การสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทย และแสดงผลด้วยภาพ เพื่อใช้ประกอบการตัดสินใจในการเลือกงาน

กานต์ สิริกานต์กุล

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2566

EXTRACTING DATA FROM IT JOB SEARCH WEBSITES IN THAILAND AND CREATE DATA VISUALIZATION FOR DECISION-MAKING IN JOB SELECTION.

Karn Sirikarnkul

A Term Paper Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology
Graduate Studies
Thai-Nichi Institute of Technology
Academic Year 2023

หัวข้อสารนิพนธ์	การสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทย และแสดงผลด้วยภาพเพื่อใช้ประกอบการตัดสินใจในการเลือกงาน
โดย	กานต์ สิริกานต์กุล
สาขาวิชา	เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษาสารนิพนธ์	ดร.ศรายุทธ นนท์ศิริ

บัณฑิตศึกษา สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้นับสารนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....รองอธิการบดีฝ่ายวิชาการ

(รองศาสตราจารย์ ดร.วรากร ศรีเชวงทรัพย์)

วันที่.....เดือน.....พ.ศ. .....

คณะกรรมการสอบสารนิพนธ์

.....ประธานกรรมการ

(ดร.ประมุข บุญเสี่ยง)

.....กรรมการ

(ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์)

.....อาจารย์ที่ปรึกษาสารนิพนธ์

(ดร.ศรายุทธ นนท์ศิริ)

กานต์ สิริกานต์กุล: การสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทย และแสดงผลด้วยภาพเพื่อใช้ประกอบการตัดสินใจในการเลือกงาน อาจารย์ที่ปรึกษา: ดร. ศรายุทธ นนท์ศิริ, 114 หน้า.

ในปัจจุบันเทคโนโลยีสารสนเทศได้พัฒนาอย่างรวดเร็วแบบก้าวกระโดดจึงมีบทบาทสำคัญอย่างมากในการใช้ชีวิตประจำวันของผู้คน โดยเฉพาะอย่างยิ่งในปัจจุบันเกิดวิกฤตจากการแพร่ระบาดโควิด 19 ทำให้องค์กรต่าง ๆ ได้นำเทคโนโลยีสารสนเทศมาช่วยเพิ่มช่องทางการทำงานในบริษัท และการบริการลูกค้าขึ้นอย่างมากจนทำให้ปัจจุบันปริมาณการขยายตัวของตลาดแรงงานอาชีพสาย IT มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ส่งผลให้คนสนใจประกอบอาชีพสาย IT มากขึ้น ดังนั้นจึงมีการเปลี่ยนสายงานจากสายอื่นมายังสายงาน IT แต่ยังคงสับสนว่าตนเองเหมาะสมกับตำแหน่งงานใด และตำแหน่งงานไหนมีความต้องการในตลาดแรงงาน

งานวิจัยนี้จึงมีวัตถุประสงค์พัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติ และแสดงผลข้อมูลด้วยภาพ แนวโน้มของตำแหน่งงาน และภาษาโปรแกรมคอมพิวเตอร์ที่มีความต้องการในประเทศไทย เพื่อให้บุคคลที่ต้องการสมัครงานสาย IT ใช้ Dashboard ที่มีข้อมูลอัพเดทได้อัตโนมัติในการประกอบการตัดสินใจในการเลือกงาน การศึกษาต่อ หรือ การเข้าร่วมการอบรมสัมมนาต่าง ๆ แล้วนำความรู้ไปใช้สมัครทำงานสาย IT ที่สนใจ ผลการวิจัยพบว่าการนำระบบอัตโนมัติมาใช้เก็บข้อมูล และสร้าง Dashboard สามารถมองเห็นข้อมูลเชิงลึกของเทรนด์งานอาชีพสาย IT รวมทั้งสามารถลดเวลาในการหาข้อมูลเชิงลึกได้ พบว่าผู้ทดสอบที่สนใจ และเกี่ยวข้องกับงาน IT ได้ตอบแบบสอบถามแล้วมีความพึงพอใจด้านการออกแบบ Dashboard ในเกณฑ์ดี และผู้ทดสอบต้องการข้อมูลที่แสดงใน Dashboard มาใช้พิจารณาตัดสินใจในการเลือกงาน

บัณฑิตศึกษา ลายมือชื่อนักศึกษา.....

สาขาวิชาเทคโนโลยีสารสนเทศ ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา 2566

KARN SIRIKARNKUL: EXTRACTING DATA FROM IT JOB SEARCH WEBSITES IN THAILAND AND CREATE DATA VISUALIZATION FOR DECISION-MAKING IN JOB SELECTION.ADVISOR: DR. SARAYUT NONSIRI, 114 PP.

Nowadays, information technology has developed at an exponential rate, so it plays a very important role in people's daily lives. At present, there is a crisis due to the COVID-19 outbreak. Organizations have greatly increased the use of information technology to help in their work. At present, the amount of expansion of the IT career labor market tends to increase. As a result, more people are interested in careers in the IT field. Therefore, they change careers from other fields to the IT field but are still confused about which job positions they are suitable for and which jobs have demand in the labor market.

The objective of this research is to develop an automated system for extracting job application data from job search websites in Thailand and create a trends of demand job positions and computer programming languages in Thailand dashboard. So that people who want to apply for IT jobs can use the dashboard that automatically updates information to make decisions about applying for the interesting IT jobs. The results of the research found that using automation to collect data and create dashboards can reduce the time needed to find insights into IT career trends. Testers who are interested or involved in IT are highly satisfied with the dashboard design. Testers want the information displayed on the dashboard to be used for consideration in their decision-making process for job applications.

Graduate Studies Student's Signature.....

Field of Engineering of Technology Advisor's Signature.....

Academic Year 2023

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดี จากความกรุณาของ ดร.ศรายุทธ นนท์ศิริอาจารย์ที่ปรึกษาสารนิพนธ์ฉบับนี้ รวมทั้งคณะกรรมการทุกท่านอย่าง ดร.ประมุข บุญเสี่ยง และว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์ ที่ได้ให้คำแนะนำ เพื่อแก้ไขปรับปรุงเนื้อหาสารนิพนธ์นี้ให้ถูกต้องและครบถ้วนจนสารนิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดี จึงขอกราบขอบพระคุณอาจารย์เป็นอย่างสูง

ขอบคุณเจ้าหน้าที่บัณฑิตศึกษา คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี ไทย-ญี่ปุ่น ในการความช่วยเหลือเรื่องต่าง ๆ กับผู้วิจัย รวมทั้งเพื่อนร่วมชั้นเรียน สถาบันเทคโนโลยีไทย-ญี่ปุ่นทุกคนในการช่วยเหลือผู้วิจัยในการเรียน

กานต์ สิริกานต์กุล

สารบัญ

สารบัญ (ต่อ)

สารบัญตาราง

สารบัญตาราง (ต่อ)

สารบัญรูป

สารบัญรูป (ต่อ)

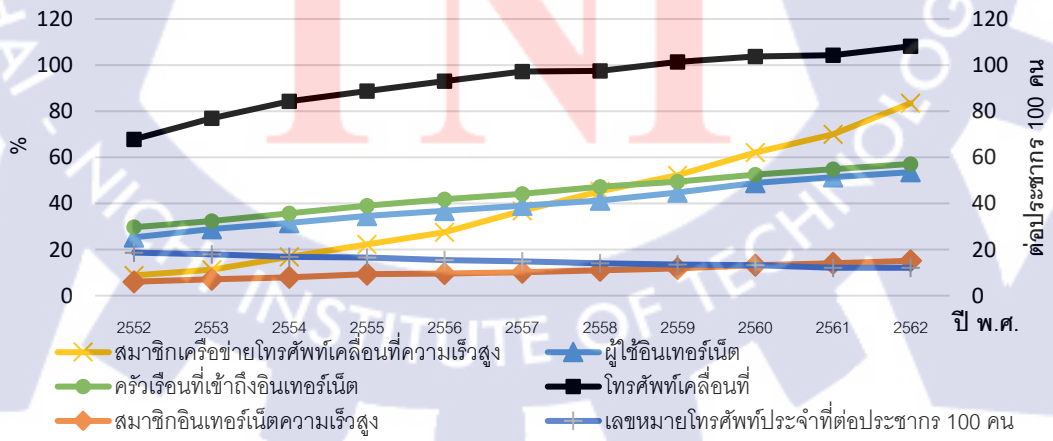
สารบัญรูป (ต่อ)

บทที่1
บทนำ

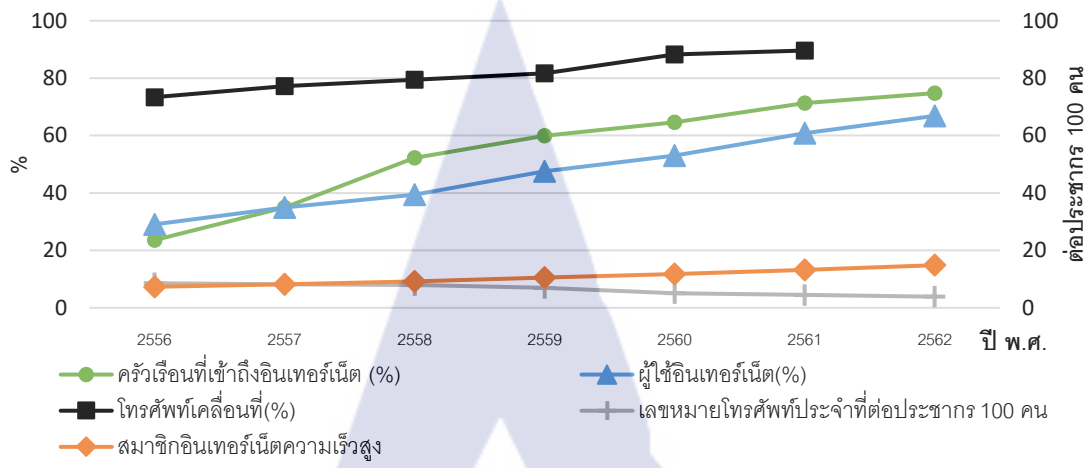
1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันเทคโนโลยีสารสนเทศได้พัฒนาอย่างรวดเร็ว และก้าวกระโดด จึงมีบทบาทสำคัญอย่างมากในการใช้ชีวิตประจำวันของผู้คน เช่น การสั่งซื้อสินค้า และบริการต่าง ๆ ผ่านทางออนไลน์ การติดต่อสื่อสารประชุมทางไกลผ่านระบบอินเตอร์เน็ต ทำให้ภาครัฐนำเทคโนโลยีสารสนเทศมาใช้ขับเคลื่อนนโยบายทางเศรฐกิจของประเทศ และทางเอกชนก็นำเทคโนโลยีสารสนเทศอย่างการนำระบบคอมพิวเตอร์มาใช้ช่วยในการออกแบบ การผลิต และการติดต่อสื่อสารงานบริการลูกค้า โดยเฉพาะอย่างยิ่งในปัจจุบันเกิดวิกฤตจากการแพร่ระบาดโควิด 19 ทำให้องค์กรต่าง ๆ ได้นำเทคโนโลยีสารสนเทศมาช่วยเพิ่มช่องทางการทำงานในบริษัท และการบริการลูกค้าขึ้นอย่างมาก

จากข้อมูลจากสำนักงานสถิติแห่งชาติ ได้แสดงข้อมูลในปี พ.ศ. 2552 – 2562 เกี่ยวกับการใช้เทคโนโลยีสารสนเทศ และการสื่อสารของทั่วโลกรวมทั้งภาคครัวเรือนของประเทศไทยด้วย แสดงให้เห็นถึงการพัฒนาการใช้เทคโนโลยีสารสนเทศ และการสื่อสารของโลกมีแนวโน้มที่เพิ่มมากขึ้นในทุก ๆ ปี โดยเฉพาะสมาชิกเครือข่ายโทรศัพท์เคลื่อนที่ความเร็วสูงเพิ่มขึ้นอย่างรวดเร็วที่สุด ส่วนผู้ใช้โทรศัพท์เคลื่อนที่ ผู้ใช้อินเทอร์เน็ต สมาชิกอินเทอร์เน็ตความเร็วสูง และครัวเรือนที่เข้าถึงอินเทอร์เน็ตยังคงเติบโตอย่างต่อเนื่อง ซึ่งสวนทางกับการลดลงของเลขหมายโทรศัพท์ประจำที่ ดังรูปที่ 1.1 การพัฒนา ICT ของทั่วโลก พ.ศ. 2552 – 2562 [1] ส่วนภาคครัวเรือนของประเทศไทยมีแนวโน้มเหมือนกับของทั่วโลก ดังรูปที่ 1.2 การพัฒนา ICT ของประเทศไทย พ.ศ. 2556 – 2562 [1] ซึ่งรายละเอียดนี้บ่งบอกว่าเทคโนโลยีสารสนเทศในประเทศไทยมีบทบาทสำคัญมากขึ้นต่อชีวิตประจำวัน

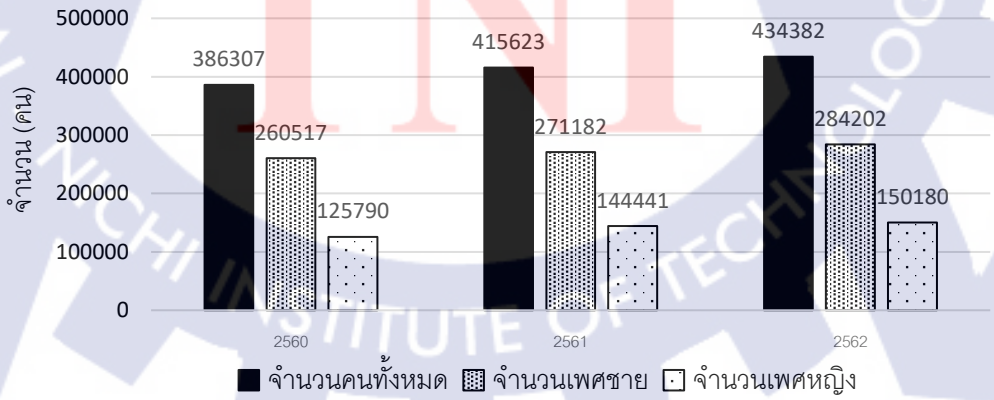


รูปที่ 1.1 การพัฒนา ICT ของทั่วโลก พ.ศ. 2552 – 2562 [1]



รูปที่ 1.2 การพัฒนา ICT ของประเทศไทย พ.ศ. 2556 – 2562 [1]

ข้อมูลจากสำนักงานสถิติแห่งชาติ ได้แสดงข้อมูลจำนวนผู้ทำงานด้านเทคโนโลยีสารสนเทศและการสื่อสาร พ.ศ. 2560-2562 แสดงให้ทราบได้ว่าผู้ทำงานด้านเทคโนโลยีสารสนเทศ และการสื่อสารมีแนวโน้มเพิ่มขึ้นตามการพัฒนา และการใช้เทคโนโลยีสารสนเทศ และการสื่อสารในประเทศไทยที่มีแนวโน้มเพิ่มมากขึ้นในทุก ๆ ปี ดังรูปที่ 1.3 จำนวนผู้ทำงานด้านเทคโนโลยีสารสนเทศ และการสื่อสาร พ.ศ. 2560 – 2562 [1] และข้อมูลจากสำนักงานสถิติแห่งชาติ ยังได้แสดงข้อมูลสัดส่วนผู้ทำงานด้านเทคโนโลยีสารสนเทศ และการสื่อสารจำแนกตามการศึกษา แสดงให้เห็นว่าผู้ทำงานด้านเทคโนโลยีสารสนเทศ และการสื่อสารที่สำเร็จการศึกษาในด้านนี้มีสัดส่วน 44.8% และที่ไม่ได้สำเร็จการศึกษาในด้านนี้มีสัดส่วนมากถึง 55.2% จากข้อมูลดังกล่าวแสดงให้เห็นว่าผู้ทำงานสายงานอื่นได้เปลี่ยนสายงานมายังสายงาน IT เป็นจำนวนมาก



รูปที่ 1.3 จำนวนผู้ทำงานด้านเทคโนโลยีสารสนเทศ และการสื่อสาร พ.ศ. 2560 – 2562 [1]

ข้อมูลจากสำนักงานสภานโยบายการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรมแห่งชาติ ได้แสดงข้อมูลแนวโน้มความต้องการบุคลากรในอุตสาหกรรม และทิศทางนโยบายการพัฒนากำลังคนของประเทศ พ.ศ. 2563-2567 [2] ซึ่งในอุตสาหกรรมดิจิทัลที่กล่าวถึงแบ่งออกเป็น 6 อุตสาหกรรมย่อย ซึ่งในแต่ละอุตสาหกรรมมีแนวโน้มการเติบดังนี้

1. อุตสาหกรรมการพัฒนาและให้บริการซอฟต์แวร์ มีแนวโน้มการเติบโตอย่างต่อเนื่อง โดยเฉพาะจากการเติบโตของการใช้อุปกร์เคลื่อนที่ต่าง ๆ เช่น สมาร์ทวอทช์ สมาร์ทโฟน และ แท็บเล็ต รวมทั้งนำซอฟต์แวร์ไปใช้ในอุตสาหกรรมต่าง ๆ เพื่อเตรียมความพร้อมเข้าสู่ยุค Thailand 4.0

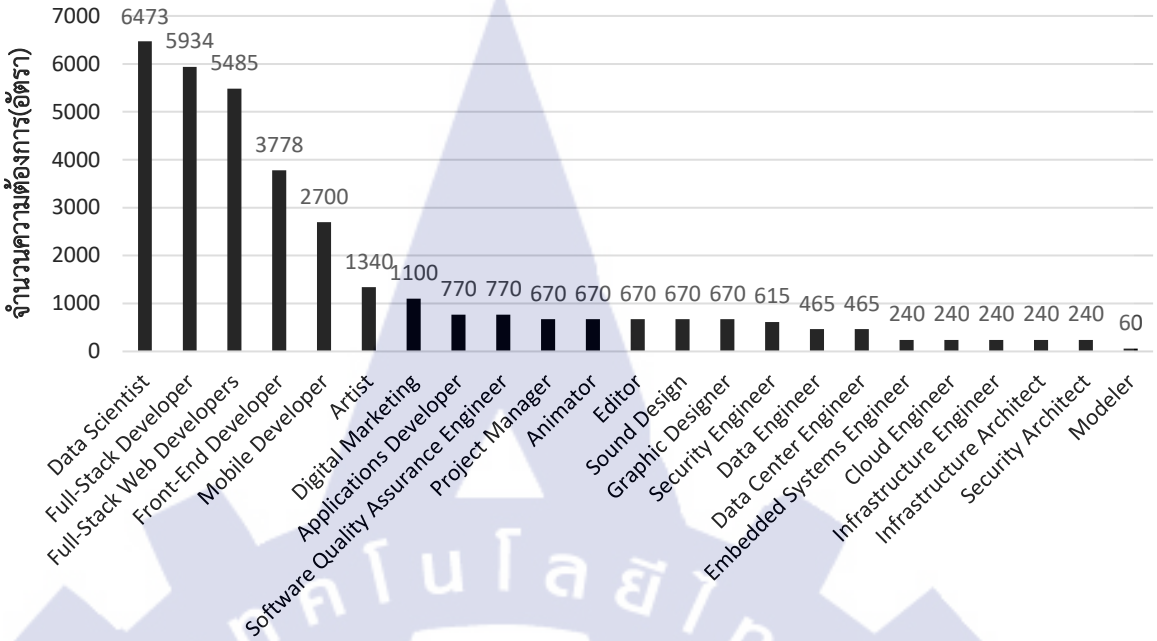
2. อุตสาหกรรมพาณิชย์อิเล็กทรอนิกส์ทั้งในและต่างประเทศ มีแนวโน้มการเติบโตอย่างต่อเนื่องจากพฤติกรรมการซื้อสินค้าของผู้บริโภคที่นิยมซื้อสินค้าผ่านระบบอินเตอร์เน็ตมากขึ้น การเกิดขึ้นของพร้อมเพย์ และสังคมไร้เงินสด

3. อุตสาหกรรมการจัดตั้งศูนย์รวบรวมและวิเคราะห์ข้อมูลผู้บริโภค และอุตสาหกรรมการให้บริการเกี่ยวกับหน่วยจัดเก็บข้อมูล การประมวลผลออนไลน์ และการป้องกันอันตรายจากอินเทอร์เน็ต มีแนวโน้มการเติบโตอย่างต่อเนื่องจากแต่ละบริษัทมีการสำรองข้อมูลในจำนวนมหาศาล ทำให้บริษัทไทยที่มีความสนใจ Data Center และ Cloud Service ในไทยมากขึ้น

4. อุตสาหกรรมการพัฒนาเมืองอัจฉริยะ มีแนวโน้มการเติบโตอย่างต่อเนื่องจากการที่ภาครัฐมีนโยบายเร่งด่วนในการพัฒนาเมืองอัจฉริยะจากการเกิดขึ้นของ ASEAN Smart Cities Network (ASCN) ซึ่งเป็นความร่วมมือของผู้นำอาเซียนร่วมกันก่อตั้งขึ้นมีเป้าหมายหลักในการสร้างเครือข่ายเมืองอัจฉริยะอาเซียนระหว่างประเทศสมาชิกอาเซียน เพื่อความเป็นอยู่ที่ดีของประชาชนในอาเซียนด้วยการใช้เทคโนโลยีในการขับเคลื่อน

5. อุตสาหกรรมสื่อสร้างสรรค์และแอนิเมชัน มีแนวโน้มการเติบโตอย่างต่อเนื่องจากการได้รับการยอมรับจากต่างประเทศในด้านคุณภาพของแอนิเมชัน

สำนักงานสภานโยบายการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรมแห่งชาติ มีการคาดการว่าอุตสาหกรรมดิจิทัลโดยภาพรวมจะมีความต้องการบุคลากรในอนาคตอยู่ที่ประมาณ 34,505 ตำแหน่ง ในอีก 5 ปีข้างหน้า โดยมีตำแหน่งงานและจำนวนความต้องการในแต่ละตำแหน่ง ดังรูปที่ 1.5 Digital Projected Talent Demand [2]



รูปที่ 1.5 Digital Projected Talent Demand [2]

จากทั้งหมดที่กล่าวมาแสดงให้เห็นว่าปัจจุบันปริมาณการขยายตัวของตลาดแรงงานอาชีพสาย IT มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ส่งผลให้คนสนใจจะประกอบอาชีพสาย IT มากขึ้น ดังนั้นจึงมีการเปลี่ยนสายงานจากสายอื่นมายังสายงาน IT แต่ยังคงสับสนว่าตนเองเหมาะสมกับตำแหน่งงานใดและตำแหน่งงานไหนมีความต้องการในตลาดแรงงาน

ในงานสารนิพนธ์นี้จึงนำเสนอการสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทยสำหรับทำสถิติเพื่อใช้ประกอบการตัดสินใจในการเลือกงาน โดยอาศัยเทคนิคการสกัดข้อมูลบนเว็บไซต์ (web scraping) จากเว็บไซต์หางานภายในประเทศไทยเพื่อสกัดข้อมูลต่าง ๆ เกี่ยวกับตำแหน่งงานที่บริษัทต้องการรับสมัครนำมาจัดกลุ่มข้อมูล และแสดงผลข้อมูลด้วยภาพ เพื่อใช้ประกอบการตัดสินใจในการเลือกงาน และตัดสินใจในการศึกษาต่อ หรือ ทำการเข้าร่วมการอบรมสัมมนาต่าง ๆ เพื่อนำความรู้ไปใช้สมัครทำงานสาย IT ที่เลือก

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ ออกแบบ และพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติ

2. เพื่อแสดงผลข้อมูลด้วยภาพ แนวโน้มของตำแหน่งงาน และภาษาโปรแกรมคอมพิวเตอร์ที่มีความต้องการในประเทศไทย

1.3 ขอบเขตของการวิจัย

1. เพื่อออกแบบระบบการสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทย 4 เว็บไซต์ ดังนี้

1.1. https://www.jobbkk.com/

1.2. https://th.jobsdb.com/th/th

1.3. https://www.jobthai.com

1.4. https://jobs.blognone.com/search

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ระบบสามารถรวบรวมข้อมูลสถิติงาน IT ล่าสุดแบบอัตโนมัติได้

2. ระบบสามารถแสดงข้อมูลสถิติงาน IT ในรูปแบบการแสดงผลด้วยภาพให้ผู้สนใจที่จะเปลี่ยนสายงานมาทำงานสาย IT ได้นำไปใช้ประกอบการตัดสินใจในการเลือกงาน และการตัดสินใจในการศึกษาต่อ

1.5 นิยามศัพท์เฉพาะ

1. การสกัดข้อมูลบนเว็บไซต์ (Web scraping) [3] หมายถึง เทคนิคการสกัดข้อมูลต่าง ๆจากหน้าเว็บไซต์ที่เปิดเผยต่อสาธารณะ เช่น ตัวเลข ข้อความ และรูปภาพ

2. การแสดงข้อมูลด้วยภาพ (Data Visualization) [3][4] หมายถึง การนำข้อมูลที่ผ่านการวิเคราะห์ และต้องการมานำเสนอผ่านภาพ อย่างเช่น กราฟ แผนภูมิ แผนที่ และรูปภาพ ทำให้สามารถสื่อสารทำความเข้าใจต่อข้อมูลปริมาณมาก ๆ ที่ต้องการแสดงได้อย่างรวดเร็วมองเห็นภาพรวมได้ชัดเจน

บทที่ 2
หลักการ ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ปัจจุบันปริมาณการขยายตัวของตลาดแรงงาน IT มีเพิ่มขึ้นอย่างต่อเนื่องทำให้คนสนใจจะประกอบอาชีพสาย IT มากขึ้น จึงต้องการข้อมูลเกี่ยวกับอาชีพสาย IT เพื่อใช้ในใช้ประกอบการตัดสินใจในการเลือกงาน และตัดสินใจในการศึกษาต่อ หรือ ทำการเข้าร่วมการอบรมสัมมนาต่อ จึงจัดทำงานวิจัยนี้ขึ้นผู้ศึกษาจึงได้สืบค้นหลักการ ทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยมีหัวข้อดังต่อไปนี้

2.1 ปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงาน

จากการศึกษางานวิจัยจำนวนหนึ่งทำให้ได้ทราบปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงาน โดยงานวิจัยด้านสังคมศาสตร์เหล่านี้ได้อ้างอิงถึงทฤษฎี ดังนี้

2.1.1 หลักการ และทฤษฎีด้านลักษณะประชากรศาสตร์

หลักการ และทฤษฎีด้านลักษณะประชากรศาสตร์ [5][6] เป็นตัวบ่งบอกลักษณะของประชากรในแต่ละตัวบุคคล และมีอิทธิพลอย่างมากต่อการจูงใจ อำนาจ และการเมืองหรือวัฒนธรรมขององค์กร โดยประกอบด้วย เพศ อายุ ระดับการศึกษา สถานภาพ ประสบการณ์ในการทำงาน และรายได้

2.1.2 หลักการ และทฤษฎีแรงจูงใจ

หลักการ และทฤษฎีแรงจูงใจ [7][8] ที่งานวิจัยต่าง ๆ อ้างอิง คือ ทฤษฎีลำดับขั้นความต้องการซึ่งเป็นแนวคิดของมาสโลว์ (Maslow’s Hierachy Of Needs Theory) เป็นการอธิบายพฤติกรรมของมนุษย์ว่ามีความต้องการตามลำดับขั้น 5 ขั้น ดังนี้ ความต้องการของร่างกาย ความต้องการในความมั่นคง และความปลอดภัย ความต้องการที่ได้รับการยอมรับจากสังคม ความภาคภูมิใจในตนเอง และความสำเร็จในชีวิตตนเอง ตามลำดับ

2.1.3 หลักการการตัดสินใจเลือกอาชีพ

หลักการการตัดสินใจเลือกอาชีพ [9] การตัดสินใจเลือกอาชีพ เป็นการตัดสินใจที่สำคัญมากในชีวิต ซึ่งเป็นการวางแผนชีวิตที่จำเป็นต้องเริ่มตั้งแต่วัยเรียนโดยมีปัจจัยในการตัดสินใจที่สำคัญ 2 ข้อ ดังนี้

1. ปัจจัยภายนอก ซึ่งเป็นข้อมูลเกี่ยวกับอาชีพ ได้แก่ แนวโน้มของตลาดแรงงาน ลักษณะงาน สภาพแวดล้อมของงาน คุณสมบัติของผู้ประกอบอาชีพ การเข้าประกอบอาชีพ รายได้ ความก้าวหน้า และการกระจายของผู้ประกอบอาชีพ

2. ปัจจัยภายใน ซึ่งเป็นลักษณะส่วนบุคคล ได้แก่ ความสนใจ บุคลิกภาพ ความถนัด ทักษะ ประสบการณ์ วุฒิภาวะทางอาชีพ เพศ อายุ สุขภาพ และความแข็งแรงของร่างกาย

โดยในแต่ละงานวิจัยที่ผู้วิจัยศึกษาได้พบปัจจัยต่าง ๆ ที่มีผลต่อแรงจูงใจในการพิจราณา และตัดสินใจเลือกทำงานดังนี้

พรทวี เถื่อนคำแสน และ บุญญรัตน์ สัมพันธ์วัฒนชัย [5] ศึกษาปัจจัยเพื่อหาปัจจัยที่มีผลต่อแรงจูงใจในการพิจราณา และตัดสินใจเลือกทำงานของพนักงานในนิคมอุตสาหกรรมเหมราช ผู้วิจัยพบว่ามีปัจจัย 3 อย่างดังนี้ อย่างแรกปัจจัยด้านประชากรศาสตร์ คือ เงินเดือนที่ได้ต่อเดือน ระดับวุฒิการศึกษา อายุการทำงาน เพศสภาพ อายุ และสถานภาพสมรส อย่างที่สองปัจจัยด้านจูงใจ คือ ลักษณะหน้าที่ที่ได้รับของงาน ความสำเร็จในการทำงาน ความก้าวหน้าในอาชีพ และการยกย่องเชิดชู อย่างสุดท้ายปัจจัยค้ำจุน คือ เงินเดือนที่ได้ต่อเดือน นโยบายการทำงานของบริษัทที่ทำงาน ความมั่นคงในอาชีพการทำงาน และสภาพในการทำงาน

พัชรินทร์ ชุติพงศ์รุ่งโรจน์ [6] ศึกษาปัจจัยเพื่อหาปัจจัยที่มีผลต่อการพิจราณาเลือกเข้าทำงานกับบริษัทงานตรวจทาน และทำบัญชีซึ่งเป็นบริษัทใหญ่ 4 บริษัท พบว่ามีปัจจัย 2 อย่างดังนี้ ปัจจัยด้านประชากรศาสตร์ ได้แก่ เพศ ผลการเรียน ระดับรายได้ต่อเดือนของครอบครัว และภูมิลำเนา อีกปัจจัยคือ ปัจจัยด้านการจูงใจในการตัดสินใจเลือกงาน ได้แก่ ลักษณะงาน โอกาสที่ได้เพิ่มประสบการณ์ และพัฒนาตนเอง โอกาสก้าวหน้าในหน้าที่การงาน ความมั่นคงในงาน ความสัมพันธ์กับเพื่อนร่วมงานผลตอบแทน และสวัสดิการ ชื่อเสียงองค์กร และค่านิยม

พรพิมล ชราศักดิ์ [7] ได้ศึกษาหาปัจจัยที่มีผลต่อพฤติกรรมการเลือกประกอบอาชีพของคนว่างงาน พบว่ามีปัจจัย 3 อย่างดังนี้ การตัดสินใจเลือกประกอบอาชีพของคนว่างงาน ได้แก่ การตระหนักในปัญหา การระบุ และวิเคราะห์สาเหตุของปัญหา การแสวงหาทางเลือก การประเมิน และตัดสินใจเลือก และติดตามการดำเนินงาน ปัจจัยด้านพฤติกรรมได้แก่ ด้านความรู้สึก ด้านทัศนคติ และด้านภูมิหลังครอบครัว และปัจจัยส่วนบุคคล ได้แก่ เพศ อายุ ระดับการศึกษา สถานภาพ ประสบการณ์การทำงาน และรายได้

ณัฐกร มั่นจิตต์ และ กตัญญู หิรัญญสมบูรณ์ [8] ศึกษาปัจจัยเพื่อหาปัจจัยที่มีผลต่อแรงจูงใจในการพิจราณา และตัดสินใจเลือกทำงานของคนไทยที่ต้องการไปทำงานในประเทศเกาหลีอย่างถูกต้องตามกฎหมาย พบว่ามีปัจจัย 2 อย่างดังนี้ อย่างแรกปัจจัยส่วนบุคคล ได้แก่ ระดับรายได้ต่อเดือนในปัจจุบันที่ได้ในประเทศไทย ประเภทอาชีพที่ต้องการสมัครไปทำงานในประเทศเกาหลี เพศ สภาพ อายุ และสถานภาพสมรส อย่างที่สองปัจจัยด้านแรงจูงใจ ได้แก่ ลักษณะของงานโอกาสได้

เรียนรู้ความรู้ใหม่และการพัฒนาตนเอง ความมั่นคง ผลตอบแทน และสวัสดิการ โอกาสเจริญก้าวหน้าในหน้าที่การงาน และค่านิยม

2.2 การสกัดข้อมูลจากอินเทอร์เน็ต

การสกัดข้อมูลจากอินเทอร์เน็ต [3][10][11] เป็นวิธีการที่อ่านข้อมูลปริมาณมากจากเว็บไซต์ที่มีอยู่ในอินเทอร์เน็ต เช่น อ่านหน้าเว็บไซต์ย่อยทุกหน้าที่มีในที่อยู่เว็บไซต์ที่ต้องการแล้วทำการคัดลอกลงที่จัดเก็บ จะได้ข้อมูลดิบทั้งหน้าเว็บไซต์แต่ละหน้าซึ่งข้อมูลจัดเก็บจะมีได้ทั้งรูปภาพและข้อความ เทคนิคนี้จะเรียกว่า Web scraping [3] เป็นเทคนิคการดึงข้อมูลพวกข้อความ และรูปภาพทุกหน้าจากเว็บไซต์ทั่วไปไปจัดเก็บเว็บไซต์ทั่วไปหมายถึงเว็บไซต์ที่จะเปิดดูหน้าเว็บไซต์ได้ด้วยเว็บบราวเซอร์เท่านั้น และเป็นเว็บไซต์ที่เปิดเผยต่อสาธารณะ อย่างเช่น เว็บไซต์หางาน โดยมีหลักการทำงานคือ ใช้โปรแกรมที่จำลองการอ่านหน้าเว็บไซต์โดยเริ่มต้นจากการอ่านข้อมูลในหน้าหลักของเว็บไซต์ และหาลิงค์ (link) ทุกอันที่มีอยู่บนหน้าเว็บไซต์นั้น ๆ เพื่อเข้าไปอ่านข้อมูลหน้าเว็บไซต์ย่อยอื่น ๆ จนสุดทางแล้วกลับไปหน้าแรกใหม่ และทำอย่างนี้ไปเรื่อย ๆ จนครบแล้วทำการจัดเก็บข้อมูลแต่ละหน้าเว็บไซต์ที่อ่านในรูปแบบ HTML

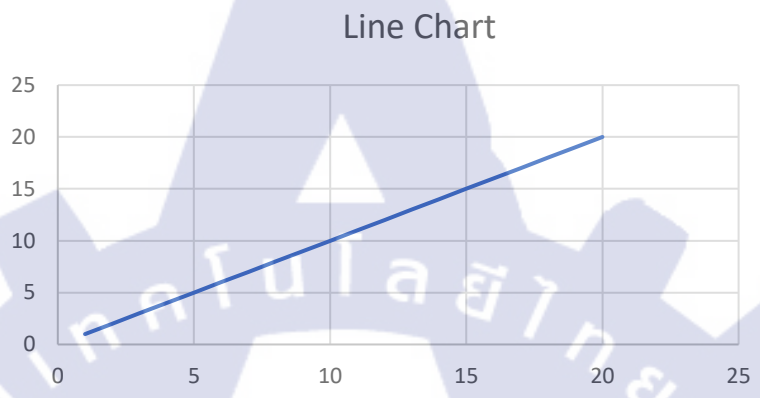
ส่วนการดึงข้อมูลมาโดยอัตโนมัติจากเว็บไซต์ที่มีส่วนต่อประสานโปรแกรมประยุกต์ (API) [12] ซึ่งเป็นเว็บไซต์ที่สามารถใช้งานผ่านเว็บบราวเซอร์ตามปกติแล้ว จะยังมี API ที่ให้ข้อมูลกับโปรแกรมประยุกต์หรือเว็บไซต์อื่นเรียกใช้บริการต่าง ๆ ได้ สามารถทำได้โดยวิเคราะห์ API ของเว็บไซต์ และรูปแบบข้อมูลที่จะได้รับ แล้วพัฒนาโปรแกรมที่เรียกใช้ API ของเว็บไซต์นั้น ๆ เพื่อดึงข้อมูลมาเป็นระยะ ๆ แปลงไฟล์ และจัดเก็บ

2.3 Data Visualization

เป็นการนำข้อมูลที่ผ่านการวิเคราะห์ และต้องการมานำเสนอผ่านภาพ อย่างเช่น กราฟ แผนภูมิ แผนที่ และรูปภาพ การแสดงข้อมูลเป็นภาพทำให้สามารถสื่อสารทำความเข้าใจต่อข้อมูลปริมาณมาก ๆ ที่ต้องการแสดงได้อย่างรวดเร็ว และทำความเข้าใจได้ในเวลาอันสั้น มองเห็นภาพรวมได้ชัดเจน ง่ายต่อการจดจำ รวมทั้งช่วยในการคาดคะเน และพยากรณ์ สิ่งที่จะเกิดขึ้นในอนาคต และยังช่วยแสดงปัญหาที่เกิดขึ้นได้ การทำ Data Visualization ให้เลือกประเภทของ Data Visualization ให้เข้ากับลักษณะความสัมพันธ์ของข้อมูลที่ต้องการแสดง เนื่องจากเครื่องมือแสดง Data Visualization แต่ละอย่างมีจุดเด่นในการแสดงลักษณะความสัมพันธ์ของข้อมูลที่แตกต่างกัน โดยในส่วนนี้จะยกตัวอย่างรูปกระบวนการทำ Data Visualization [3][4][13] ดังรูปต่อไปนี้

2.3.1 แผนภูมิเส้น (Line Chart)

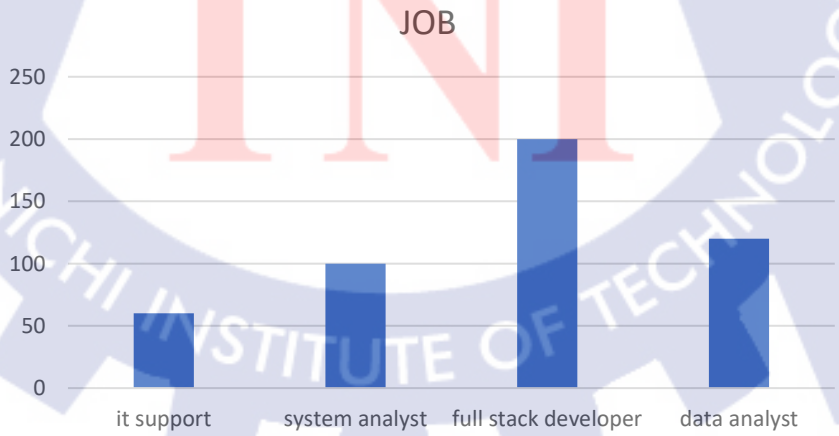
แผนภูมิเส้นเป็นวิธีการนำเสนอข้อมูล โดยใช้จุด และเส้นตรงที่ลากเชื่อมต่อจุดแต่ละจุด ซึ่งบอกแต่ละจุดจะบอกค่าของข้อมูล แผนภูมิเส้นเหมาะกับข้อมูลที่มีความต่อเนื่อง และแสดงการเปลี่ยนแปลงอย่างต่อเนื่อง



รูปที่ 2.1 แผนภูมิเส้น (Line Chart)

2.3.2 แผนภูมิแท่ง (Bar Chart)

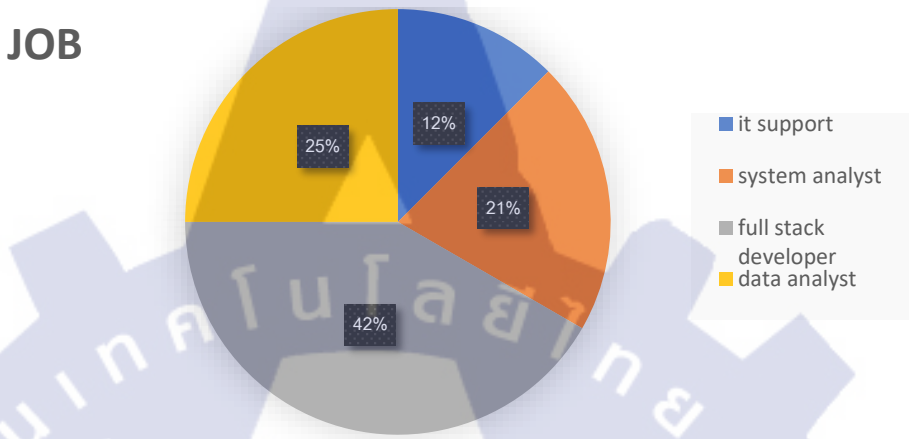
แผนภูมิแท่งเป็นแผนภูมิที่ประกอบด้วยแกนแนวนอนซึ่งบ่งบอกถึงชื่อชุดข้อมูล และแนวตั้งซึ่งบ่งบอกถึงปริมาณข้อมูล แผนภูมิแท่งเหมาะสำหรับใช้ในการเปรียบเทียบจำนวนของข้อมูลในแต่ละชุด



รูปที่ 2.2 แผนภูมิแท่ง (Bar Chart)

2.3.3 แผนภูมิวงกลม (Pie Chart)

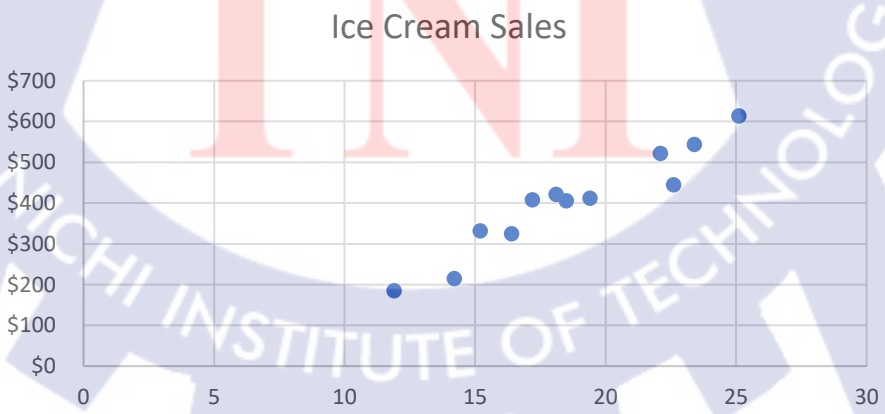
แผนภูมิวงกลมเป็นแผนภูมิที่มีลักษณะเป็นวงกลม และแบ่งเป็นช่องต่าง ๆ มักใช้แสดงสัดส่วนข้อมูลต่าง ๆ ต่อข้อมูลทั้งหมดเป็นเปอร์เซ็นต์ จำนวณสัดส่วนที่เหมาะสมมีตั้งแต่ 2 – 5 สัดส่วนหากมากเกินไปจะอ่านภาพลำบาก



รูปที่ 2.3 แผนภูมิวงกลม (Pie Chart)

2.3.4 แผนภาพกระจาย (Scatter Plot)

แผนภาพกระจายเป็นแผนภูมิที่ใช้จุดแสดงความสัมพันธ์ระหว่างตัวแปรข้อมูล 2 ตัว ข้อมูลแกน x และ แกน y เหมาะสำหรับใช้แสดงความสัมพันธ์ของข้อมูลที่มีปริมาณมาก ๆ และดูการกระจายตัวของข้อมูล



รูปที่ 2.4 แผนภาพกระจาย Scatter Plot

2.3.5 ตารางข้อมูล

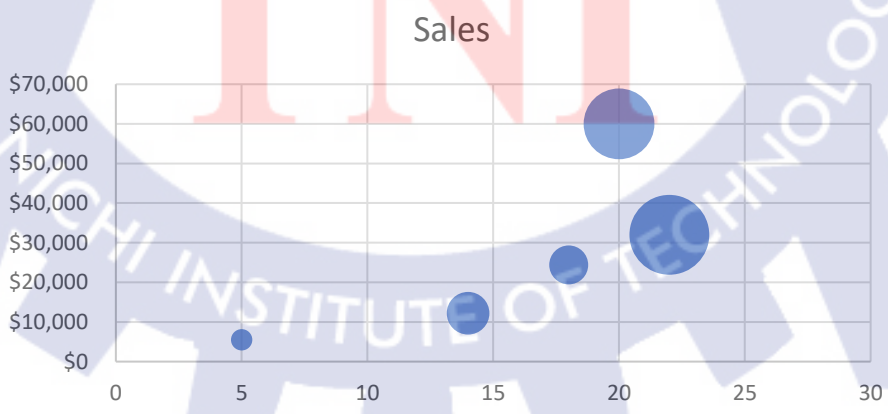
ตารางข้อมูลเป็นการจัดเรียงของข้อมูลในแถว และคอลัมน์ ซึ่งเหมาะสำหรับใช้เก็บข้อมูล และแสดงความสัมพันธ์ของข้อมูล

Temperature °C	Ice Cream Sales
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408

รูปที่ 2.5 ตารางข้อมูล

2.3.6 แผนภูมิฟองอากาศ (Bubble Chart)

แผนภูมิฟองอากาศเป็นแผนภูมิที่ใช้ฟองอากาศแสดงความสัมพันธ์ระหว่างตัวแปรข้อมูล 2 ตัว ข้อมูลแกน x และ แกน y คล้ายกับแผนภาพกระจายเพียงแต่ใช้ฟองอากาศแทนจุด ซึ่งตัวขนาดฟองอากาศยังบอกถึงค่าข้อมูลในแต่ละฟองอากาศ



รูปที่ 2.6 แผนภูมิฟองอากาศ (Bubble Chart)

2.3.7 แผนภูมิความร้อน (heat map)

แผนภูมิความร้อนเป็นแผนภูมิที่แสดงความสัมพันธ์ของข้อมูลโดยใช้คู่สี และความเข้มข้นของสีแทนปริมาณหรือความถี่ เช่น ตารางแสดงข้อมูลอุณหภูมิในแต่ละพื้นที่ในช่วงเวลาของแต่ละเดือน

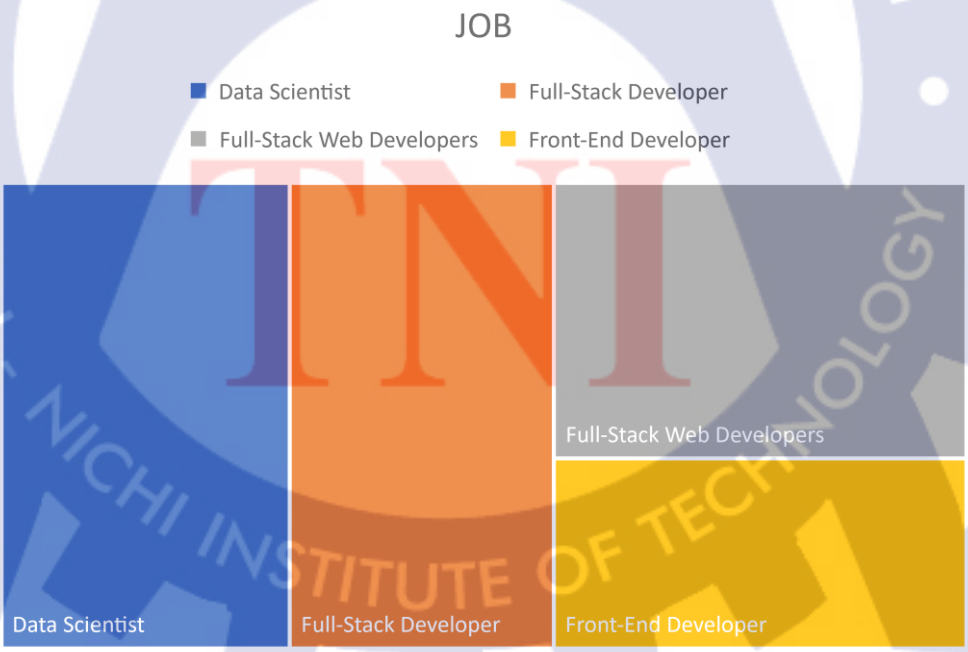
Air Temperature (°F)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Avg.	40.2	41.8	47.4	56.3	65	71.3	78.5	79.5	73.6	64.3	54.2	44.4
Min.	35.6	36.5	41.1	49.6	58.8	66.5	73.8	74.8	69.2	59.7	49.5	40
Max.	45.8	48.4	55	64.3	72.3	77.2	84.2	85.3	78.9	69.8	59.8	49.7

รูปที่ 2.7 แผนภูมิความร้อน (heat map)

2.3.8 แผนภูมิต้นไม้ (Tree Map)

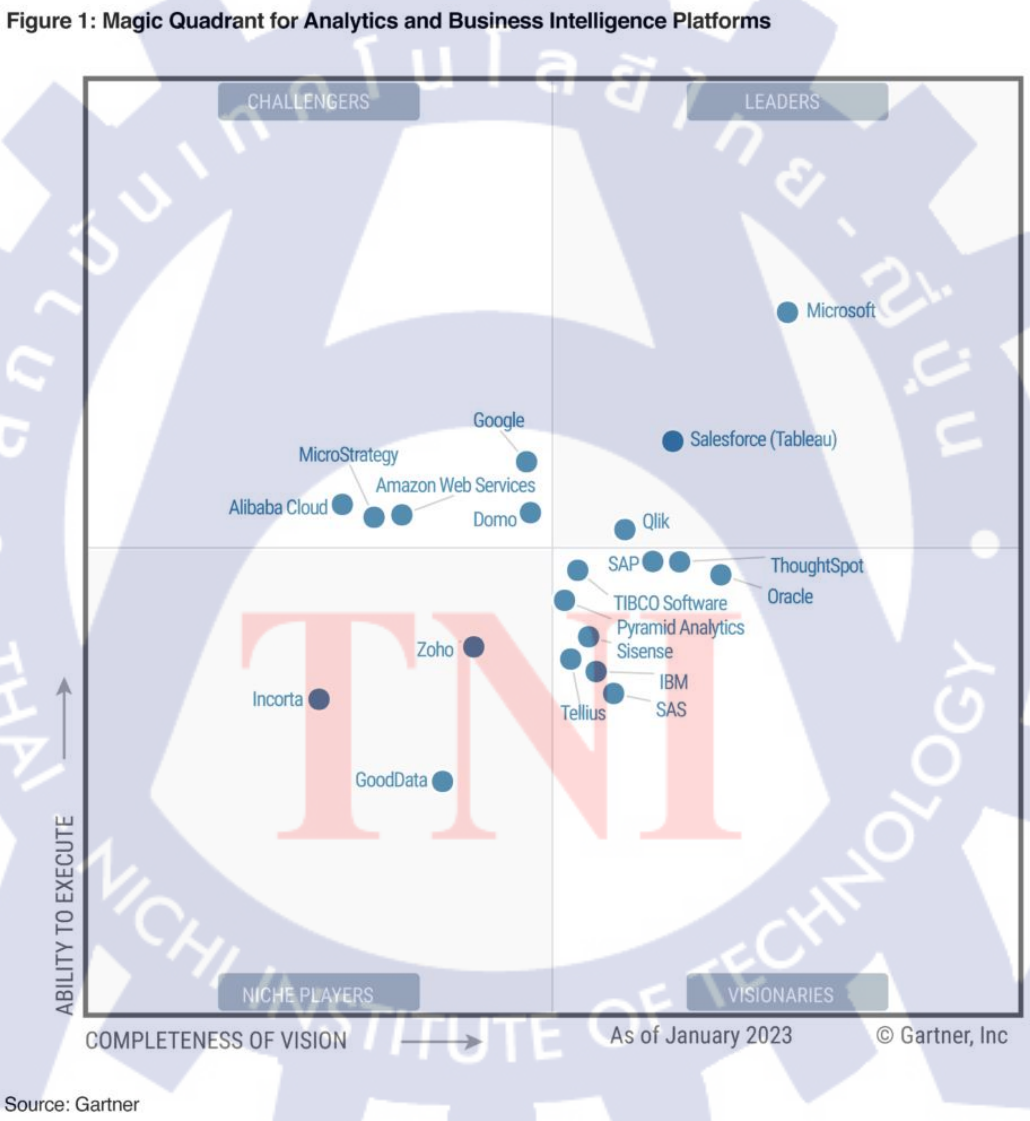
แผนภูมิต้นไม้เป็นแผนภูมิที่มีนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ เหมาะสำหรับการเปรียบเทียบข้อมูลที่ซับซ้อน และมีหลายปัจจัย และต้องการจัดกลุ่มข้อมูล และแบ่งตามความสำคัญของข้อมูล



รูปที่ 2.8 แผนภูมิต้นไม้ (Tree Map)

2.4 Power BI

Power BI [14][15][16][17] เป็นโปรแกรมสำหรับใช้ในการวิเคราะห์ข้อมูลเชิงธุรกิจ (Business Intelligence) และสร้างรายงานการนำเสนอผ่านภาพ (Visualization) ได้อย่างรวดเร็วมีตัวเลือกหลากหลาย และทันสมัยสามารถอัพเดทข้อมูลได้รวดเร็วสามารถเชื่อมต่อแหล่งข้อมูลที่เป็นที่นิยมจำนวนมาก เช่น Excel, CSV, Database, Website และ File รูปแบบต่าง ๆ อีกทั้งสามารถรองรับการดูรายงานได้จากหลากหลายอุปกรณ์ทั้ง คอมพิวเตอร์ โน๊ตบุ๊ค สมาร์ทโฟน และ แท็บเล็ต Power BI ยังสามารถทำงานร่วมกับโปรแกรมในกลุ่ม Microsoft Power Platform ได้อีกด้วย และยังเป็นโปรแกรมที่ได้รับความนิยมอันดับต้น ๆ ดังรูปที่ 2.9



รูปที่ 2.9 magic quadrant for analytics and business intelligence platforms [14]

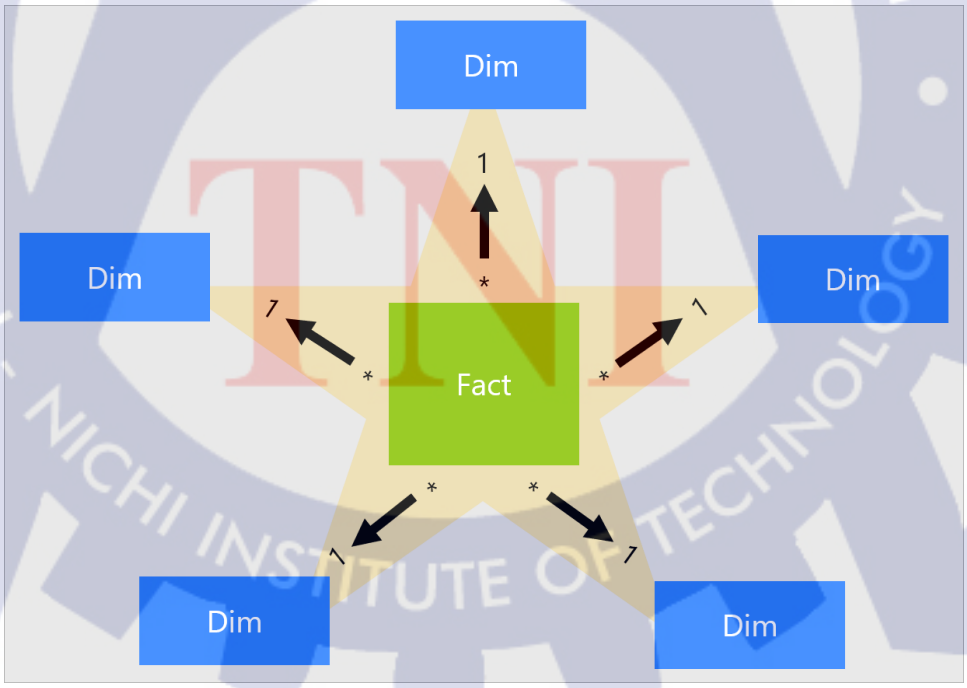
รูปแบบการทำงานพื้นฐานของ Power BI เกิดจากองค์ประกอบหลายอย่างที่ทำงานร่วมกัน 3 อย่าง ดังนี้

2.4.1 Power BI Desktop

Power BI Desktop ซึ่งเป็นโปรแกรมบนเครื่องคอมพิวเตอร์ใช้สำหรับสร้างรายงานซึ่งมีกระบวนการ 4 ขั้นตอนดังนี้

1. นำเข้าข้อมูล [18] เป็นขั้นตอนการนำเข้าข้อมูลจากแหล่งข้อมูลต่าง ๆ ที่เป็นที่นิยมจำนวนมาก เช่น Excel CSV XML JSON PDF Database MySQL Oracle IBM และ Website แหล่งข้อมูลที่ใช้ในขั้นตอนนำเข้าข้อมูล มีข้อจำกัดบางประการขนาดของชุดข้อมูลนำเข้าต้องไม่เกินพื้นที่จัดเก็บใน Power BI Service ซึ่งมีขนาดเพียง 1 GB ถ้าจะใช้เกินความจุ 1 GB ต้องสมัครใช้ Power BI Premium

2. สร้าง Data Model [19] เป็นขั้นตอนการจัดระเบียบข้อมูลต่าง ๆ เพื่อแสดงความสัมพันธ์ของข้อมูล ซึ่งการออกแบบ Data Model ที่ดี จะช่วยให้ประสิทธิภาพในการแสดงผลเร็วขึ้น และสามารถเขียนสูตร DAX ได้ง่ายขึ้น โดย Data Model ที่เหมาะสมกับ Power BI คือ Data Model Star Schema ซึ่งมีลักษณะมีตารางหลัก 1 ตารางเรียกว่า Fact อยู่ตรงกลาง แล้วล้อมรอบด้วยตารางอ้างอิง เรียกว่า Dimension ดังรูปภาพที่ 2.10



รูปภาพที่ 2.10 ลักษณะของ Data Model Star Schema [19]

3. เขียนสูตร DAX (Data Analysis Expressions) [20] เป็นขั้นตอนการเขียนชุดคำสั่งเพื่อใช้ในการคำนวณ และวิเคราะห์ข้อมูล ซึ่งหลังจากสร้าง Data Model แล้วทำสามารถเขียนชุดคำสั่งสร้างคอลัมน์คำนวณ ตรวจสอบ หรือ กรองค่าต่าง ๆ ที่ต้องการเพิ่มข้ามตารางตามความสัมพันธ์ของข้อมูล

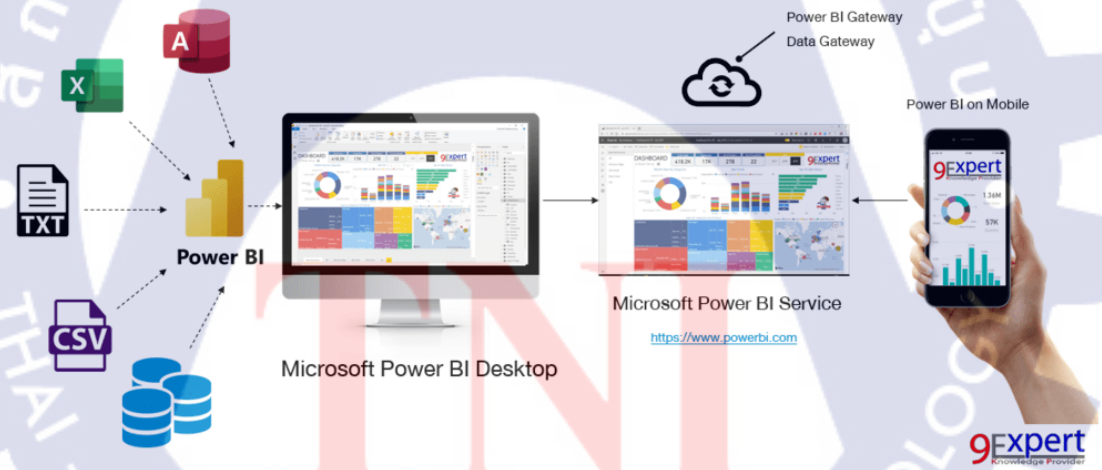
4. สร้าง Report เป็นขั้นตอนสร้าง Dashboard เพื่อแสดงผลข้อมูลที่ต้องการนำเสนอโดยใช้ Visualization รูปแบบต่างเพื่อให้ดูง่าย และสามารถตีความสั้น ๆ ในพริบตา

2.4.2 Power BI Service

Power BI Service คือส่วนของ Cloud เพื่อใช้เก็บข้อมูล และเผยแพร่รายงานให้ผู้ใช้ทางธุรกิจ หลังจากต้องการเผยแพร่รายงานที่สร้างจาก Power BI Desktop

2.4.3 Power BI Mobile

Power BI Mobile คือโปรแกรมที่แสดงผลรายงานในอุปกรณ์เคลื่อนที่ เช่น โน๊ตบุ๊ต สมาร์ทโฟน และแท็บเล็ต



รูปที่ 2.11 แสดงองค์ประกอบของ Power BI ที่ทำงานร่วมกัน 3 อย่าง [15]

2.6 งานวิจัยที่เกี่ยวข้อง

สุรจิต ภูมิคง, ธราอั่งสกุล, และจิติมนต์อั่งสกุล [21] ได้ทำงานวิจัยวิธีการสกัดข่าวด้านความปลอดภัย มีจุดประสงค์เพื่อพัฒนาวิธีการสกัดข้อมูลข่าวสารแบบอัตโนมัติให้ทันสมัย และระบบที่สามารถใช้ช่วยวางแผนในการท่องเที่ยวแบบออนไลน์ในเงื่อนไขต้องท่องเที่ยวอย่างปลอดภัย ผู้วิจัยจะออกแบบระบบสกัดข้อมูลข่าวด้านความมั่นคงความปลอดภัยอัตโนมัติ โดยต้องการดึงข้อมูลข่าวที่เกี่ยวข้องกับความมั่นคงความปลอดภัยจากเว็บไซต์ที่รวบรวมข่าวจากนั้นนำไปวิเคราะห์ตัดสินเป็นระดับความอันตรายของข่าวในเหตุการณ์ และคาดการณ์ระยะเวลาที่เหตุการณ์ในข่าวจะสิ้นสุด ผลลัพธ์ที่ได้ระบบที่ออกแบบสามารถสกัดข้อมูลข่าวความปลอดภัยจากเว็บไซต์รวบรวมข่าว และมีความแม่นยำในการสกัดคำที่เกี่ยวกับข่าวความมั่นคงเท่ากับ 82.84% รวมทั้งมีความคลาดเคลื่อนในการคาดเดาระดับความอันตรายของข่าวในเหตุการณ์เท่ากับ 0.35

จักรินทร์ สันติรัตนภักดี [22] ได้ทำงานวิจัยการออกแบบ และพัฒนาระบบสกัดข้อมูลกึ่งโครงสร้างตามกระบวนการอีแอลที และค้นคืนสารสนเทศด้วยการเสนอภาพข้อมูลกรณีศึกษารายงานอุบัติเหตุทางถนนรายใหญ่บนเว็บไซต์ มีจุดประสงค์เพื่อออกแบบ และพัฒนาการสกัดข้อมูลกึ่งมีโครงสร้างตามกระบวนการอีแอลทีค้นคืนสารสนเทศด้วยการนำเสนอภาพข้อมูลแบบปฏิสัมพันธ์กับผู้ใช้ และนำข้อมูลมาช่วยในการวางแผน ป้องกันอุบัติเหตุทางถนน ผู้วิจัยทำการออกแบบ และพัฒนาการสกัดข้อมูลด้วยการสร้างคลังข้อมูลคำศัพท์ร่วมกับการรู้จำเอนทิตี้มาสกัดข้อมูลอุบัติเหตุทางถนนที่เว็บไซต์ Thairsc เพื่อสร้างระบบค้นคืนสารสนเทศด้วยการนำเสนอภาพข้อมูลปฏิสัมพันธ์กับผู้ใช้ผ่านทางเว็บไซต์ วัดผลด้วยผู้เชี่ยวชาญ 5 คน จำนวน 77 ครั้งตามจำนวนจังหวัดในประเทศไทย ทำซ้ำจนครบ 3 ครั้งแล้วนำมาหาค่าเฉลี่ย พบว่าค่าความถูกต้อง ค่าความครบถ้วน และค่าประสิทธิภาพโดยรวม เท่ากับ 0.87, 0.85 และ 0.86 ตามลำดับ

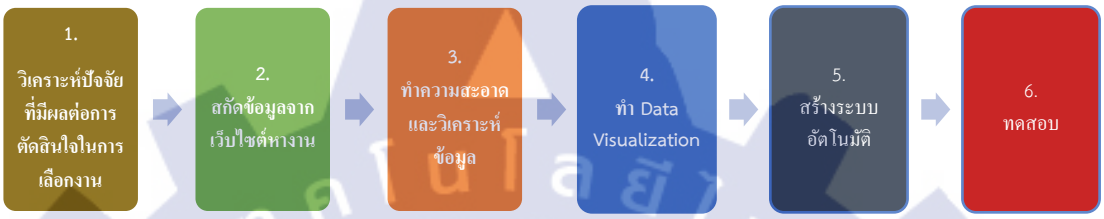
R. B. Mbah, M. Rege, and B. Misra [12] ได้ทำงานวิจัยค้นหาแนวโน้มของตลาดแรงงานด้วยการวิเคราะห์ข้อความ มีจุดประสงค์เพื่อนำเสนอวิธีการวิเคราะห์ข้อมูลตลาดแรงงานด้าน IT และติดตามข้อมูลตลาดแรงงานด้าน ผู้วิจัยใช้ภาษา JAVA เก็บข้อมูลบริษัทจากเว็บไซต์ Glassdoor โดยใช้ API ของเว็บไซต์ ถัดมาใช้ข้อมูลบริษัทมาค้นหาข้อมูลงานในเว็บไซต์ Indeed เก็บข้อมูลโดยใช้ API ได้ข้อมูลมาเป็น JSON จากนั้นทำความสะอาดข้อมูล สุดท้ายใช้ภาษา R ในการวิเคราะห์ข้อมูล และแสดงผลภาพ ระบบสามารถทำงานได้แสดงข้อมูลการวิเคราะห์โดยใช้กราฟแท่งในหัวข้อ Top 10 ตำแหน่งงานที่ต้องการ Top 10 Certificates ที่ต้องการ Top 10 บริษัทที่ต้องการแรงงาน IT มากที่สุด Top 10 อุตสาหกรรมที่ต้องการแรงงาน IT มากที่สุด และ Top 10 Technologies ที่ต้องการมากที่สุด ใช้ Word clouds ในการแสดงข้อมูลทักษะในแต่ละตำแหน่งงาน สุดท้ายในกราฟเส้นในการแสดงข้อมูลแนวโน้มของความต้องการในหัวข้อต่าง ๆ ในแต่ละเดือน งานวิจัยนี้ยังมีข้อจำกัดเป็นข้อมูลทักษะมีแต่ข้อมูลทักษะภาษาคอมพิวเตอร์ และเครื่องมือ Software

A. Phaphuangwittayakul และคณะ [23] ได้ทำงานวิจัยวิเคราะห์ความต้องการแรงงานของตลาดแรงงานในประเทศไทยจากเว็บไซต์หางาน มีจุดประสงค์เพื่อวิเคราะห์ และนำเสนอข้อมูลแนวโน้มความต้องการแรงงานในตลาดแรงงานของประเทศไทยผ่านการสำรวจข้อมูลจากเว็บไซต์หางาน ผู้วิจัยใช้ภาษา Python, BeautifulSoup, Scrapy และ Requests เป็นเครื่องมือในการทำ web scraping เก็บรวบรวมข้อมูลรับสมัครงานในเว็บไซต์หางาน Job Thai และเว็บไซต์หางาน jobsDB ลง Database MySQL และใช้อัลกอริทึม RAKE ในการดึงคำสำคัญเพื่อทำความสะอาดข้อมูล จากนั้นทำการวิเคราะห์ข้อมูล และแสดงผลภาพโดยใช้โปรแกรม Power BI ใช้ Tree map แสดงข้อมูลตำแหน่งงาน และใช้ Word clouds แสดงข้อมูลทักษะมีต้องการจากคุณสมบัติผู้สมัคร และรายละเอียดงาน ระบบสามารถใช้งานได้ช่วยให้ระบบการศึกษาเข้าใจทักษะที่ต้องการในตลาดแรงงาน แต่ระบบการดึงคำสำคัญยังดึงคำที่ไม่มีความหมายปะปนมาด้วย

D. M. Y. Lee และคณะ [24] ได้ทำงานวิจัยการวิเคราะห์สังคมเครือข่ายของตำแหน่งงานและทักษะ มีจุดประสงค์เพื่อวิเคราะห์ความสัมพันธ์ของตำแหน่งงาน ทักษะงาน และอุตสาหกรรมต่าง ๆ ในสิงคโปร์ ผู้วิจัยใช้ Octoparse เป็นเครื่องมือในการทำ web scraping เก็บรวบรวมข้อมูลรับสมัครงานในเว็บไซต์หางาน Indeed และเว็บไซต์หางาน jobsDB ถัดมาทำความสะอาดข้อมูลด้วยมนุษย์ สุดท้ายทำการวิเคราะห์สังคมเครือข่ายจากข้อมูลที่เก็บรวบรวมมา และแสดงผลภาพความสัมพันธ์

บทที่ 3
วิธีการดำเนินงาน

จากการศึกษา หลักการ ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อเป็นความรู้ในการสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทย และแสดงผลด้วยภาพเพื่อใช้ประกอบการตัดสินใจในการเลือกงาน มีวิธีดำเนินงาน ดังต่อไปนี้



3.1 วิเคราะห์ปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงาน

ศึกษาปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานจากงานวิจัยต่าง ๆ ที่มีคนได้วิจัยปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานมาก่อนจากนั้นนำปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานในแต่ละงานวิจัยมาเปรียบเทียบกับเนื้อหารับสมัครงานในเว็บไซต์หางานที่เลือกไว้ทั้ง 4 เว็ปไซต์ เลือกเปรียบเทียบจาก 2 ลำดับเนื้อหารับสมัครงานที่พบในหน้าเว็บไซต์ โดยทั้ง 4 เว็บไซต์ได้เลือกจากการอ้างอิงเว็บไซต์ที่แนะนำเว็บไซต์หางาน มีดังนี้

- https://www.jobbkk.com/
- https://th.jobsdb.com/th/th
- https://www.jobthai.com
- https://jobs.blognone.com/search

หลังจากเปรียบเทียบว่ามีข้อมูลใดที่จัดกลุ่มเข้ากับหัวข้อปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานที่ได้เลือก จะตัดหัวข้อปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานที่ไม่มีข้อมูลเกี่ยวข้องออก สรุปหัวข้อปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานที่เลือกแล้วนำไปหา Key Word ที่จะใช้สกัดคำปัจจัยที่มีผลต่อการตัดสินใจในการเลือกงานที่พบทั้ง 4 เว็บไซต์ คือ ตำแหน่งงาน เงินเดือน ที่อยู่องค์กร รายละเอียดงาน ระดับการศึกษา ประสบการณ์การทำงานและชื่อองค์กร แต่จากการวิเคราะห์ปัจจัยด้านรายละเอียดงานมีผลต่อการตัดสินใจมากเนื่องมาจากเกี่ยวข้องกับทักษะที่ต้องใช้ทำงานถ้าขาดทักษะที่เกี่ยวข้องจะไม่มีโอกาศได้รับเลือกเข้าทำงานจึงเน้นที่ปัจจัยด้านรายละเอียดงาน ซึ่งเป็นที่มาของวัตถุประสงค์ของการวิจัยจะหาแนวโน้มของตำแหน่งงาน และภาษาโปรแกรมคอมพิวเตอร์ที่มีความต้องการในประเทศไทย

ตารางที่ 3.1 เปรียบเทียบปัจจัยที่มีผลต่อการเลือกงานกับเนื้อหารับสมัครงานในเว็บไซต์หางาน

ปัจจัยในการเลือกงานที่พบในงานวิจัยที่ผู้วิจัยทำการศึกษา	**เว็บไซต์หางาน**				**ความถี่**
	jobsblognone	jobsdb	jobbkk	jobthai	
เงินเดือน	√	√	√	√	4
ระดับการศึกษา	√	√	√	√	4
ประสบการณ์การทำงาน	√	√	√	√	4
เพศ			√		1
อายุ			√	√	2
สถานภาพสมรส					0
ลักษณะหน้าที่ที่ได้รับของงาน	√	√	√	√	4
ความสำเร็จในการทำงาน					0
ความก้าวหน้าในอาชีพ					0
การยกย่องเชิดชู					0
นโยบายการทำงานของบริษัท					0
ความมั่นคงในอาชีพการทำงาน					0
สภาพในการทำงาน					0
โอกาสเพิ่มประสบการณ์พัฒนาตน					0
ความสัมพันธ์กับเพื่อนร่วมงาน					0
ผลตอบแทนและสวัสดิการ			√	√	2
ชื่อเสียงองค์กร	√	√	√	√	4
สถานที่ปฏิบัติงาน	√	√	√	√	4
ค่านิยม					0
การตระหนักในปัญหา					0
การวิเคราะห์สาเหตุของปัญหา					0
การแสวงหาทางเลือก					0
การประเมินและตัดสินใจเลือก					0
ติดตามการดำเนินงาน					0
ด้านภูมิหลังครอบครัว					0
ด้านความรู้สึก					0
ด้านทัศนคติ					0

ตารางที่ 3.2 ลำดับปัจจัยที่มีผลต่อการเลือกงานที่มีในเนื้อหารับสมัครงานในเว็บไซต์หางาน

ลำดับ	ปัจจัยในการเลือกงาน	ความถี่
1	เงินเดือน	4
2	ลักษณะหน้าที่ที่ได้รับของงาน	4
3	ชื่อเสียงองค์กร	4
4	สถานที่ปฏิบัติงาน	4
5	ระดับการศึกษา	4
6	ประสบการณ์การทำงาน	4
7	อายุ	2
8	ผลตอบแทนและสวัสดิการ	2
9	เพศ	1

3.2 สกัดข้อมูลจากเว็บไซต์หางาน

ในขั้นตอนที่ 2 เป็นการจัดการกับแหล่งข้อมูลที่นำมาใช้ ใช้เครื่องมือคือ Visual Studio Code, Python 3.10.5 และ Libary pandas 1.5.2 ซึ่งข้อมูลได้นำมาจากเว็บไซต์หางาน 4 เว็บ ดังนี้

3.2.1 เว็บไซต์ jobthai

ในการสกัดข้อมูลที่ต้องการนำมาใช้งานจากเว็บไซต์หางานอย่างเว็บไซต์ jobthai มีขั้นตอนดังนี้

1. ทำการศึกษาโครงสร้างของเว็บไซต์หางาน https://www.jobthai.com ว่าต้องเข้าใช้อย่างไรจึงจะเข้าถึงหน้าตำแหน่งงาน และรายละเอียดงาน เริ่มจากหน้าHomeของเว็บไซต์จะมีตัวเลือกให้เลือกประเภทงานเมื่อเลือก งานคอมพิวเตอร์/ไอที จะเข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาซึ่งมีหลายหน้าสามารถเลือกหน้าถัดไปหรือหน้าสุดท้ายที่แถบ Pagination และแต่ละหน้ามีตำแหน่งงาน 20 รายการใน 1 หน้า เมื่อกดที่ตำแหน่งงานจะเข้าสู่หน้ารายละเอียดของบริษัทรับสมัคร เมื่อกดที่ ดูรายละเอียดงาน จะเข้าสู่หน้ารายละเอียดงานเป็นหน้าที่มีข้อมูลที่ต้องการ เส้นทางการเข้าถึงรายละเอียดงานทั้งหมดเป็นไปตาม รูปที่ 3.1

2. หลังจากศึกษาโครงสร้างของเว็บไซต์แล้ว ให้ออกแบบขั้นตอนการเข้าถึงหน้ารายละเอียดงาน โดยเริ่มจากการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรง ต่อมาต้องเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เพื่อจะได้ URL เข้าหน้าตำแหน่งงานทุกตำแหน่งโดยข้ามหน้ารายละเอียดของบริษัทรับสมัครแล้วเข้าหน้ารายละเอียดงานเลย จึงทำการสกัดข้อมูลเก็บไว้ในไฟล์ Excel

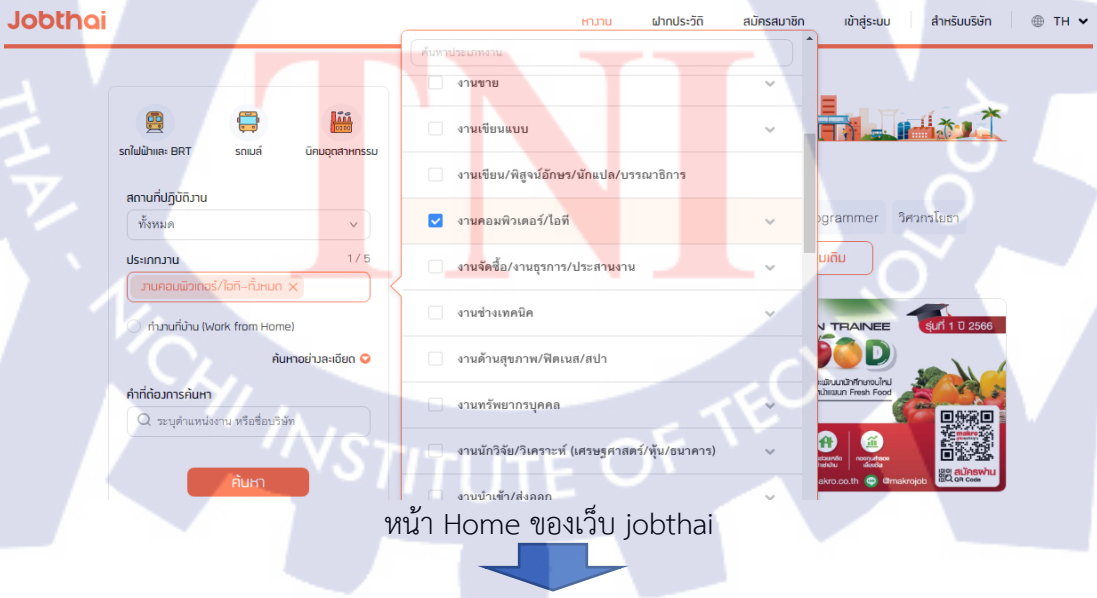
3. การทำงานในขั้นตอนการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหา โดยตรงทำได้โดยใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหาซึ่งคือ https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/ และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ

4. การทำงานในขั้นตอนเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เริ่มจากวิเคราะห์ลักษณะของ URL ที่ใช้เข้าแต่ละหน้า ซึ่งมีลักษณะดังนี้

หน้าที่ 1 => https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/1

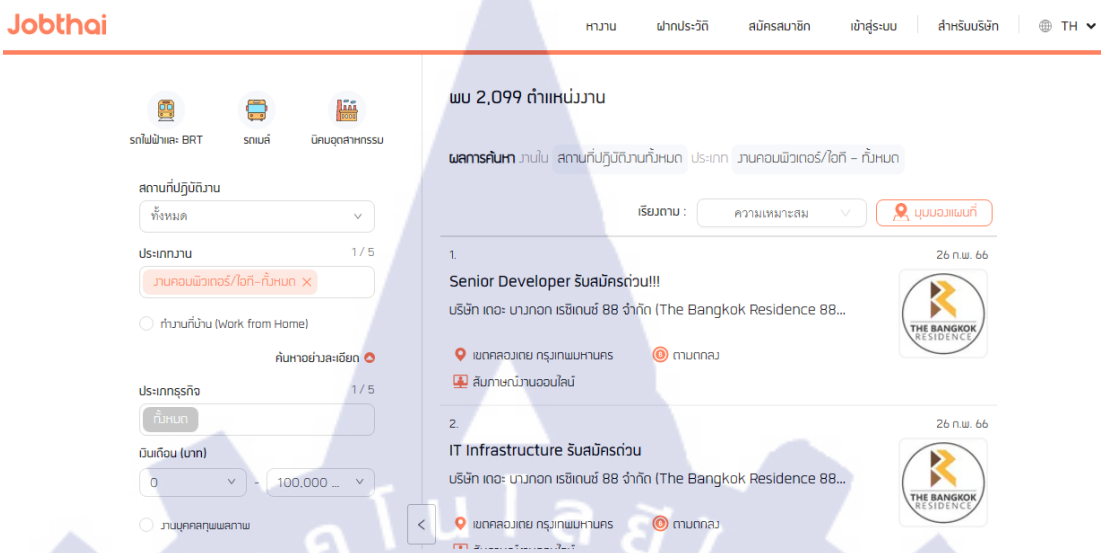
หน้าที่ 2 => https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/2

จะสังเกตุได้ว่า URL จะแตกต่างแค่เลขหน้าท้าย URL จึงนำข้อมูลทั้งหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาที่ถูกสกัดมาสกัดเฉพาะในส่วนของแถบ Pagination เนื่องจากไม่มีเลขหน้าสุดท้ายจึงต้องดึง URL หน้าสุดท้ายจากปุ่มกดไปหน้าสุดท้ายแล้วค่อยเก็บค่าเฉพาะเลขหน้าสุดท้ายเพื่อจะไปใช้สร้างเลขหน้าทุกหน้าจากนั้นจึงสร้างค่าเลขหน้าทั้งหมดตั้งแต่เลข 1 จนถึงเลขหน้าสุดท้ายโดยใช้ for loop แล้วนำค่าไปวางไว้เป็น list ต่อมาให้สร้าง URL สำหรับเข้าหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาทุกหน้าโดยใช้ for loop นำค่าที่เก็บไว้เป็น list มาต่อท้าย URL => https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/ เช่น นำค่าเลข 10 ที่เก็บเป็น list มาต่อจะได้ URL=> https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/10 เป็นต้น สุดท้ายใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหา และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา

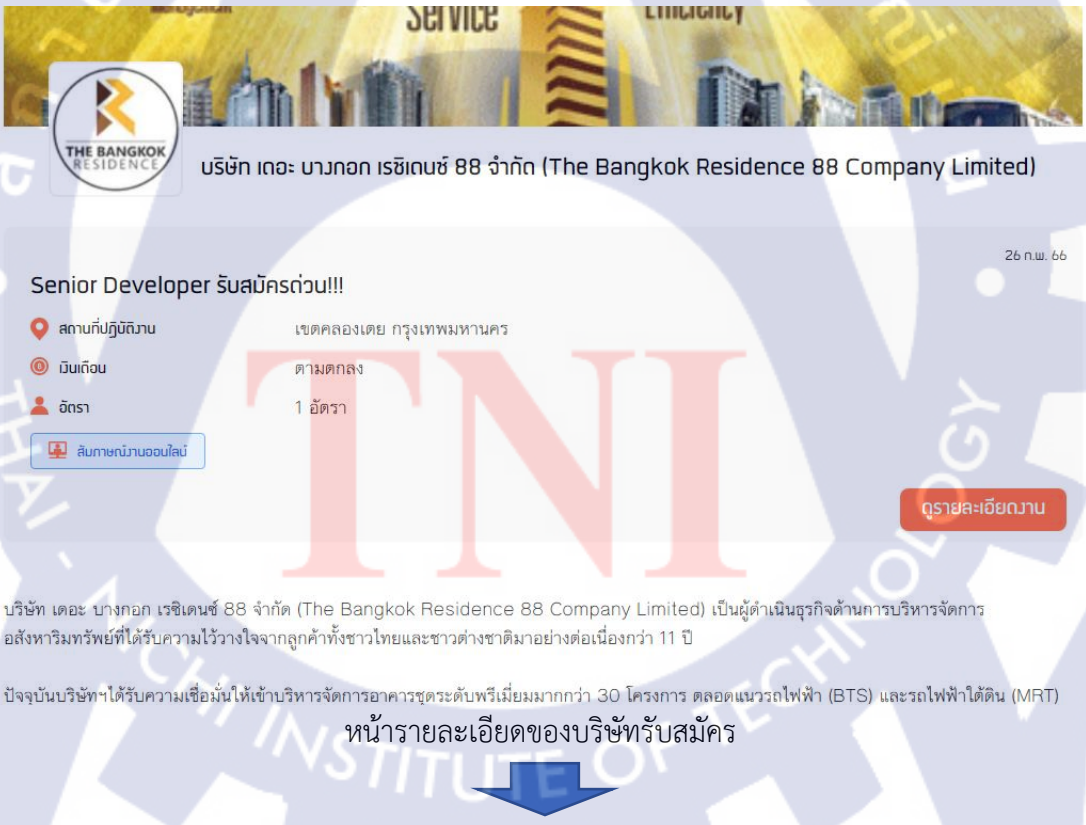


หน้า Home ของเว็บ jobthai

รูปที่ 3.1 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobthai

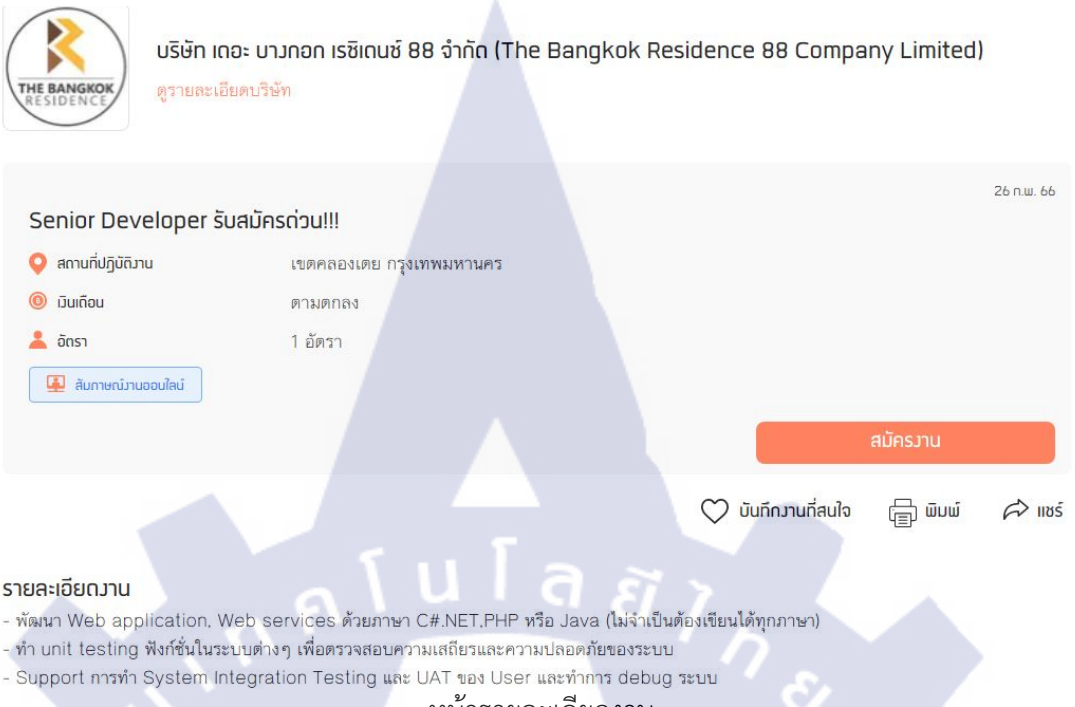


หน้ารายการตำแหน่งงานหลังค้นหา



หน้ารายละเอียดของบริษัทรับสมัคร

รูปที่ 3.1 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobthai (ต่อ)



หน้ารายละเอียดงาน

รูปที่ 3.1 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobthai (ต่อ)

5. การทำงานในขั้นตอนหา URL เข้าหน้ารายละเอียดงานโดยข้ามหน้ารายละเอียดของบริษัทรับสมัคร เริ่มจากวิเคราะห์ URL ของหน้ารายละเอียดของบริษัทรับสมัครกับหน้ารายละเอียดงาน ดังนี้

หน้ารายละเอียดของบริษัทรับสมัคร URL=>https://www.jobthai.com/th/company/job/930141

หน้ารายละเอียดงาน URL => https://www.jobthai.com/th/job/930141

ความแตกต่างของ URL ของหน้ารายละเอียดของบริษัทรับสมัครกับหน้ารายละเอียดงาน คือการที่มี company/ แทรกอยู่ใน URL จึงใช้ for loop ทำซ้ำการใช้ Library BeautifulSoup สกัด URL ของหน้ารายละเอียดของบริษัทรับสมัครแล้วดึงเลข ID Job ที่ท้าย URL ของหน้ารายละเอียดของบริษัทรับสมัคร มาต่อกับ URL => https://www.jobthai.com/th/job/ จะได้ URL ของหน้ารายละเอียดงาน

6. การทำงานในขั้นตอนเข้าหน้ารายละเอียดงาน และทำการสกัดข้อมูล เริ่มจากใช้ for loop ทำการวนซ้ำคำสั่งโดยแต่ละ loop จะใช้ Library requests เข้า URL ของหน้ารายละเอียดงาน และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ จากนั้นสกัดข้อมูลในหน้าเว็บแยกตามหัวข้อต่าง ๆ 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary,

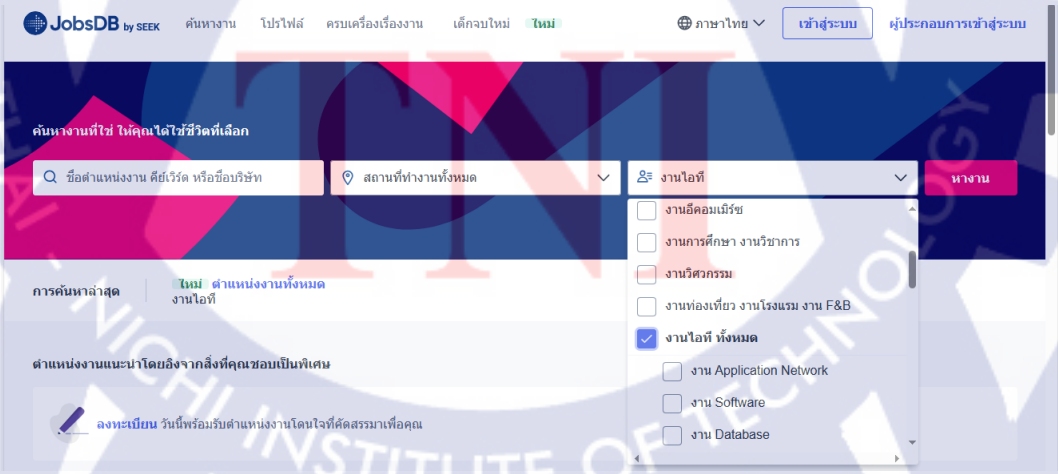
Job Description และ Candidate Qualifications โดยใช้คำสั่ง try และ except ในการข้ามขั้นตอนกรณีไม่มีข้อมูลทำซ้ำจนครบทุก URL ที่ได้มา

7. การทำงานในขั้นตอนการเก็บข้อมูลไว้ในไฟล์ Excel จะใช้ Library pandas สร้าง dataframe โดยมี 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications ทำการหาค่าวันที่ที่ทำการสกัดข้อมูลด้วย Library datetime จากนั้นเพิ่มคอลัมน์ date และใส่ค่าวันที่ที่ทำการสกัดข้อมูลที่หามาสุดท้ายทำการ save dataframe เป็นไฟล์ Excel ด้วย Library pandas

3.2.2 เว็บไซต์ jobsDB

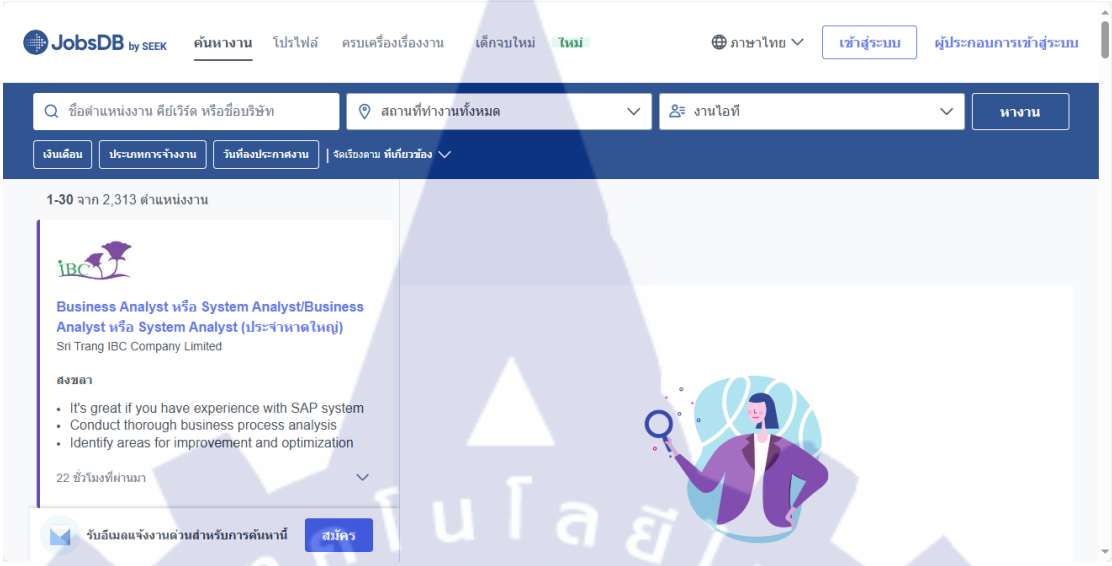
ในการสกัดข้อมูลที่ต้องการนำมาใช้งานจากเว็บไซต์หางานอย่างเว็บไซต์ jobsDB มีขั้นตอนดังนี้

1. ทำการศึกษาโครงสร้างของเว็บไซต์หางาน https://th.jobsdb.com/th/th ว่าต้องเข้าใช้อย่างไรจึงจะเข้าถึงหน้าตำแหน่งงาน และรายละเอียดงาน เริ่มจากหน้าHomeของเว็บไซต์จะมีตัวเลือกให้เลือกประเภทงานเมื่อเลือก งานคอมพิวเตอร์/ไอที จะเข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาซึ่งมีหลายหน้าสามารถเลือกหน้าถัดไปหรือหน้าสุดท้ายที่แถบ Pagination และแต่ละหน้ามีตำแหน่งงาน 30 รายการใน 1 หน้า เมื่อกดที่ตำแหน่งงานหน้าต่างด้านขวาจะแสดงรายละเอียดงาน หรือ ต้องคลิ๊กขวาจึงจะเปิดหน้าต่างใหม่เข้าสู่หน้ารายละเอียดงานเป็นหน้าที่มีข้อมูลที่ต้องการ เส้นทางการเข้าถึงรายละเอียดงานทั้งหมดเป็นไปตามรูปที่ 3.2



หน้า Home ของเว็บ jobsDB

รูปที่ 3.2 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobsDB



หน้ารายการตำแหน่งงานหลังค้นหา

หน้ารายละเอียดงาน

รูปที่ 3.2 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobsDB (ต่อ)

2. หลังจากศึกษาโครงสร้างของเว็บไซต์แล้วให้ออกแบบขั้นตอนการเข้าถึงหน้ารายละเอียดงาน โดยเริ่มจากการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรง ต่อมาต้องเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เพื่อจะได้ URL เข้าหน้ารายละเอียดงานของทุกตำแหน่งจึงทำการสกัดข้อมูลเก็บไว้ในไฟล์ Excel

3. การทำงานในขั้นตอนการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรงทำได้โดยใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหาซึ่งคือ https://th.jobsdb.com/th/th/jobs/งานไอที/ และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ

4. การทำงานในขั้นตอนเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เริ่มจากวิเคราะห์ลักษณะของ URL ที่ใช้เข้าแต่ละหน้า ซึ่งมีลักษณะดังนี้

หน้าที่ 1 => https://th.jobsdb.com/th/th/jobs/งานไอที/1

หน้าที่ 2 => https://th.jobsdb.com/th/th/jobs/งานไอที/2

จะสังเกตุได้ว่า URL จะแตกต่างแค่เลขหน้าท้าย URL จึงนำข้อมูลทั้งหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาที่สกัดมาดึงเลขหน้าสุดท้ายในส่วนของแถบ Pagination เพื่อจะไปใช้สร้างเลขหน้าทุกหน้าโดยใช้ Library BeautifulSoup กรณีนี้เป็น drop down ซึ่งมีเลขหน้าให้ดึงออกมาได้เลยจากนั้นจึงสร้างค่าเลขหน้าทั้งหมดตั้งแต่เลข 1 จนถึงเลขหน้าสุดท้ายโดยใช้ for loop แล้วนำค่าไปวางไว้เป็น list ต่อมาให้สร้าง URL สำหรับเข้าหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาทุกหน้าโดยใช้ for loop นำค่าที่เก็บไว้เป็น list มาต่อท้าย URL => https://th.jobsdb.com/th/th/jobs/งานไอที/ เช่น นำค่าเลข 12 ที่เก็บเป็น list มาต่อจะได้ URL=> https://th.jobsdb.com/th/th/jobs/งานไอที/12 เป็นต้น สุดท้ายใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหา และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา

5. การทำงานในขั้นตอนหา URL เข้าหน้ารายละเอียดงาน เริ่มจากวิเคราะห์ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา ดังนี้

หน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา URL=>/th/th/job/developer-power-bi-300003002984929?token=0~48092626-333b-4a81-ba6c-23b13466ba9a&sectionRank=6&jobId=jobsdb-th-job-300003002984929

หน้ารายละเอียดงาน URL => https://th.jobsdb.com/th/th/job/developer-power-bi-300003002984929?token=0~48092626-333b-4a81-ba6c-23b13466ba9a&sectionRank=6&jobId=jobsdb-th-job-300003002984929

ความแตกต่างของ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา กับ หน้ารายละเอียดงาน คือการที่มีแต่ค่าบอกตำแหน่งเข้าหน้ารายละเอียดงาน จึงใช้ Library BeautifulSoup และ for loop ทำซ้ำสกัดค่าบอกตำแหน่งเข้าหน้ารายละเอียดงานมาต่อกับ URL => https://th.jobsdb.com จะได้ URL ของหน้ารายละเอียดงาน

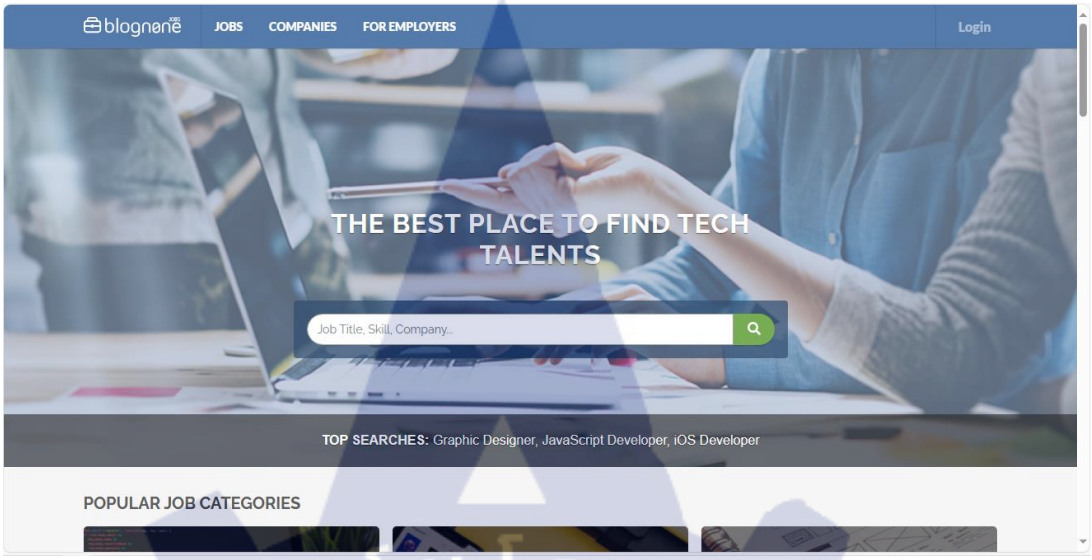
6. การทำงานในขั้นตอนเข้าหน้ารายละเอียดงาน และทำการสกัดข้อมูล เริ่มจากใช้ for loop ทำการวนซ้ำคำสั่งโดยแต่ละ loop จะใช้ Library requests เข้า URL ของหน้ารายละเอียดงาน และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บจากนั้นสกัดข้อมูลในหน้าเว็บแยกตามหัวข้อต่าง ๆ 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications โดยใช้คำสั่ง try และ except ในการข้ามขั้นตอนกรณีไม่มีข้อมูล ทำซ้ำจนครบทุก URL ที่ได้มา

7. การทำงานในขั้นตอนการเก็บข้อมูลไว้ในไฟล์ Excel จะใช้ Library pandas สร้าง dataframe โดยมี 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications ทำการหาค่าวันที่ที่ทำการสกัดข้อมูลด้วย Library datetime จากนั้นเพิ่มคอลัมน์ date และใส่ค่าวันที่ที่ทำการสกัดข้อมูลที่หามา สุดท้ายทำการ save dataframe เป็นไฟล์ Excel ด้วย Library pandas

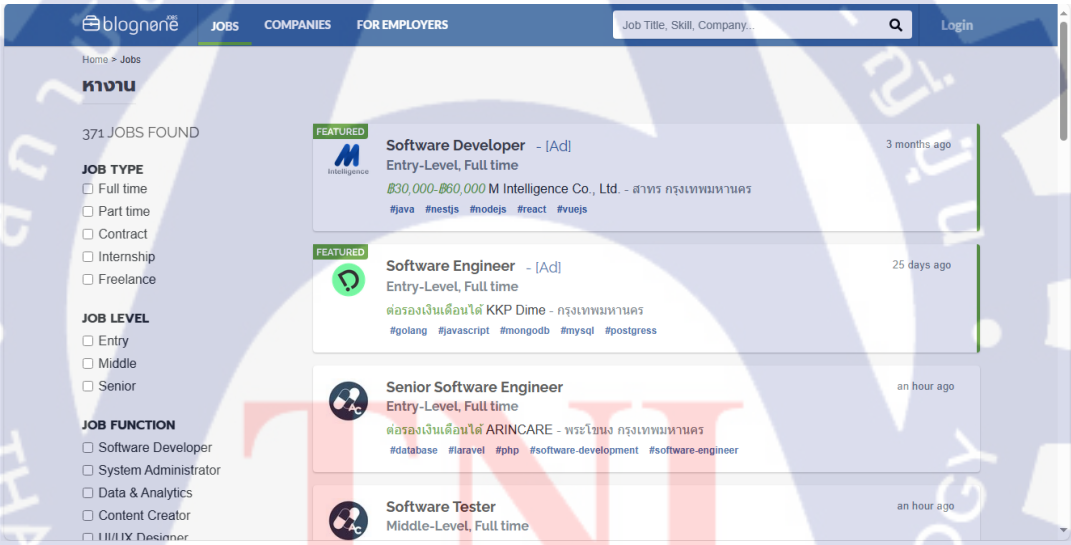
<u>3.2.3 เว็บไซต์ jobsblognone</u>

การสกัดข้อมูลที่ต้องการนำมาใช้งานจากเว็บไซต์หางานอย่างเว็บไซต์ jobsblognone มีขั้นตอนดังนี้

1. ทำการศึกษาโครงสร้างของเว็บไซต์หางาน https://jobs.blognone.com/ ว่าต้องเข้าใช้อย่างไรจึงจะเข้าถึงหน้าตำแหน่งงาน และรายละเอียดงาน เริ่มจากหน้าHomeของเว็บไซต์จะมีตัวเลือกให้เลือกดูตำแหน่งงานหรือดูบริษัทที่รับสมัครงานเมื่อเลือกตำแหน่งงาน จะเข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาซึ่งมีหลายหน้าสามารถเลือกหน้าถัดไปได้อย่างเดียวที่แถบ Pagination และแต่ละหน้ามีตำแหน่งงาน 13 รายการใน 1 หน้า เมื่อกดที่ตำแหน่งงานจะเข้าสู่หน้ารายละเอียดงานเป็นหน้าที่มีข้อมูลที่ต้องการ เส้นทางการเข้าถึงรายละเอียดงานทั้งหมดเป็นไปตามรูปที่ 3.3

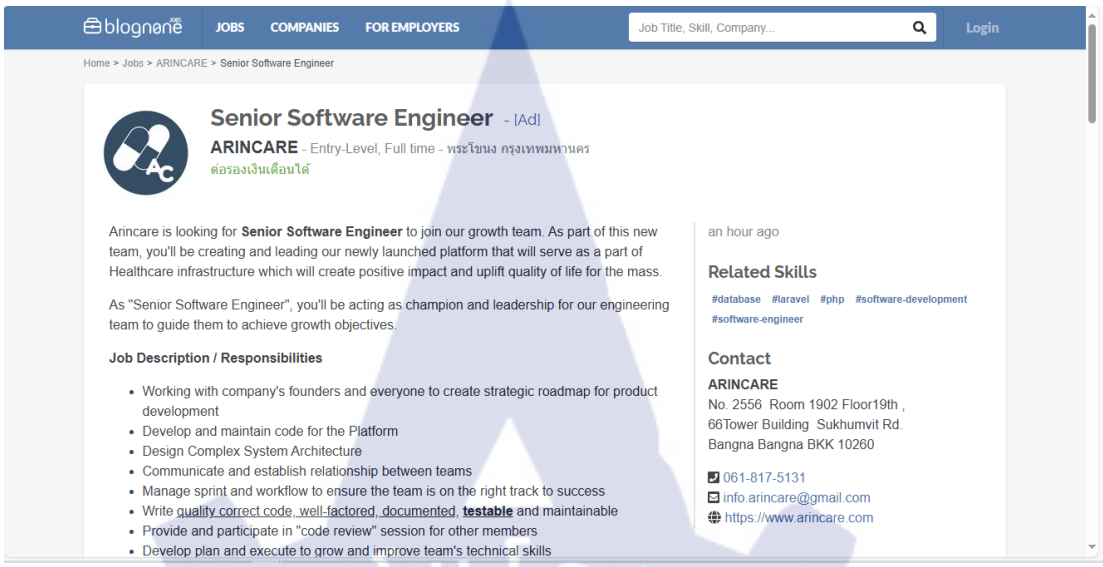


หน้า Home ของเว็บ jobsblognone



หน้ารายการตำแหน่งงานหลังค้นหา

รูปที่ 3.3 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobsblognone



หน้ารายละเอียดงาน

รูปที่ 3.3 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobsblognone (ต่อ)

2. หลังจากศึกษาโครงสร้างของเว็บไซต์แล้ว ให้ออกแบบขั้นตอนการเข้าถึงหน้ารายละเอียดงาน โดยเริ่มจากการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรง ต่อมาต้องเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เพื่อจะได้ URL เข้าหน้ารายละเอียดงานของทุกตำแหน่ง จึงทำการสกัดข้อมูลเก็บไว้ในไฟล์ Excel

3. การทำงานในขั้นตอนการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรงทำได้โดยใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหาซึ่งคือ https://jobs.blognone.com/search?page= และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ

4. การทำงานในขั้นตอนเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหาเริ่มจากวิเคราะห์ลักษณะของ URL ที่ใช้เข้าแต่ละหน้าซึ่งมีลักษณะดังนี้

หน้าที่ 1 => https://jobs.blognone.com/search?page=1

หน้าที่ 2 => https://jobs.blognone.com/search?page=2

จะสังเกตุได้ว่า URL จะแตกต่างแค่เลขหน้าท้าย URL จึงนำข้อมูลทั้งหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาที่สกัดมาดึงเลขหน้าสุดท้ายในส่วนของแถบ Pagination แต่เว็บนี้ไม่มีเลขหน้าสุดท้ายในแถบ Pagination จึงต้องใช้ Library BeautifulSoup และ regex ดึงจำนวนตำแหน่งงานทั้งหมดที่เว็บแสดงแล้วนำมาคำนวนหาเลขหน้าสุดท้าย จำนวนตำแหน่งงานทั้งหมดหารด้วยจำนวนที่แสดงในหน้ารายการ โดยใช้ Library math ปัดเศษขึ้น เพื่อนำไปใช้สร้างเลขหน้าทุก

หน้าจึงมีเศษไม่ได้จากนั้นจึงสร้างค่าเลขหน้าทั้งหมดตั้งแต่เลข 1 จนถึงเลขหน้าสุดท้ายโดยใช้ for loop แล้วนำค่าไปวางไว้เป็น list ในขั้นตอนถัดไปให้สร้าง URL สำหรับเข้าหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาทุกหน้าโดยใช้ for loop นำค่าที่เก็บไว้เป็น list มาต่อท้าย URL => https://jobs.blognone.com/search?page= เช่น นำค่าเลข 2 ที่เก็บเป็น list มาต่อจะได้ URL=> https://jobs.blognone.com/search?page=2 เป็นต้น ในขั้นตอนสุดท้ายใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหา และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา

5. การทำงานในขั้นตอนหา URL เข้าหน้ารายละเอียดงาน เริ่มจากวิเคราะห์ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา ดังนี้

หน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา URL=> /company/arincare/job/development-team-lea-xLn4

หน้ารายละเอียดงาน URL => https://jobs.blognone.com/company/arincare/job/development-team-lea-xLn4

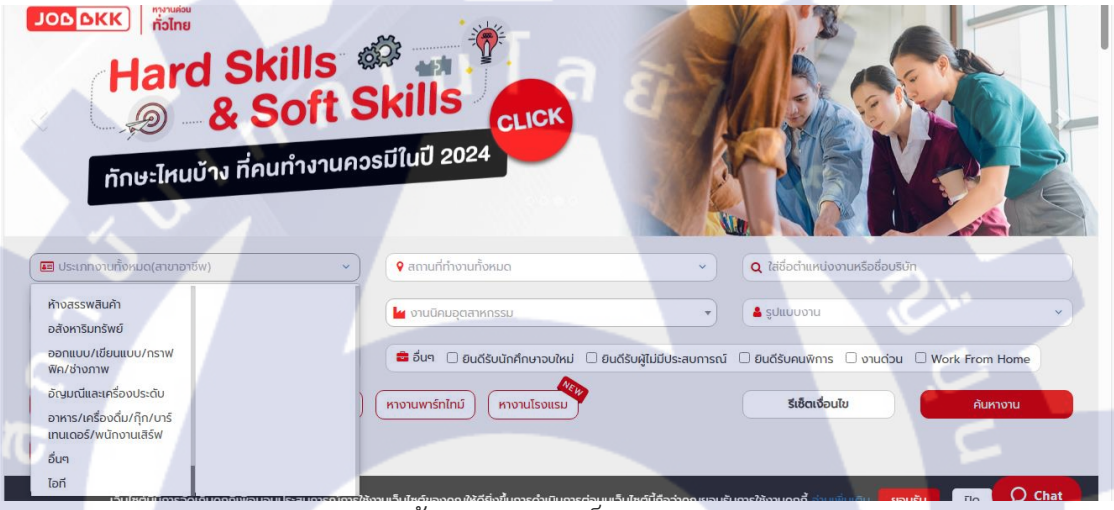
ความแตกต่างของ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหากับหน้ารายละเอียดงาน คือการที่มีแต่ค่าบอกตำแหน่งเข้าหน้ารายละเอียดงานจึงใช้ Library BeautifulSoup และ for loop ทำซ้ำสกัดค่าบอกตำแหน่งเข้าหน้ารายละเอียดงานมาต่อกับ URL => https://jobs.blognone.com จะได้ URL ของหน้ารายละเอียดงาน

6. การทำงานในขั้นตอนเข้าหน้ารายละเอียดงาน และทำการสกัดข้อมูล เริ่มจากใช้ for loop ทำการวนซ้ำคำสั่งโดยแต่ละ loop จะใช้ Library requests เข้า URL ของหน้ารายละเอียดงานและใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ จากนั้นสกัดข้อมูลในหน้าเว็บแยกตามหัวข้อต่าง ๆ 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications โดยใช้คำสั่ง try และ except ในการข้ามขั้นตอนกรณีไม่มีข้อมูลทำซ้ำจนครบทุก URL ที่ได้มา

7. การทำงานในขั้นตอนการเก็บข้อมูลไว้ในไฟล์ Excel จะใช้ Library pandas สร้าง dataframe โดยมี 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications ทำการหาค่าวันที่ที่ทำการสกัดข้อมูลด้วย Library datetime จากนั้นเพิ่มคอลัมน์ date และใส่ค่าวันที่ที่ทำการสกัดข้อมูลที่หามา สุดท้ายทำการ save dataframe เป็นไฟล์ Excel ด้วย Library pandas

3.2.4 เว็บไซต์ jobBBK

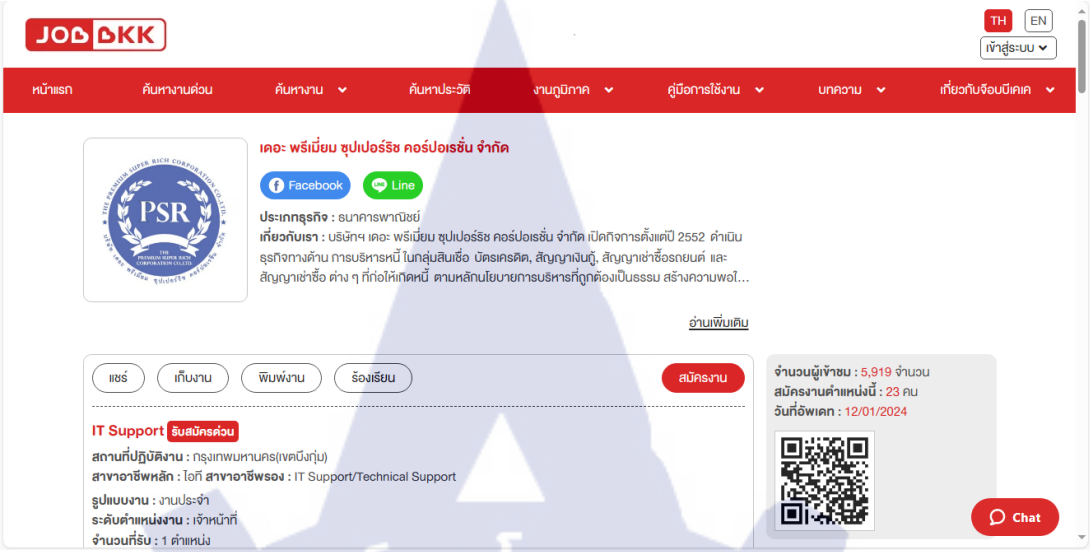
1. ทำการศึกษาโครงสร้างของเว็บไซต์หางาน https://www.jobbkk.com/หางาน/ไอที/ ว่าต้องเข้าใช้อย่างไรจึงจะเข้าถึงหน้าตำแหน่งงาน และรายละเอียดงาน เริ่มจากหน้าHomeของเว็บไซต์จะมีตัวเลือกให้เลือกดูตำแหน่งงานหรือดูบริษัทที่รับสมัครงานเมื่อเลือกตำแหน่งงาน จะเข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาซึ่งมีหลายหน้าสามารถเลือกหน้าถัดไปได้อย่างเดียวที่แถบ Pagination และแต่ละหน้ามีตำแหน่งงาน 25 รายการใน 1 หน้า เมื่อกดที่ตำแหน่งงานจะเข้าสู่หน้ารายละเอียดงานเป็นหน้าที่มีข้อมูลที่ต้องการ เส้นทางการเข้าถึงรายละเอียดงานทั้งหมดเป็นไปตามรูปที่ 3.4



หน้า Home ของเว็บ jobBBK

หน้ารายการตำแหน่งงานหลังค้นหา

รูปที่ 3.4 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobBBK



หน้ารายละเอียดงาน

รูปที่ 3.4 ตัวอย่างขั้นตอนการเข้าถึงข้อมูลรายละเอียดงานของเว็บ jobBBK (ต่อ)

2. หลังจากศึกษาโครงสร้างของเว็บไซต์แล้วให้ออกแบบขั้นตอนการเข้าถึงหน้ารายละเอียดงาน โดยเริ่มจากการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรง ต่อมาต้องเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เพื่อจะได้ URL เข้าหน้ารายละเอียดงานของทุกตำแหน่ง จึงทำการสกัดข้อมูลเก็บไว้ในไฟล์ Excel

3. การทำงานในขั้นตอนการข้ามหน้า Home เข้าสู่หน้ารายการตำแหน่งงานหลังค้นหาโดยตรงทำได้โดยใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหาซึ่งคือ https://www.jobbkk.com/หางาน/ไอที/ และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ

4.การทำงานในขั้นตอนเข้าทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา เริ่มจากวิเคราะห์ลักษณะของ URL ที่ใช้เข้าแต่ละหน้า ซึ่งมีลักษณะดังนี้

หน้าที่ 1 => https://www.jobbkk.com/หางาน/ไอที/1

หน้าที่ 2 => https://www.jobbkk.com/หางาน/ไอที/2

จะสังเกตุได้ว่า URL จะแตกต่างแค่เลขหน้าท้าย URL จึงนำข้อมูลทั้งหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาที่สกัดมาดึงเลขหน้าสุดท้ายในส่วนของแถบ Pagination แต่เว็บนี้ไม่มีเลขหน้าสุดท้ายในแถบ Pagination จึงต้องใช้ Library BeautifulSoup และ regex ดึงจำนวนตำแหน่งงานทั้งหมดที่เว็บแสดงแล้วนำมาคำนวนหาเลขหน้าสุดท้าย จำนวนตำแหน่งงานทั้งหมดหารด้วย 25 (ตำแหน่งงานที่มีใน1หน้า) โดยใช้ Library math ปัดเศษขึ้น เพื่อนำไปใช้สร้างเลขหน้าทุก

หน้าจึงมีเศษไม่ได้ จากนั้นจึงสร้างค่าเลขหน้าทั้งหมดตั้งแต่เลข 1 จนถึงเลขหน้าสุดท้ายโดยใช้ for loop แล้วนำค่าไปวางไว้เป็น list ต่อมาให้สร้าง URL สำหรับเข้าหน้าเว็บหน้ารายการตำแหน่งงานหลังค้นหาทุกหน้าโดยใช้ for loop นำค่าที่เก็บไว้เป็น list มาต่อท้าย URL => https://www.jobbkk.com/หางาน/ไอที/ เช่น นำค่าเลข 3 ที่เก็บเป็น list มาต่อจะได้ URL=> https://www.jobbkk.com/หางาน/ไอที/3 เป็นต้น สุดท้ายใช้ Library requests เข้า URL ของหน้ารายการตำแหน่งงานหลังค้นหา และใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บทุกหน้าของหน้ารายการตำแหน่งงานหลังค้นหา

5. การทำงานในขั้นตอนหา URL เข้าหน้ารายละเอียดงาน เริ่มจากวิเคราะห์ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา ดังนี้

หน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหา URL=> https://www.jobbkk.com/jobs/detail/19162/407563

หน้ารายละเอียดงาน URL => https://www.jobbkk.com/jobs/detail/19162/407563

ไม่มีความแตกต่างของ URL ของหน้ารายละเอียดงานในหน้ารายการตำแหน่งงานหลังค้นหากับหน้ารายละเอียดงาน จึงใช้ Library BeautifulSoup และ for loop ทำซ้ำสกัด URL ของหน้ารายละเอียดงานทั้งหมด

6. การทำงานในขั้นตอนเข้าหน้ารายละเอียดงาน และทำการสกัดข้อมูล เริ่มจากใช้ for loop ทำการวนซ้ำคำสั่งโดยแต่ละ loop จะใช้ Library requests เข้า URL ของหน้ารายละเอียดงานและใช้ Library BeautifulSoup ในการสกัดข้อมูลทั้งหน้าเว็บ จากนั้นสกัดข้อมูลในหน้าเว็บแยกตามหัวข้อต่าง ๆ 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications โดยใช้คำสั่ง try และ except ในการข้ามขั้นตอนกรณีไม่มีข้อมูล ทำซ้ำจนครบทุก URL ที่ได้มา

7. การทำงานในขั้นตอนการเก็บข้อมูลไว้ในไฟล์ Excel จะใช้ Library pandas สร้าง dataframe โดยมี 7 คอลัมน์คือ url, Company Name, Job Position, Company Address, Salary, Job Description และ Candidate Qualifications

ทำการหาค่าวันที่ที่ทำการสกัดข้อมูลด้วย Library datetime จากนั้นเพิ่มคอลัมน์ date และใส่ค่าวันที่ที่ทำการสกัดข้อมูลที่หามา สุดท้ายทำการ save dataframe เป็นไฟล์ Excel ด้วย Library pandas

ตารางที่ 3.3 แสดงรายละเอียดคอลัมน์ 8 อย่างที่ใช้เก็บข้อมูล

ชื่อคอลัมน์	รายละเอียดข้อมูลที่เก็บในแต่ละคอลัมน์
date	วันที่ที่ทำการสกัดข้อมูล
url	ลิ้งค์หน้ารายละเอียดงาน
Company Name	ชื่อบริษัท
Job Position	ชื่อตำแหน่งงาน
Company Address	ที่อยู่
Salary	เงินเดือน
Job Description	รายละเอียดงาน
Candidate Qualifications	คุณสมบัติผู้สมัคร

3.3 วิเคราะห์ และทำความสะอาดข้อมูล

ในขั้นตอนที่ 3 จะทำการวิเคราะห์ข้อมูลดิบจากขั้นตอนที่ 2 ก่อนจะนำไปทำความสะอาดข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการนำไปใช้งานต่อ

3.3.1 ทำการวิเคราะห์ข้อมูลดิบ

เครื่องมือที่นำมาใช้ คือ โปรแกรม Microsoft Excel และภาษา Python ใช้ Libary pandas ทำการวิเคราะห์ข้อมูลดิบจากขั้นตอนที่ 2 ที่สกัดข้อมูลจากเว็บไซต์หางานทั้ง 4 เว็บ โดยวัตถุประสงค์ของการวิจัยจะหาแนวโน้มของตำแหน่งงาน และภาษาโปรแกรมคอมพิวเตอร์ที่มีความต้องการในประเทศไทย จึงทำการวิเคราะห์ข้อมูลที่คอลัมน์ ชื่อตำแหน่งงาน รายละเอียดงาน และคุณสมบัติผู้สมัคร เป็นหลักซึ่งพบว่าข้อมูลในคอลัมน์ชื่อตำแหน่งงานส่วนใหญ่เป็นภาษาอังกฤษมีบางตำแหน่งเป็นภาษาไทยคำภาษาอังกฤษมีตัวพิมพ์เล็ก และพิมพ์ใหญ่ปนกัน รวมทั้งยังมีบางแถวมีชื่อตำแหน่งงานมากกว่า 1 ตำแหน่งในแถวเดียวกัน มีข้อมูลเป็นภาษาไทยที่ไม่เกี่ยวข้องอย่างเงินเดือน ที่อยู่ รายละเอียดงาน ข้อความรับสมัครด่วน อักขระพิเศษ และช่องว่าง รวมทั้งชื่อตำแหน่งงานเดียวกันแต่เขียนต่างกัน ส่วนข้อมูลในคอลัมน์รายละเอียดงาน และ คุณสมบัติผู้สมัคร แสดงข้อมูลรายละเอียดงาน และคุณสมบัติผู้สมัครงาน ส่วนใหญ่เป็นภาษาไทยปนกับภาษาอังกฤษคำภาษาอังกฤษมีตัวพิมพ์เล็ก และพิมพ์ใหญ่ปนกัน มีบางแถวเป็นภาษาอังกฤษทั้งหมดแต่ภาษาโปรแกรมคอมพิวเตอร์จะเป็นภาษาอังกฤษเท่านั้น ลักษณะ URL รายละเอียดงานของเว็บ jobsDB ให้เป็นตำแหน่งงานเดียวกันแต่ละวันที่เปิด URL จะไม่เหมือนกัน

ตารางที่ 3.4 ตัวอย่างลักษณะข้อมูลของคอลัมน์รายละเอียดงาน และ คุณสมบัติผู้สมัคร

ลักษณะข้อมูล	รูปแบบข้อมูลที่เก็บมาได้
มีข้อมูลภาษาไทยปนกับภาษาอังกฤษ แต่ภาษาโปรแกรมคอมพิวเตอร์จะเป็นภาษาอังกฤษ	1. พัฒนาแอพพลิเคชันทำงานบน Web Browser, Android หรือ iOS 2. พัฒนาส่วนติดต่อกับผู้ใช้ (User Interface) แบบ Hybridge ด้วย React หรือ Angular, HTML5/CSS, NodeJS 3. พัฒนาส่วนติดต่อกับผู้ใช้ (User Interface) แบบ Native ด้วย Android Java, Kotlin, Xcode, Swift หรือ Flutter 4. พัฒนาส่วน API ได้ด้วย Java / JAX-RS หรือ Express / NodeJS 5. พัฒนาส่วนจัดการข้อมูลด้วย SQL และคุ้นเคยกับ RDBMS เช่น MySQL, SQLite, Oracle หรือ SQLServer เป็นต้น
มีข้อมูลเป็นภาษาอังกฤษทั้งหมด	Back-end Developer: 1 year contract with renewal (hybrid work) Qualifications:• At least 3 years of experience in software development for backend programming languages in Java Spring Boot, J-Unit, JPA, SQL, Couchbase, NoSQL Database• Knowledge in Docker, Git & Bitbuckets.- Good command of English

ตารางที่ 3.5 ตัวอย่างลักษณะข้อมูลของคอลัมน์ชื่อตำแหน่งงาน

ลักษณะข้อมูล	รูปแบบข้อมูลที่เก็บมาได้
มีข้อมูลเป็นภาษาไทยที่ไม่เกี่ยวข้อง	System Analyst รับสมัครด่วน
ชื่อตำแหน่งงานเดียวกันแต่เขียนต่างกัน	Back-end Developer Developer Back-End - UTEL
ชื่อตำแหน่งงานมากกว่า 1 และเป็นภาษาไทย	Web programmer, โปรแกรมเมอร์ PHP
มีข้อมูลเป็นภาษาไทยที่ไม่เกี่ยวข้อง	System Analyst นนทบุรี ประชาชื่น 45K ยังไม่รวมสวัสดิการงานประจำ ติดต่อ HR เฟม

3.3.2 การทำความสะอาดข้อมูล

ถัดจากทำการวิเคราะห์ข้อมูลดิบจะทำความสะอาดข้อมูลโดยใช้ภาษา Python และใช้ Libary pandas มีขั้นตอนดังนี้

1. จากขั้นตอนสกัดข้อมูลจากเว็บไซต์หางานข้อมูลที่ได้จะเก็บในรูปแบบไฟล์ Excel 4 ไฟล์โดย 1 ไฟล์จะมีข้อมูลแค่ 1 เว็บไซต์จึงต้องรวมทั้ง 4 ไฟล์เข้าด้วยกันเป็น dataframe เดียวกัน

2. นำ dataframe ที่รวมทั้ง 4 ไฟล์เข้าด้วยกัน มาทำความสะอาดข้อมูลต่อ เนื่องจากลักษณะ URL รายละเอียดงานของเว็บ jobsDB ให้เป็นตำแหน่งงานเดียวกันแต่ละวันที่เปิด URL จะไม่เหมือนกัน จึงต้องนำส่วนที่ไม่ซ้ำกันออกก่อนซึ่งเป็นส่วนหลังคำว่า token เนื่องจากต้องใช้ URL ในการลบข้อมูลที่ซ้ำกันออก

3. ทำการลบค่าทั้งแถวเมื่อแถวนั้นไม่มีค่าในคอลัมน์ Job Position จากนั้นลบค่าทั้งแถวเมื่อมีค่าซ้ำกันในคอลัมน์ url, Company Name และ Job Position จากนั้นเปลี่ยนตัวพิมพ์ใหญ่ให้กลายเป็นพิมพ์เล็ก (Lowercasing) ในคอลัมน์ Job Position, Company Address, Job Description และ Candidate Qualifications ก่อนจะทำการแทนค่าว่าง ให้เป็น NaN

4. ทำการสร้าง dataframe2 เพื่อใช้เก็บค่าที่ต้องการใช้งาน

5. ทำการ tag ชื่อตำแหน่งงาน สถานที่ทำงาน ทักษะ และเครื่องมือที่ใช้ จาก dataframeอันแรกมาใส่ใน dataframe2 เพื่อจะบ่งบอกตำแหน่งงาน สถานที่ทำงาน ทักษะ และเครื่องมือที่ใช้ ในแถวข้อมูลรับสมัครงานได้ชัดเจนให้ข้อมูลที่มีความหมายเดียวกันแต่เขียนต่างกันให้เป็นรูปแบบเดียวกัน โดยใช้คำสั่ง contrains ซึ่งสามารถค้นหาข้อมูลบางส่วนโดยไม่ต้องเหมือนกันทั้งหมดในคอลัมน์นั้น ๆ จาก dataframe อันแรกแล้วทำการแทนค่าลงคอลัมน์ที่ต้องการใน dataframe2 อย่างการ tag ชื่อตำแหน่งงานจะใช้ key word ซึ่งเป็นส่วนหนึ่งในตำแหน่งงานมาคัดกรองตำแหน่งงานในคอลัมน์ Job Position เช่น iot เป็น key word ที่มีบางส่วนในข้อมูลตำแหน่งงานประเภท iot engineer เป็นต้น ถ้ามี key word มากกว่า 1 คำให้ใช้ & เพื่อเพิ่ม key word ที่ต้องเจอทั้งหมดในชื่อตำแหน่งงาน แล้วใส่ประเภทตำแหน่งงานที่ต้องการลงในคอลัมน์ tag_Jobname

การ tag ชื่อสถานที่ทำงาน ค่า tag ที่ต้องการคือชื่อจังหวัดจึงใช้ key word ที่เป็นคำส่วนหนึ่งของชื่อจังหวัด และ อำเภอของจังหวัดปริมณฑล โดยใช้คำสั่ง contrains และทำซ้ำด้วย for loop ให้ครบทุกจังหวัด ใส่ค่า tag ลงคอลัมน์ tag_add

การ tag ชื่อทักษะ และเครื่องมือที่ใช้ เนื่องจากในรายละเอียดรับสมัครงานจะมีทักษะและเครื่องมือที่ใช้มากกว่า 1 ทำให้ต้องออกแบบเป็นสร้างคอลัมน์แล้วใส่ค่า TRUE/FALSE แทนโดยให้ 0 เป็นค่า FALSE และ 1 เป็น TRUE ตัว key word เป็นชื่อ ภาษาโปรแกรม และ ชื่อเครื่องมือที่ใช้ ตัว key word จะเป็นทั้งตัวกรอง และชื่อคอลัมน์ ใช้คำสั่ง contrains และทำซ้ำด้วย for loop ให้ครบทุก key word

การจัดการข้อมูลเงินเดือน ข้อมูลเงินเดือนมักจะแสดงเป็น ตามที่ตกลง หรือ ไม่ใส่มา ถ้ามีบอกไว้มักจะเป็นช่วงเงินเดือน 15,000 – 20,000 บาท หรือ 15k-30k จึงต้องทำการแก้ไขให้อยู่ในรูปแบบ 15000 – 20000 ก่อนจะแยกเป็นเงินเดือนขั้นต่ำใส่ลงคอลัมน์ Salary min และเงินเดือนสูงสุดใส่ลงในคอลัมน์ Salary max ใน dataframe2

6. เครื่องมือที่จะใช้ทำ Data Visualization คือ power bi data model ที่เหมาะสมที่สุดคือ star schema จึงต้องออกแบบฐานข้อมูล ให้มีตารางข้อมูลเป็น Fact คือตารางหลักมี 1 ตาราง และ Dimension คือ ตารางอ้างอิงมีหลายตาราง dataframe2 คือตารางข้อมูล Fact ซึ่งต้องเปลี่ยนค่าจากการ tag ข้อมูลในคอลัมน์ tag_Jobname และ tag_add ทำการเปลี่ยนค่า tag ข้อมูลที่เป็นคำให้เป็นตัวเลขเพื่อลดความซ้ำซ้อนของข้อมูลทำให้ขนาดข้อมูลลดลง และ ทำให้ power bi อัพเดท และแสดงผลข้อมูลได้รวดเร็ว

เริ่มจากสร้างตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_Jobname เป็นไฟล์ Excel มี 3 คอลัมน์ ดังนี้

tag_Jobname มีค่าเป็นตัวเลข

tag_Jobname_name มีค่าเป็นประเภทตำแหน่งงานที่มีในคอลัมน์ tag_Jobname ของตาราง Fact

Jobname มีค่าเป็นประเภทตำแหน่งงานที่ใช้แสดงผลในแดชบอร์ด

ตารางที่ 3.6 ลักษณะตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_Jobname

tag_Jobname	tag_Jobname_name	Jobname
1	3d model	3D Model
2	account manager	Account Manager
3	application consultant	Application Consultant
...	...	...
87	web programmer	Web Programmer

ทำการอ่านไฟล์ Excel ตาราง Dimension สร้างเป็น dataframe จากนั้นใช้คำสั่ง contrains ค้นหาค่า tag ในคอลัมน์ tag_Jobname ของตาราง Fact ด้วย key word ในคอลัมน์ tag_Jobname_name แล้วแทนค่า tag_Jobname ของตาราง Fact ด้วยตัวเลขในคอลัมน์ tag_Jobname ของตาราง Dimension ใช้จาก และทำซ้ำด้วย for loop ให้ครบทุก key word

สร้างตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_add เป็นไฟล์ Excel มี 3 คอลัมน์ ดังนี้

tag_add มีค่าเป็นตัวเลข

tag_add_name มีค่าประเภทตำแหน่งงานที่มีในคอลัมน์ tag_add ของตาราง Fact

ทำการอ่านไฟล์ Excel สร้างเป็น dataframe จากนั้นใช้คำสั่ง contrains ค้นหาค่า tag ในคอลัมน์ tag_add ของตาราง Fact ด้วย key word ในคอลัมน์ tag_add_name แล้วแทนค่าเป็นตัวเลขในคอลัมน์ tag_add ของตาราง Dimension ใช้จาก และทำซ้ำด้วย for loop ให้ครบทุก key word

ตารางที่ 3.7 ลักษณะตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_add

tag_add	tag_add_name
1	นครราชสีมา
2	เชียงใหม่
3	กาญจนบุรี
...	...
77	สมุทรสงคราม

ทำการอ่านไฟล์ Excel ตาราง Dimension สร้างเป็น dataframe จากนั้นใช้คำสั่ง contrains ค้นหาค่า tag ในคอลัมน์ tag_add ของตาราง Fact ด้วย key word ในคอลัมน์ tag_add_name แล้วแทนค่า tag ในคอลัมน์ tag_add ของตาราง Fact ด้วยตัวเลขในคอลัมน์ tag_add ของตาราง Dimension และทำซ้ำด้วย for loop ให้ครบทุก key word

3.3.3 ระบบอัพเดทฐานข้อมูล

เครื่องมือที่จะใช้ทำ Data Visualization คือ power bi data model ที่เหมาะสมที่สุดคือ star schema จึงต้องออกแบบฐานข้อมูล ให้มีตารางข้อมูลเป็น Fact คือตารางหลักมี 1 ตาราง และ Dimension คือ ตารางอ้างอิงมีหลายตาราง แต่ลักษณะของข้อมูลที่มีไม่เหมาะสมกับ data model star schema เพราะจำเป็นต้องแยกข้อมูลทักษะ และเครื่องมือมากลับด้านข้อมูลคอลัมน์ทำให้มี Fact 2 ตารางจึงกลายเป็น data model snowflake จึงมีตารางฐานข้อมูลที่รวมข้อมูลทั้งหมด และตารางข้อมูล Fact 2 ตาราง มีขั้นตอนดังนี้

1. ระบบอัพเดทตารางฐานข้อมูลที่รวมข้อมูลทั้งหมดโดยใช้วิธีอ่านไฟล์ CSV ตารางฐานข้อมูลที่รวมข้อมูลทั้งหมดแล้วสร้างเป็น dataframe แล้วนำมารวมกับ dataframe ข้อมูลใหม่ที่ผ่านการทำความสะอาดข้อมูลมาแล้ว จากนั้นลบแถวข้อมูลที่ไม่มีข้อมูลในคอลัมน์ tag_Jobname เนื่องจากตำแหน่งงานไม่เกี่ยวข้องกับงานสายไอที แล้วค่อยลบแถวข้อมูลที่มีข้อมูลในคอลัมน์ url,

Company Name และ Job Position ซ้ำทิ้งทำการสร้าง index ใหม่ และให้ค่า index คือ ค่าในคอลัมน์ ID_Job แล้วเซฟไฟล์เป็น CSV ทับของเก่า

2. ทำการสร้างตารางข้อมูล Fact ทักษะ และเครื่องมือโดยทำการสร้าง dataframe ใหม่สำหรับทักษะ และเครื่องมือมีคอลัมน์ ID_Job เป็นตัวเชื่อมระหว่างตาราง Fact ทั้ง 2 ตาราง

ทำการกลับด้านคอลัมน์ด้วยคำสั่ง melt โดยให้ ID_Job เป็นคอลัมน์ที่ไม่กลับด้าน ให้คอลัมน์ทักษะ และเครื่องมือทั้งหมดกลับด้านโดยชื่อของคอลัมน์ที่เก็บค่าชื่อทักษะและเครื่องมือมีชื่อว่า tool และ คอลัมน์ที่เก็บค่า 1 หรือ 0 มีชื่อว่า value เปลี่ยนประเภทข้อมูลทั้ง dataframe ให้เป็นสตริงค์แล้วทำการคัดกรองแถวข้อมูลที่มีค่าข้อมูลในคอลัมน์ value เป็น 1 ซึ่งหมายถึงรายละเอียดรับสมัครงานมีทักษะ และเครื่องมือนั้น

เริ่มจากสร้างตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact ทักษะ และเครื่องมือ เป็นไฟล์ Excel มี 3 คอลัมน์ ดังนี้

tool มีค่าเป็นตัวเลข

tool_tag มีค่าเป็นประเภทตำแหน่งงานที่มีในคอลัมน์ tool ของตาราง Fact

tool_name มีค่าเป็นประเภทตำแหน่งงานที่ใช้แสดงผลในแดชบอร์ด

ตารางที่ 3.8 ลักษณะตาราง Dimension สำหรับอ้างอิงข้อมูลในตาราง Fact ทักษะ และเครื่องมือ

tool	tool_tag	tool_name
1	abap	Abap
2	.net	.Net
3	kotlin	Kotlin
...	...	...
190	elk	ELK

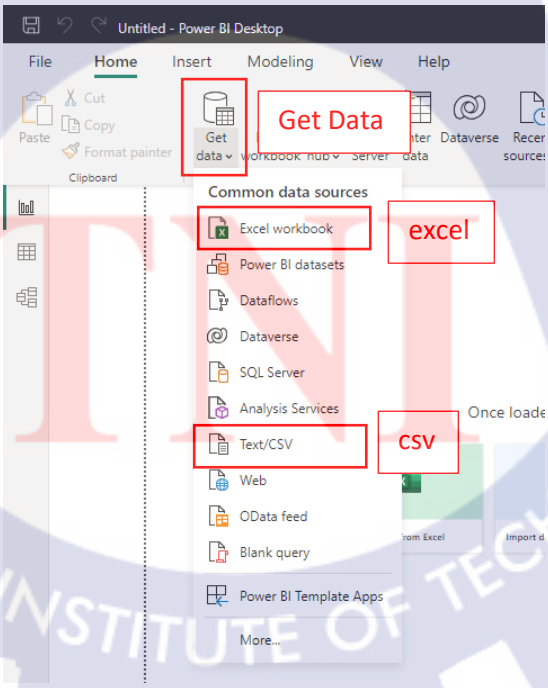
ทำการอ่านไฟล์ Excel ตาราง Dimension ทักษะ และเครื่องมือ สร้างเป็น dataframe จากนั้นใช้คำสั่ง contrains ค้นหาค่า tag ในคอลัมน์ tool ของตาราง Fact ทักษะ และเครื่องมือ ด้วย key word ในคอลัมน์ tool_tag ของตาราง Dimension ทักษะ และเครื่องมือ แล้วแทนค่าค่า tag ในคอลัมน์ tool ของตาราง Fact ด้วยตัวเลขในคอลัมน์ tool ของตาราง Dimension และทำซ้ำด้วย for loop ให้ครบทุก key word แล้วเซฟไฟล์เป็น CSV

3. ทำการสร้างตารางข้อมูล Fact ข้อมูลรับสมัครงานโดยใช้คำสั่งเซฟไฟล์เป็น CSV ที่มีเงื่อนไขเซฟเฉพาะคอลัมน์ date, ID_Job, url, Company Name, Job Position, tag_Jobname, tag_add, Salary min, Salary max

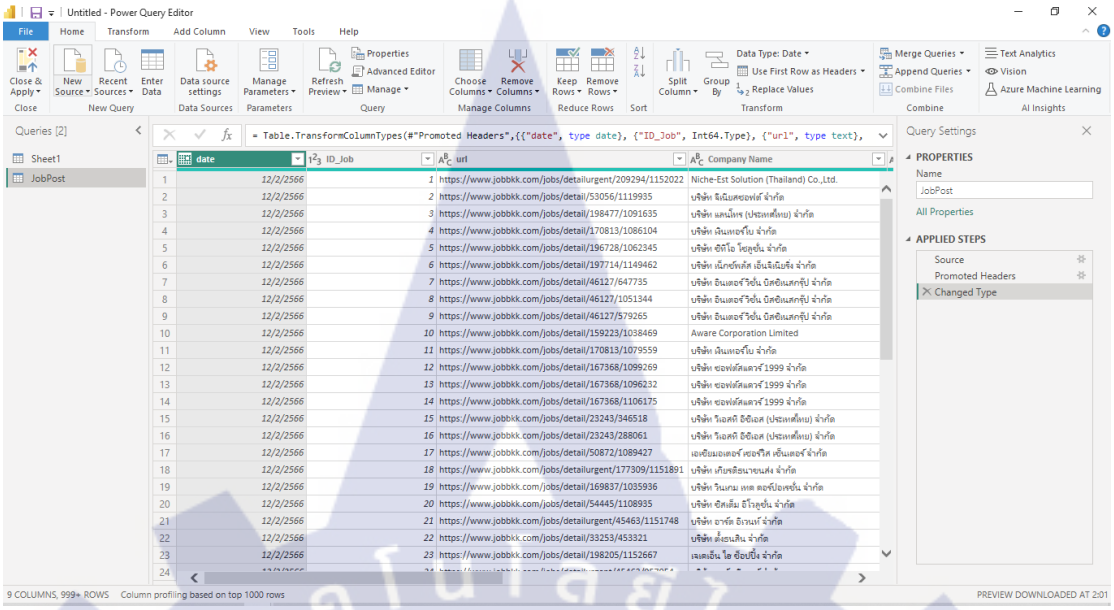
3.4.ทำ Data Visualization

ในขั้นตอนที่ 4 จะใช้โปรแกรม Microsoft Power BI ในการทำ Data Visualization ซึ่งโปรแกรม Microsoft Power BI เป็นโปรแกรมที่ได้รับความนิยมอันดับต้น ๆ ในการการวิเคราะห์ข้อมูลเชิงธุรกิจ (Business Intelligence) และการทำ Data Visualization

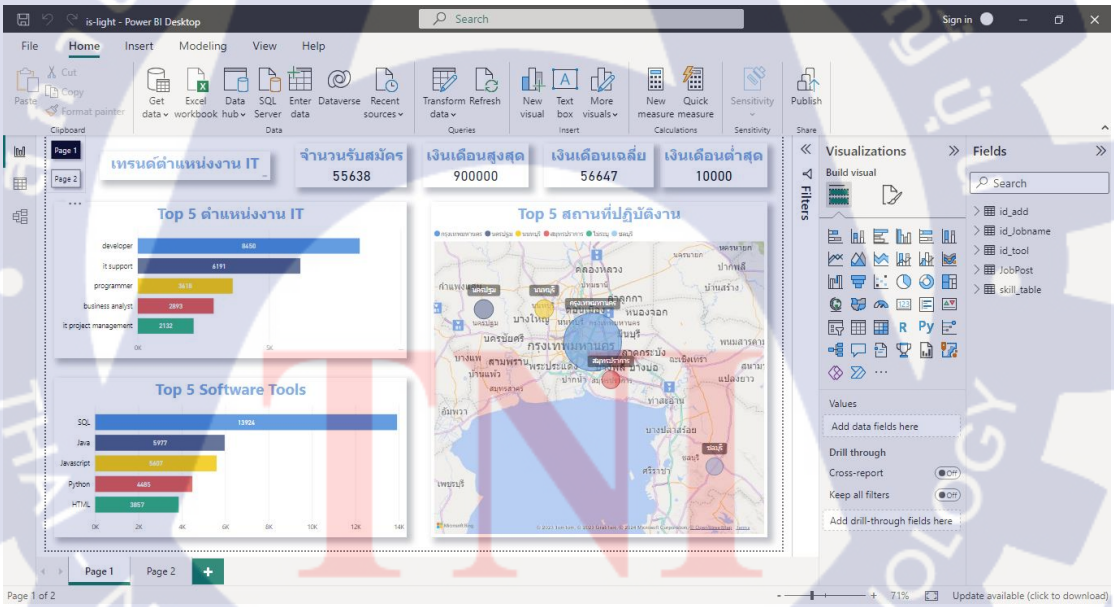
โดยชุดข้อมูลที่เริ่มต้นใช้มาจากการสกัดข้อมูลทั้ง 4 เว็บไซต์หางานในวันที่ 12/2/2566 มีข้อมูลรวมกันมากกว่า 6,000 รายการ และจะเพิ่มมากขึ้นในอนาคต เริ่มจากนำเข้าข้อมูลด้วยคำสั่ง Get Data > Excel คำสั่งนี้ทำให้นำเข้าข้อมูลไฟล์ Excel ต้องทำซ้ำเพื่อนำเข้าทุกไฟล์ตาราง Dimension จากนั้นใช้คำสั่ง Get Data > CSV คำสั่งนี้ทำให้นำเข้าข้อมูลไฟล์ CSV ต้องทำซ้ำเพื่อนำเข้าทุกไฟล์ตาราง Fact เข้าจากนั้นเข้า Power Query ทำการนำแถวข้อมูลแถวแรกเป็นชื่อคอลัมน์และเปลี่ยนประเภทข้อมูลทำการ Apply&Close แล้วทำ Data Visualization และแชร์ Data Visualization โดยคำสั่ง Publish



รูปที่ 3.5 นำเข้าข้อมูลด้วยคำสั่ง Get Data > Excel หรือ Get Data > CSV



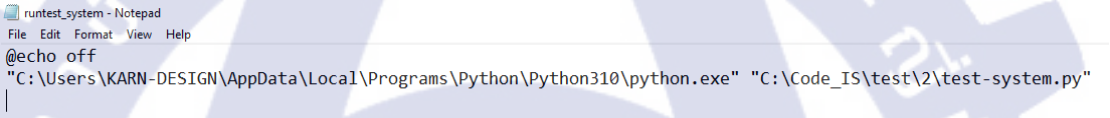
รูปที่ 3.6 หน้า Power Query ที่ใช้ปรับแต่งตารางข้อมูล



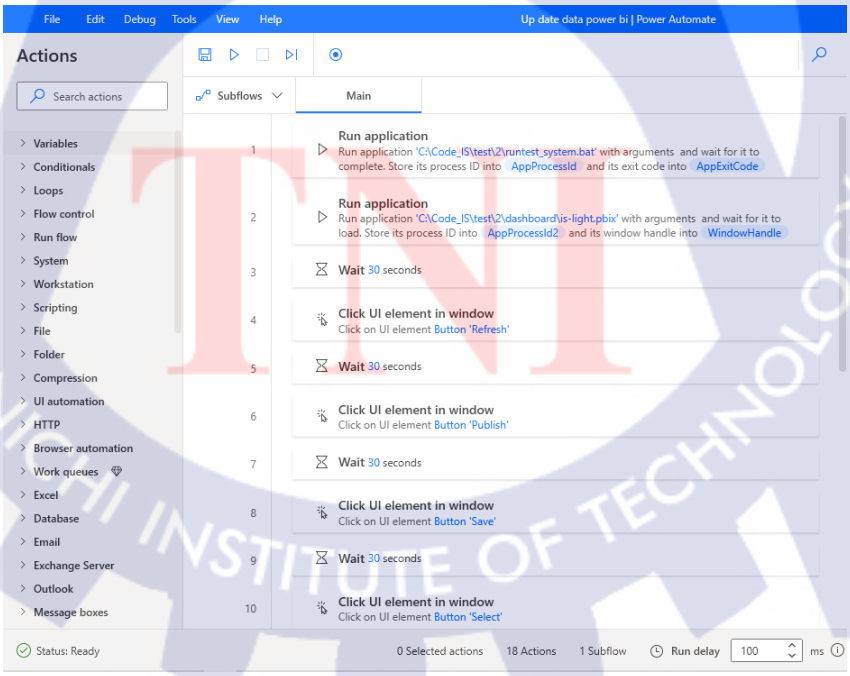
รูปที่ 3.7 Dashboard เทรนด์ตำแหน่งงาน IT

3.5.สร้างระบบให้ทำงานอัตโนมัติ

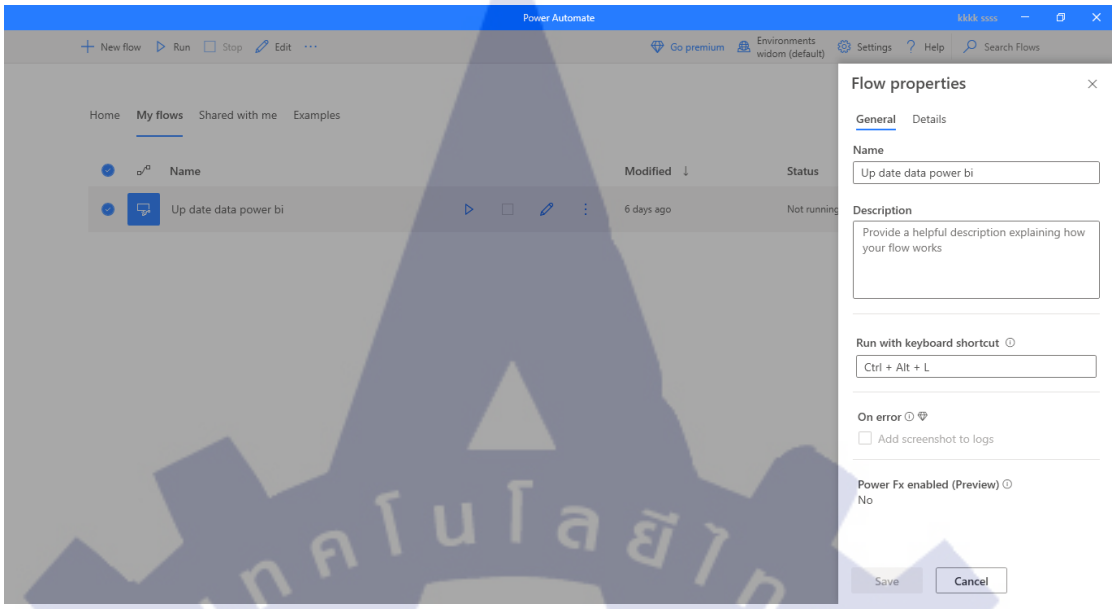
ในขั้นตอนที่ 5 เป็นการสร้างระบบให้ทำงานอัตโนมัติ เริ่มจากการสร้างไฟล์ .BAT สำหรับรัน Python Script ที่ใช้สกัดข้อมูลจากเว็บไซต์หางานทั้ง 4 เว็บ และทำความสะอาดข้อมูล ใช้ RPA อย่าง Power Automate สร้าง flow สั่งทำงานไฟล์ .BAT สำหรับรัน Python Script ที่ใช้สกัดข้อมูลจากเว็บไซต์หางานทั้ง 4 เว็บ และทำความสะอาดข้อมูลเมื่อทำงานเสร็จแล้วจะสั่งให้เปิดไฟล์ Dashboard เข้า Power BI Desktop รอ 30 วินาที สั่งกดปุ่ม Reflesh data รอ 30 วินาที สั่งกดปุ่ม Publish รอ 30 วินาที สั่งกดปุ่ม Save รอ 30 วินาที สั่งกดปุ่ม Select รอ 30 วินาที สั่งกดปุ่ม Replace รอ 30 วินาที สั่งกดปุ่ม Got it รอ 30 วินาที ปิด Power BI Desktop จึงจบการทำงาน จากนั้นตั้งค่าให้รัน flow ด้วย key ลัด Ctrl + Alt + L แล้วสร้างไฟล์ powershell Script ที่ใช้กด key ลัด สร้างไฟล์ .BAT ที่ใช้สำหรับรัน powershell Script จากนั้นให้ใช้โปรแกรม Task Scheduler ซึ่งมีใน window 10 อยู่แล้วในการตั้งเวลาให้รันไฟล์ .BAT ที่ใช้สำหรับรัน powershell Script ที่ใช้รัน flow เวลา 20.00 น. ทุกวันเสาร์



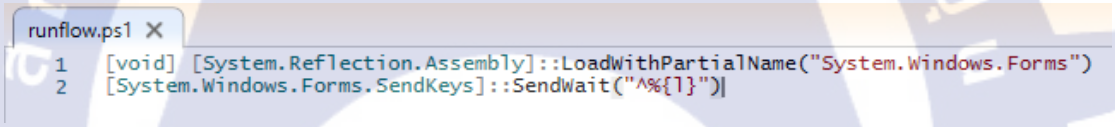
รูปที่ 3.8 การสร้างไฟล์ .BAT รัน Python Script ที่ใช้สกัดข้อมูล และทำความสะอาดข้อมูล



รูปที่ 3.9 Power Automate Desktop สำหรับสร้าง flow อัพเดทข้อมูลอัตโนมัติ

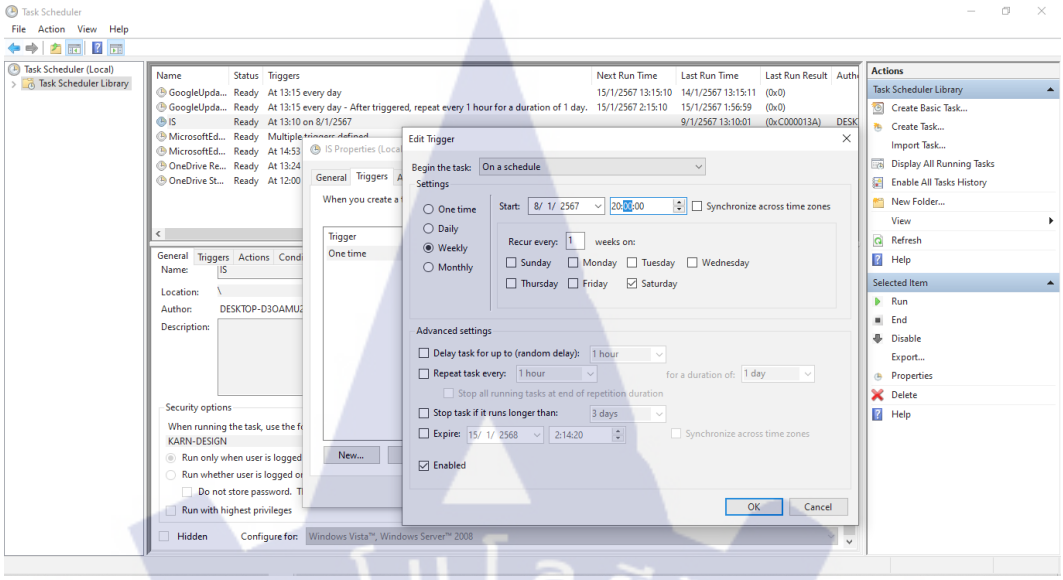


รูปที่ 3.10 การตั้งค่าให้รัน flow ด้วย key ลัด Ctrl + Alt + L



รูปที่ 3.11 การสร้างไฟล์ .BAT ที่ใช้สำหรับรัน powershell Script

รูปที่ 3.12 การสร้างไฟล์ powershell Script ที่ใช้กด key ลัด

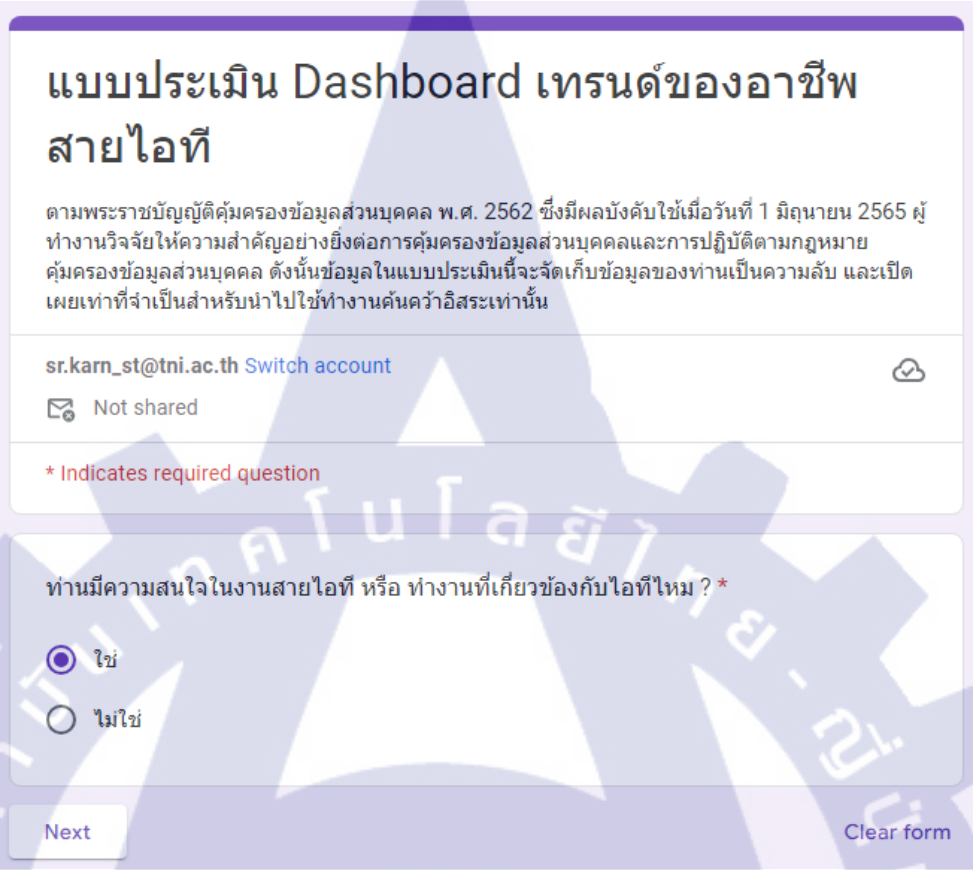


รูปที่ 3.13 การตั้งเวลาให้รันไฟล์ .BAT ด้วยโปรแกรม Task Scheduler ใน window 10

3.6. ทดสอบระบบ

ทดสอบระบบมี 2 ส่วน ส่วนแรกทำการจับเวลาการทำงานของระบบในส่วนต่าง ๆ มีหัวข้อดังนี้ สกัดข้อมูลเว็บไซต์ jobsblognone สกัดข้อมูลเว็บไซต์ jobBBK สกัดข้อมูลเว็บไซต์ jobthai สกัดข้อมูลเว็บไซต์ jobsDB หน่วงเวลา รวมตารางข้อมูล ทำความสะอาดข้อมูล อัพเดทฐานข้อมูล ข้อมูล อัพเดทข้อมูลใน Dashboard และระยะเวลาทั้งหมด ส่วนที่สองทำแบบสอบถามความพึงพอใจของผู้ที่ได้เข้ามาเยี่ยมชมต่อ Dashboard ด้วย Google Form ในส่วนนี้จะมีหัวข้อด้านความครบถ้วนของข้อมูล กับด้านการออกแบบ Dashboard ซึ่งจะมี 2 แบบให้ดู

รูปที่ 3.14 หน้าต่างการทำงานของระบบในส่วนต่าง ๆ



รูปที่ 3.15 ตัวอย่าง Google Form แบบสอบถามความพึงพอใจผู้เยี่ยมชมต่อ Dashboard

บทที่ 4
ผลการวิจัย

การสกัดข้อมูลบนเว็บไซต์หางานอาชีพสาย IT ในประเทศไทย และแสดงผลด้วยภาพเพื่อใช้ประกอบการตัดสินใจในการเลือกงาน เป็นการออกแบบ และพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติรวมทั้งสร้างรายงานการวิเคราะห์ข้อมูลโดยใช้ Data Visualization ต่าง ๆ ในรูปแบบรายงาน Dashboard เพื่อให้ผู้สนใจที่จะเปลี่ยนสายงานมาทำงานสาย IT ได้นำไปใช้ประกอบการตัดสินใจในการเลือกงาน และการตัดสินใจในการศึกษาต่อ

4.1 ลักษณะฐานข้อมูลหลัก

ข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยได้รวบรวมไว้ตั้งแต่วันที่ 12/2/2566 ถึงวันที่ 8/1/2567 โดยชุดข้อมูลที่เริ่มต้นใช้มาจากการสกัดข้อมูลทั้ง 4 เว็บไซต์หางาน หลังจากทำความสะอาดข้อมูลแล้วมีข้อมูลรวมกันมากกว่า 55,638 แถว เซฟเป็นไฟล์ CSV ฐานข้อมูลหลักมีลักษณะ ตามตารางที่ 4.1

ตารางที่ 4.1 ลักษณะตารางฐานข้อมูลหลัก

ชื่อคอลัม	ลักษณะข้อมูล
ID_Job	เลข ID โพสรายละเอียดรับสมัครงาน
date	วันที่ทำการสกัดข้อมูลจากเว็บไซต์
url	URL ของหน้าเว็บไซต์รายละเอียดงาน
Company Name	ชื่อบริษัทรับสมัครงาน
Job_Position	ตำแหน่งงานที่สกัดมาจากเว็บไซต์
tag_Jobname	เลข ID อ้างอิงประเภทตำแหน่งงาน
tag_add	เลข ID อ้างอิงจังหวัด
Salary min	ตัวเลขเงินเดือนช่วงต่ำ
Salary max	ตัวเลขเงินเดือนช่วงสูง
ชื่อเครื่องมือหรือภาษาโปรแกรมมิ่งจำนวน 190 คอลัมน์	ข้อมูลเป็น 1 (True) หรือ 0(False)

ID_Job	date	url	Company Name	Job Position	tag_Jobname	tag_add	Salary min	Salary max	abap	.net	kotlin	c	c#
1	12/2/2023	https://www.jobbkk.com/job	Niche-Est Solution (Thailand) Co.,Ltd.	system analyst	72	70	20000	45000	0	0	0	0	
2	12/2/2023	https://www.jobbkk.com/job	บริษัท จีเน็ตซอฟต์ จำกัด	automated tester	81	69			0	0	0	0	
3	12/2/2023	https://www.jobbkk.com/job	บริษัท แลนไทย (ประเทศไทย) จำกัด	it support	43	71	18000	22000	0	0	0	0	
4	12/2/2023	https://www.jobbkk.com/job	บริษัท เงินเทอร์โบ จำกัด	quality assurance engineer (ประจำจังหวัดนนทบุ	56	75			0	0	0	0	
5	12/2/2023	https://www.jobbkk.com/job	บริษัท ซิทีโอ โซลูชั่น จำกัด	.net c# programmer	54	69			0	1	0	0	
6	12/2/2023	https://www.jobbkk.com/job	บริษัท เน็กซ์พลัส เอ็นจีเนียริ่ง จำกัด	it support ประจำสำนักงานใหญ่	43	69			0	0	0	0	
7	12/2/2023	https://www.jobbkk.com/job	บริษัท อินเดอร์ วิชัน บีสซิเนสกรุ๊ป จำกัด	system analyst (รับสมัครด่วน)	72	69			0	0	0	0	
8	12/2/2023	https://www.jobbkk.com/job	บริษัท อินเดอร์ วิชัน บีสซิเนสกรุ๊ป จำกัด	php web programmer (รับสมัครด่วน)	54	69			0	0	0	0	
9	12/2/2023	https://www.jobbkk.com/job	บริษัท อินเดอร์ วิชัน บีสซิเนสกรุ๊ป จำกัด	software project manager	38	69			0	0	0	0	
10	12/2/2023	https://www.jobbkk.com/job	Aware Corporation Limited	sr. software engineer (java) - bkk	66	69			0	0	0	0	
11	12/2/2023	https://www.jobbkk.com/job	บริษัท เงินเทอร์โบ จำกัด	software engineer (ประจำจังหวัดนนทบุรี)	66	75			0	0	0	0	
12	12/2/2023	https://www.jobbkk.com/job	บริษัท ซอฟต์สแควร์ 1999 จำกัด	software developer (ประจำสาขาชลบุรี)	20	50			0	0	0	0	
13	12/2/2023	https://www.jobbkk.com/job	บริษัท ซอฟต์สแควร์ 1999 จำกัด	software developer (สนง.ใหญ่ จ.ปทุมธานี)	20	70			0	0	0	0	
14	12/2/2023	https://www.jobbkk.com/job	บริษัท ซอฟต์สแควร์ 1999 จำกัด	software developer (ประจำสาขาเชียงใหม่)	20	2			0	0	0	0	
15	12/2/2023	https://www.jobbkk.com/job	บริษัท วีเอสที อีซีเอส (ประเทศไทย) จำกัด	product manager (it products)	51	71			0	0	0	0	
16	12/2/2023	https://www.jobbkk.com/job	บริษัท วีเอสที อีซีเอส (ประเทศไทย) จำกัด	product sales (f5)	58	71			0	0	0	0	
17	12/2/2023	https://www.jobbkk.com/job	เอเชี่ยมอเตอร์ เซอร์วิส เซ็นเตอร์ จำกัด	it admin	30	70			0	0	0	0	
18	12/2/2023	https://www.jobbkk.com/job	บริษัท เกียรติธนาขนส่ง จำกัด	เจ้าหน้าที่ system analysis	72	75			0	0	0	0	
19	12/2/2023	https://www.jobbkk.com/job	บริษัท วินเกม เทค คอร์ปอเรชั่น จำกัด	front end developer	20	69	30000	60000	0	0	0	0	
20	12/2/2023	https://www.jobbkk.com/job	บริษัท ซิสเต็ม อีโวลูชั่น จำกัด	web programmer , โปรแกรมเมอร์ php	54	75	18000	50000	0	0	0	0	
21	12/2/2023	https://www.jobbkk.com/job	บริษัท อาร์ค อิเวนท์ จำกัด	it.support	43	71			0	0	0	0	
22	12/2/2023	https://www.jobbkk.com/job	บริษัท ดั๊ชธนสิน จำกัด	it support	43	70			0	0	0	0	
23	12/2/2023	https://www.jobbkk.com/job	เจเคเอ็น ไฮ ซ็อบบิ้ง จำกัด	sr. java web programmer	54	66			0	0	0	0	
24	12/2/2023	https://www.jobbkk.com/job	บริษัท อาร์ค อิเวนท์ จำกัด	programmer	54	71			0	0	0	0	
25	12/2/2023	https://www.jobbkk.com/job	บริษัท บุญถาวรเซรามิค จำกัด	digital platform development officer	43	71			0	0	0	0	
26	12/2/2023	https://www.jobbkk.com/job	แพ็คไห้	developer	20	75	25000	40000	0	1	0	0	
27	12/2/2023	https://www.jobbkk.com/job	บริษัท โคเวอ์ (ประเทศไทย) จำกัด	java programmer	54	69			0	0	0	0	
28	12/2/2023	https://www.jobbkk.com/job	บริษัท ซัน เคทเวย์ จำกัด	business analyst	11	69			0	0	0	0	
29	12/2/2023	https://www.jobbkk.com/job	บริษัท ซอฟต์สแควร์ 1999 จำกัด	project coordinator (ด้าน it)	55	70			0	0	0	0	
30	12/2/2023	https://www.jobbkk.com/job	บริษัท เทคนิคออซอฟต์ จำกัด	it support application	7	70	10000	20000	0	0	0	0	

รูปที่ 4.1 ลักษณะตารางฐานข้อมูลหลัก

ต้องสร้างตารางข้อมูลที่ใช้สร้าง Dashboard ใน Power BI สำหรับงานวิจัยนี้วิเคราะห์จากลักษณะฐานข้อมูลหลักไม่สามารถนำมาใช้ได้ทันทีต้องมีการกลับด้านคอลัมน์ (Unpivot Column) คอลัมน์ทักษะ และเครื่องมือ จึงจำเป็นต้องแยกตารางออกเป็น 2 ตารางทำให้มีโดยมีตารางข้อมูล Fact 2 ตาราง และ ตารางข้อมูล Dimension 3 ตาราง ตารางข้อมูล Fact 2 ตารางมีตารางข้อมูลรับสมัครงานมีลักษณะตามตารางที่ 4.2 และตารางข้อมูลทักษะ และเครื่องมือตามตารางที่ 4.3 ส่วนตารางข้อมูล Dimension 3 ตารางมีตารางอ้างอิงข้อมูลในตาราง Fact ทักษะ และเครื่องมือมีลักษณะตามตารางที่ 4.4 ตารางอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_Jobname มีลักษณะตามตารางที่ 4.5 และตารางอ้างอิงข้อมูลในตาราง Fact คอลัมน์ tag_add มีลักษณะตามตารางที่ 4.6

ตารางที่ 4.2 ลักษณะตาราง Fact ข้อมูลรับสมัครงาน

ชื่อคอลัม	ลักษณะข้อมูล
ID_Job	เลข ID โพสรายละเอียดรับสมัครงาน
date	วันที่ทำการสกัดข้อมูลจากเว็บไซต์
url	URL ของหน้าเว็บไซต์รายละเอียดงาน
Company Name	ชื่อบริษัทรับสมัครงาน
Job_Position	ตำแหน่งงานที่สกัดมาจากเว็บไซต์
tag_Jobname	เลข ID อ้างอิงประเภทตำแหน่งงาน
tag_add	เลข ID อ้างอิงจังหวัด
Salary min	ตัวเลขเงินเดือนช่วงต่ำ
Salary max	ตัวเลขเงินเดือนช่วงสูง

date	ID_Job	url	Company Name	Job Position	tag_Jobnan	tag_ad	Salary min	Salary max
8 ตุลาคม 2566	44681	https://www.jobthai.com/t	บริษัท ไอวีพลัสแนลเวิน จำกัด	it support - system operator(bts อ่อนนุช)	75	69	20000	25000
22 ตุลาคม 2566	45631	https://www.jobbkk.com/j	The Thai Silk Co., Ltd	it supervisor	42	69	20000	40000
22 ตุลาคม 2566	46316	https://th.jobsdb.com/th/t	ไม่เปิดเผยชื่อบริษัท	mobile test engineer (junior/senior)	11	69	20000	45000
4 พฤศจิกายน 2566	47614	https://th.jobsdb.com/th/t	TT CABLE AND DATA SOLUTION CO., LTD.	network engineer	48	69	20000	45000
4 พฤศจิกายน 2566	47619	https://th.jobsdb.com/th/t	Port-IT Co., Ltd.	it support engineer	43	69	20000	45000
12 พฤศจิกายน 256	48632	https://th.jobsdb.com/th/t	TT CABLE AMD DATA SOLUTION CO., LTD.	project engineer (it)	34	69	20000	45000
12 พฤศจิกายน 256	48651	https://th.jobsdb.com/th/t	Miwcom Co., Ltd.	it support professional/เจ้าหน้าที่ไอทีซัพพอร์ต	43	69	20000	45000
12 พฤศจิกายน 256	48747	https://th.jobsdb.com/th/t	QA Hive Company Limited	qa / software tester (manual / automated)	81	69	20000	45000
12 พฤศจิกายน 256	49264	https://www.jobthai.com/t	บริษัท ไทยรุท ดอทคอม จำกัด	senior/junior php programmer (ทำงานกรุงเทพ)	54	69	20000	60000
18 พฤศจิกายน 256	50066	https://www.jobthai.com/t	Digital Commerce Co., Ltd.	ux/ui designer	82	69	20000	27000
18 พฤศจิกายน 256	50103	https://www.jobthai.com/t	Orenburg Consulting Group	cloud operation engineer (hybird)	13	69	20000	50000
25 พฤศจิกายน 256	50794	https://th.jobsdb.com/th/t	Unity Focus Co., Ltd.	system/automation/plc/scada engineer/system/autor	73	69	20000	45000
2 ธันวาคม 2566	51069	https://www.jobbkk.com/j	บริษัท ไซเทค ซิสเต็ม จำกัด	noc & soc engineer รับสมัครด่วน	49	69	20000	45000
2 ธันวาคม 2566	51070	https://www.jobbkk.com/j	บริษัท ไซเทค ซิสเต็ม จำกัด	sale it security	61	69	20000	100000
2 ธันวาคม 2566	51072	https://www.jobbkk.com/j	บริษัท ไซเทค ซิสเต็ม จำกัด	network engineer	48	69	20000	45000
2 ธันวาคม 2566	51452	https://th.jobsdb.com/th/t	FEYVERLY CO., LTD.	software tester/qa	81	69	20000	45000
9 ธันวาคม 2566	51840	https://www.jobbkk.com/j	บริษัท สยามสไมล์ ประกันภัย จำกัด (มหาชน	เปิดรับสมัครด่วน programmer พิจารณาจบใหม่ สถานา	54	69	20000	50000
9 ธันวาคม 2566	51851	https://www.jobbkk.com/j	บริษัท จัดหางาน เพอร์ซันแนล คอนซัลแตนา	it officer เพศชาย ประสบการณ์ it support ประจำออฟฟ	43	69	20000	28000
9 ธันวาคม 2566	52004	https://www.jobthai.com/t	A.R.E.E Technology Solutions	it service help desk (bts กรุงธนบุรี)	63	69	20000	23000
16 ธันวาคม 2566	52705	https://th.jobsdb.com/th/t	QA Hive Company Limited	qa / software tester (manual / automated)	81	69	20000	45000
16 ธันวาคม 2566	52882	https://th.jobsdb.com/th/t	Unity Focus Co., Ltd.	system/automation/plc/scada engineer	73	69	20000	45000
23 ธันวาคม 2566	53912	https://www.jobthai.com/t	Bewell (Digital Commerce Co., Ltd.)	ux/ui designer	82	69	20000	27000
23 ธันวาคม 2566	53952	https://www.jobthai.com/t	1577 Infinite Co., Ltd.	automate tester (ดึก p.s. tower btsอโศก/mrtสุขุมวิท)	81	69	20000	50000
1 มกราคม 2567	54032	https://www.jobbkk.com/j	บริษัท พีเพิล พลัส ซอฟต์แวร์ จำกัด	project leader or project manager	38	69	20000	70000

รูปที่ 4.2 ตัวอย่างตารางข้อมูลรับสมัครงาน

ตารางที่ 4.3 ลักษณะตาราง Fact ข้อมูลทักษะ และเครื่องมือ

ชื่อคอลัม	ลักษณะข้อมูล
ID_Job	เลข ID โพสรายละเอียดรับสมัครงาน
tool	เลข ID อ้างอิงทักษะ และเครื่องมือ
value	ข้อมูลเป็น 1 แสดงว่าเป็น True

ID_Job	value	tool
10	1	19
12	1	19
13	1	19
14	1	19
23	1	19
33	1	19
45	1	19
56	1	19
60	1	19
61	1	19

รูปที่ 4.3 ตัวอย่างตารางข้อมูลทักษะ และเครื่องมือ

ตารางที่ 4.4 ลักษณะตารางข้อมูล Dimension ทักษะ และเครื่องมือ

ชื่อคอลัม	ลักษณะข้อมูล
tool	เลข ID อ้างอิงทักษะ และเครื่องมือ
tool_tag	ชื่อ Tag ทักษะ และเครื่องมือ
tool_name	ชื่อทักษะ และเครื่องมือที่ใช้แสดงใน Dashboard

tool_name	tool	tool_tag
Abap	1	abap
.Net	2	.net
Kotlin	3	kotlin
C	4	c
C++	5	c+
C#	6	c#
Objective-C	7	objective-c
Swift	8	swift
Java	9	java
Js	10	js
Javascript	11	javascript
HTML	12	html
CSS	13	css

รูปที่ 4.4 ตัวอย่างตารางข้อมูล Dimension ทักษะ และเครื่องมือ

ตารางที่ 4.5 ลักษณะตารางข้อมูล Dimension tag_Jobname

ชื่อคอลัม	ลักษณะข้อมูล
tag_Jobname	เลข ID อ้างอิงประเภทตำแหน่งงาน
tag_Jobname_name	ชื่อ Tag ประเภทตำแหน่งงาน
Jobname	ชื่อตำแหน่งงานที่ใช้แสดงใน Dashboard

tag_Jobname	tag_Jobname_name	Jobname
1	3d model	3D Model
2	account manager	Account Manager
3	application consultant	Application Consultant
4	application developer	Application Developer
5	application engineer	Application Engineer
6	application owner	Application Owner
7	application support	Application Support
8	automation engineer	Automation Engineer
9	back end developer	Back End Developer
10	bpr consultant	BPR Consultant

รูปที่ 4.5 ตัวอย่างตารางข้อมูล Dimension tag_Jobname

ตารางที่ 4.6 ลักษณะตารางข้อมูล Dimension tag_add

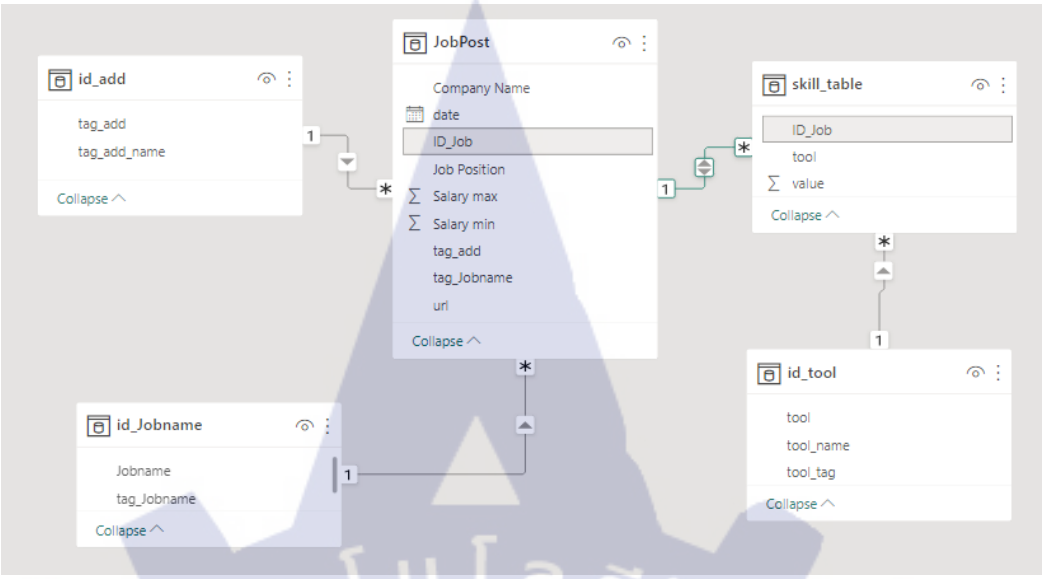
ชื่อคอลัม	ลักษณะข้อมูล
tag_add	เลข ID อ้างอิงที่อยู่ที่ทำงานเป็นจังหวัด
tag_add_name	ชื่อ Tag ที่อยู่ที่ทำงานเป็นจังหวัด

tag_add	tag_add_name
1	นครราชสีมา
2	เชียงใหม่
3	กาญจนบุรี
4	ตาก
5	อุบลราชธานี
6	สุราษฎร์ธานี
7	ชัยภูมิ
8	แม่ฮ่องสอน
9	เพชรบูรณ์
10	ลำปาง

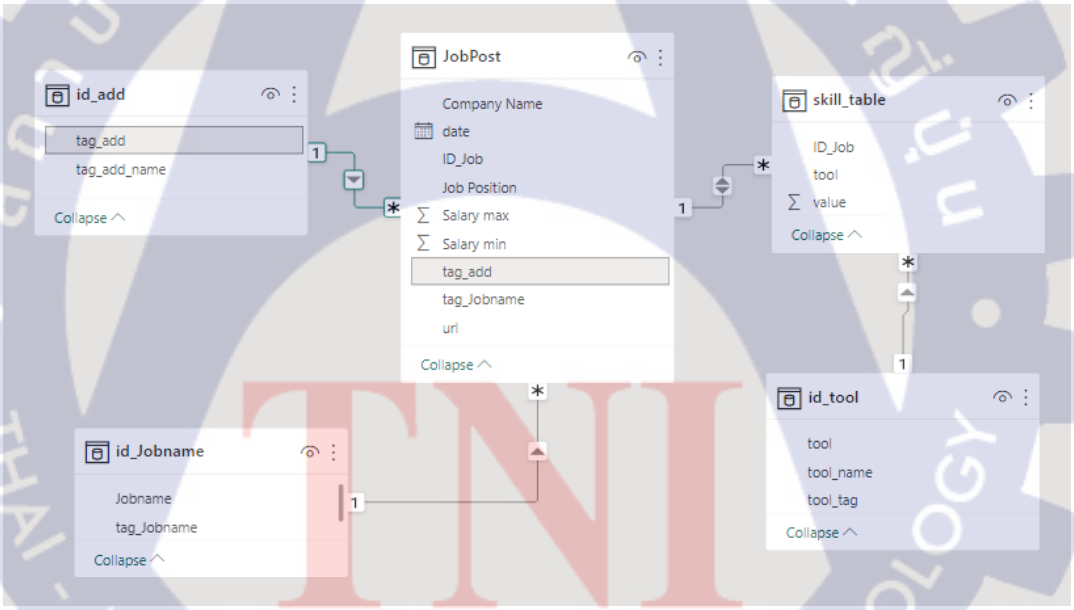
รูปที่ 4.6 ตัวอย่างตารางข้อมูล Dimension tag_add

4.2 แบบจำลองข้อมูล (Data Model)

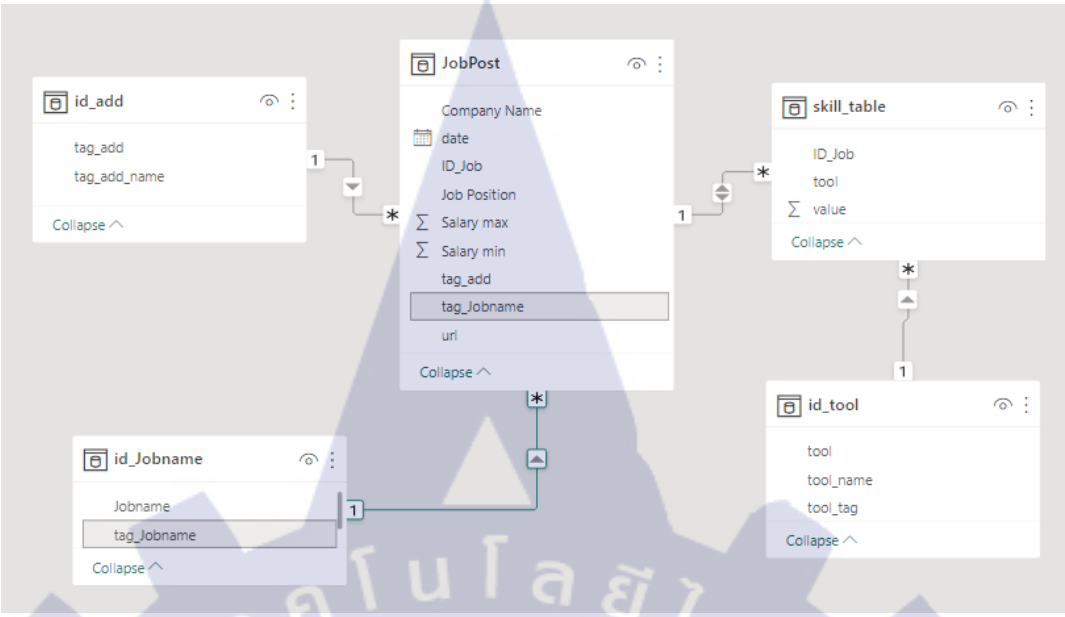
ในการออกแบบ Data Model สำหรับงานวิจัยนี้วิเคราะห์จากลักษณะฐานข้อมูลหลักไม่สามารถนำมาใช้ได้ทันทีต้องมีการกลับด้านคอลัมน์ (Unpivot Column) คอลัมน์ทักษะ และเครื่องมือ จึงจำเป็นต้องแยกตารางออกเป็น 2 ตารางทำให้มีตาราง Fact 2 ตาราง และมีตาราง Dimension 3 ตาราง Data Model จึงอยู่ในรูปแบบของ Snowflake Schema โดยระหว่างตาราง Fact 2 ตาราง เชื่อมต่อความสัมพันธ์ด้วยคอลัมน์ ID_Job แบบ Many to one และสองทางตามรูปที่ 4.7 ตาราง Fact ตารางข้อมูลรับสมัครงานกับตาราง Dimension อ้างอิงข้อมูลที่อยู่ที่ทำงานเชื่อมต่อความสัมพันธ์ด้วยคอลัมน์ tag_add แบบ Many to one และหนึ่งทาง ตามรูปที่ 4.8 ตาราง Fact ตารางข้อมูลรับสมัครงานกับตาราง Dimension อ้างอิงข้อมูลตำแหน่งงาน เชื่อมต่อความสัมพันธ์ด้วยคอลัมน์ tag_Jobname แบบ Many to one และหนึ่งทาง ตามรูปที่ 4.9 ตาราง Fact ตารางข้อมูลทักษะ และเครื่องมือ กับตาราง Dimension อ้างอิงข้อมูลทักษะ และเครื่องมือ เชื่อมต่อความสัมพันธ์ด้วยคอลัมน์ tool แบบ Many to one และหนึ่งทาง ตามรูปที่ 4.10



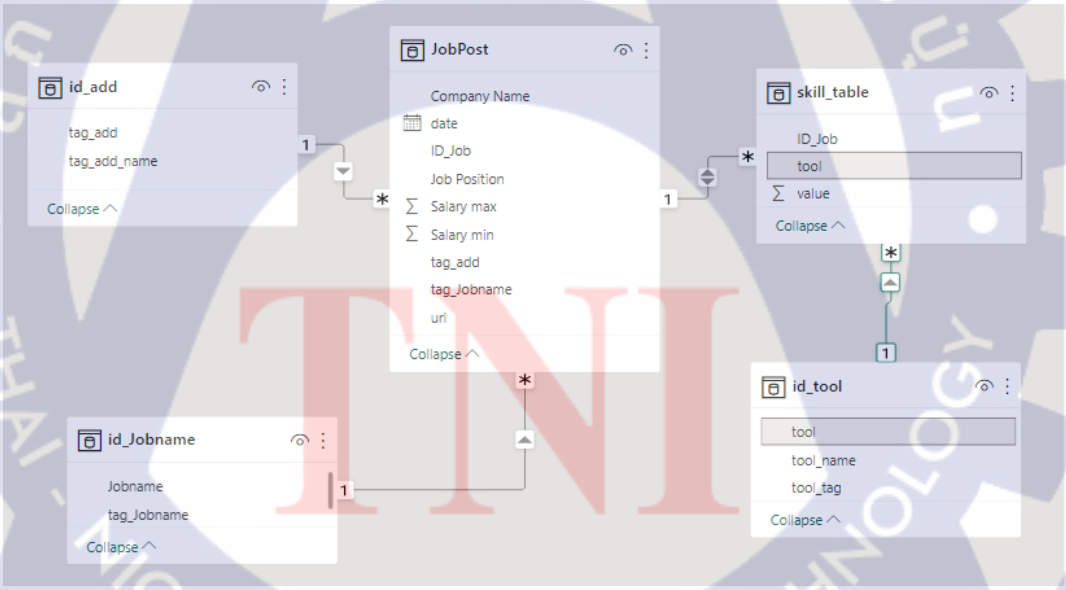
รูปที่ 4.7 เชื่อมต่อความสัมพันธ์ระหว่าง ตาราง Fact 2 ตาราง



รูปที่ 4.8 เชื่อมต่อความสัมพันธ์ระหว่างตาราง Fact ข้อมูลรับสมัครงานกับตาราง Dimension ข้อมูลที่อยู่ที่ทำงาน



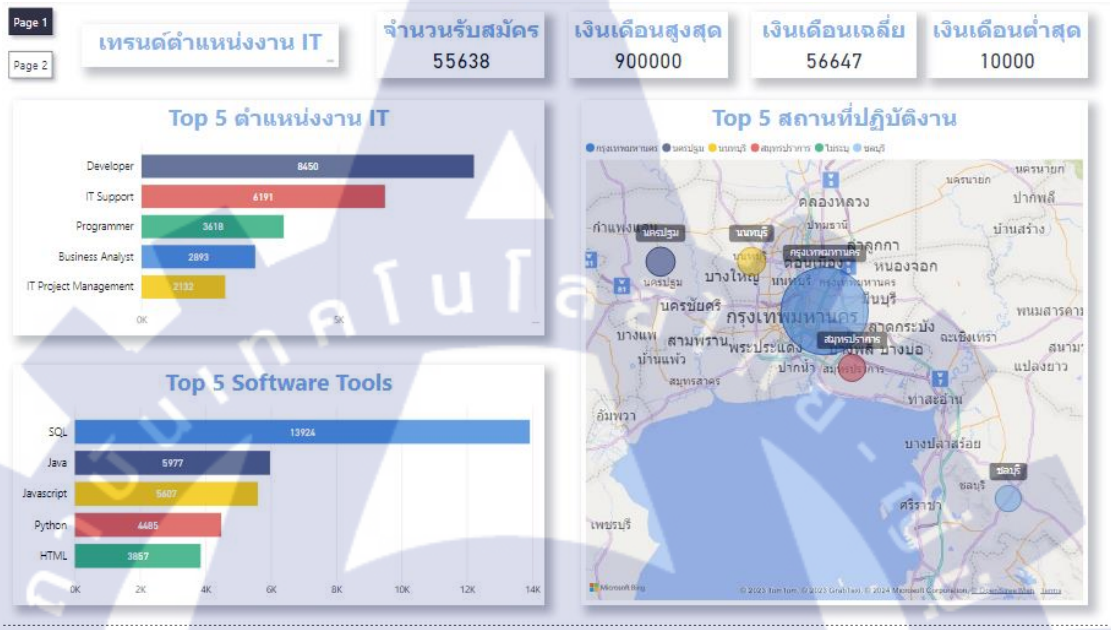
รูปที่ 4.9 เชื่อมต่อความสัมพันธ์ระหว่าง ตารางข้อมูลรับสมัครงานกับตาราง Dimension ข้อมูลตำแหน่งงาน



รูปที่ 4.10 เชื่อมต่อความสัมพันธ์ระหว่าง ตารางข้อมูลรับสมัครงานกับตาราง Dimension ข้อมูลทักษะ และเครื่องมือ

4.3 Data Visualization

ในการวิเคราะห์ข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยที่ได้รวบรวมไว้ตั้งแต่วันที่ 12/2/2566 ถึงวันที่ 8/1/2567 ได้แบ่งออกเป็น 2 ส่วน จึงได้สร้าง Dashboard ให้มี 2 หน้าในหนึ่งหน้าของ Dashboard จะแสดงข้อมูลให้จบหนึ่งส่วน



รูปที่ 4.11 Dashboard เทรนด์ตำแหน่งงาน IT หน้า 1

Dashboard เทรนด์ตำแหน่งงาน IT หน้าแรกดังรูปที่ 4.11 จะแสดงข้อมูลที่วิเคราะห์ออกมาเป็นเทรนด์ต่าง ๆ ของการรับสมัครงาน IT โดยรวมบ่งบอกแนวโน้มลักษณะความต้องการแรงงานในตลาดแรงงาน IT มี 5 หัวข้อ มีดังนี้

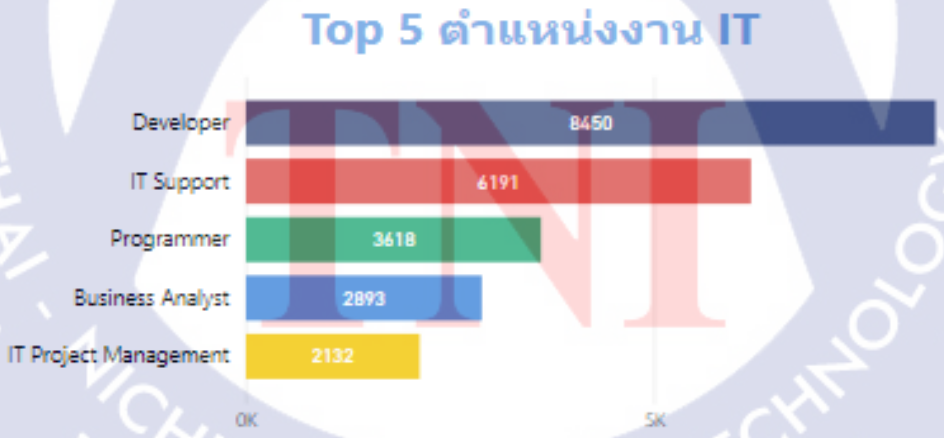
1. จำนวนรับสมัคร เป็นจำนวนรับสมัครงานตำแหน่งงาน IT ทั้งหมดที่ได้เก็บรวบรวมมาจากเว็บไซต์หางานมีจำนวน 55,638 อัตรา แสดงข้อมูลโดยในรูปแบบ card ดังรูปที่ 4.12

รูปที่ 4.12 จำนวนรับสมัคร Dashboard เทรนด์ตำแหน่งงาน IT หน้า 1

2. เงินเดือน แสดงเงินเดือนที่ทางผู้รับสมัครงานเสนอให้ผู้สมัครงานตำแหน่งงาน IT มีหน่วยเป็นบาท ซึ่งแบ่งย่อยเป็น 3 หัวข้อ เป็นเงินเดือนสูงสุดเท่าที่มีในการรับสมัครงานตำแหน่งงาน IT เป็นจำนวนเงิน 900,000 บาท เงินเดือนเฉลี่ยเป็นจำนวนเงิน 56,647 บาท และเงินเดือนต่ำสุดเท่าที่มีในการรับสมัครงานตำแหน่งงาน IT เป็นจำนวนเงิน 10,000 บาท แสดงข้อมูลโดยในรูปแบบ card ดังรูปที่ 4.13

รูปที่ 4.13 ข้อมูลสถิติเงินเดือน Dashboard เทรนด์ตำแหน่งงาน IT หน้า 1

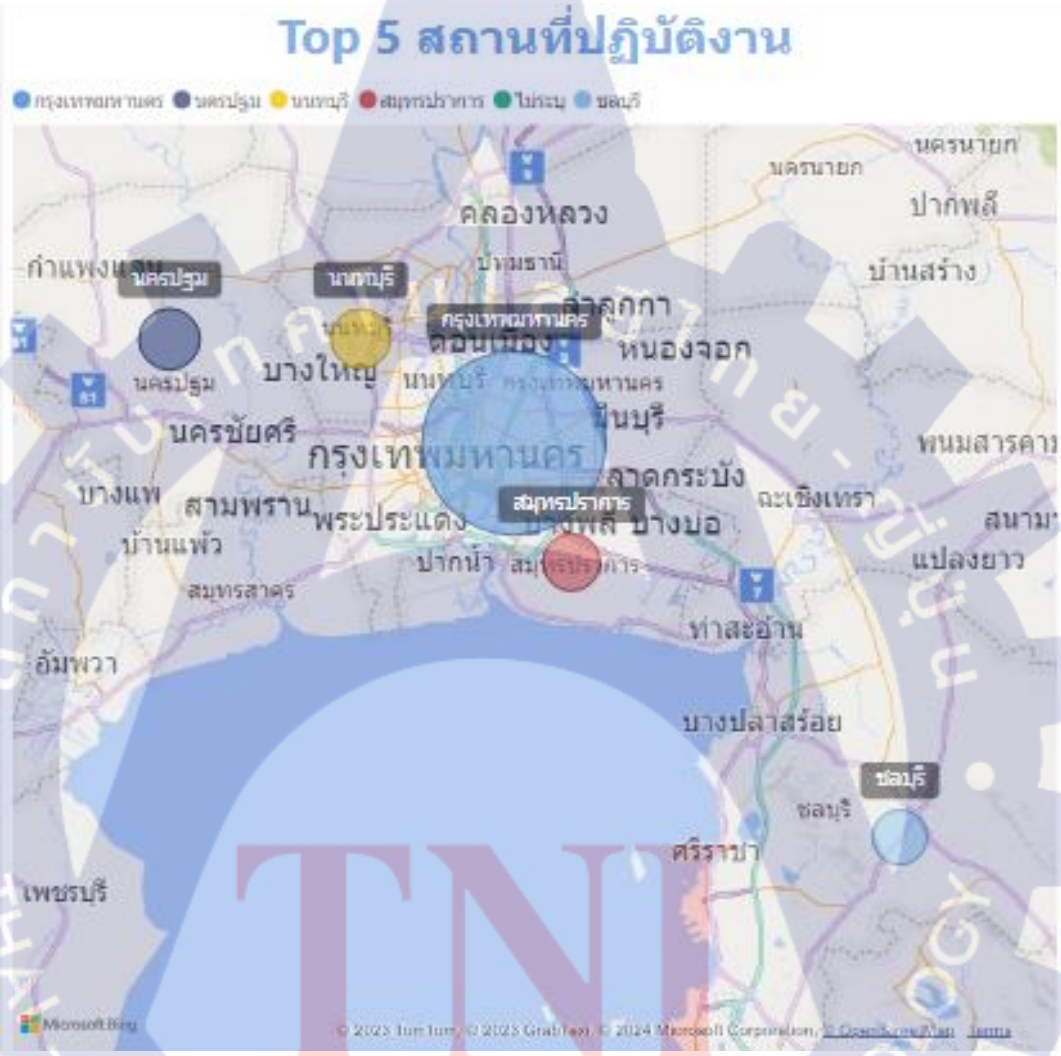
3. Top 5 ตำแหน่งงาน IT แสดงถึง 5 อันดับตำแหน่งงาน IT ที่มีจำนวนการเปิดรับสมัครมากที่สุดจากข้อมูลที่เก็บรวบรวมมาแสดงข้อมูลเป็นกราฟแท่งเพื่อให้ง่ายต่อการเปรียบเทียบ อันดับที่ 1 คือ Developer มีจำนวนการรับสมัคร 8,450 อัตรา อันดับที่ 2 คือ IT Support มีจำนวนการรับสมัคร 6,191 อัตรา อันดับที่ 3 คือ Programmer มีจำนวนการรับสมัคร 3,618 อัตรา อันดับที่ 4 คือ Business Analyst มีจำนวนการรับสมัคร 2,893 อัตรา และอันดับที่ 5 คือ IT Project Management มีจำนวนรับสมัคร 2,132 อัตรา ดังรูปที่ 4.14



รูปที่ 4.14 Top 5 ตำแหน่งงาน IT

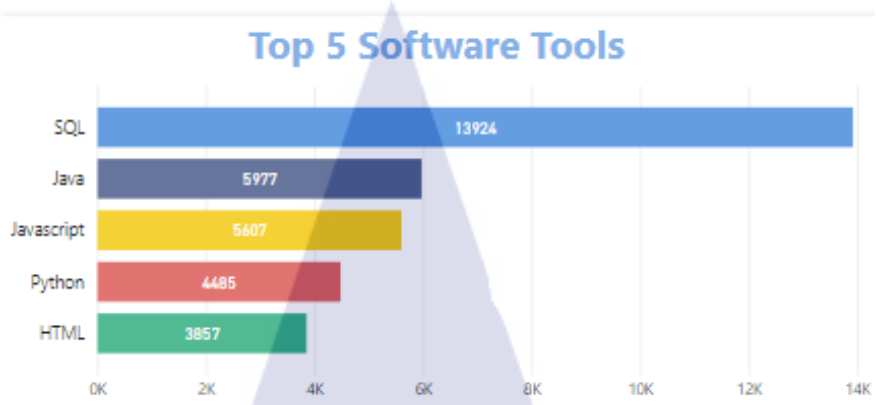
4. Top 5 สถานที่ปฏิบัติงาน แสดงถึง 5 อันดับสถานที่ปฏิบัติงานที่นิยมที่สุดเป็นจำนวนการเปิดรับสมัครงานในแต่ละจังหวัดจากข้อมูลที่เก็บรวบรวมมาแสดงข้อมูลเป็น bubble map ซึ่ง

แสดงทั้งตำแหน่งจังหวัดบนแผนที่จากจุด bubble และแสดงปริมานจากขนาด bubble อันดับที่ 1 กรุงเทพ จำนวน 38,563 อันดับที่ 2 นครปฐม จำนวน 3,283 อันดับที่ 3 นนทบุรี จำนวน 2,745 อันดับที่ 4 สมุทรปราการ จำนวน 2,139 และอันดับที่ 5 ชลบุรี จำนวน 1,482 แสดงว่างานกระจุกตัวที่ภาคกลางเป็นหลักดังรูปที่ 4.15

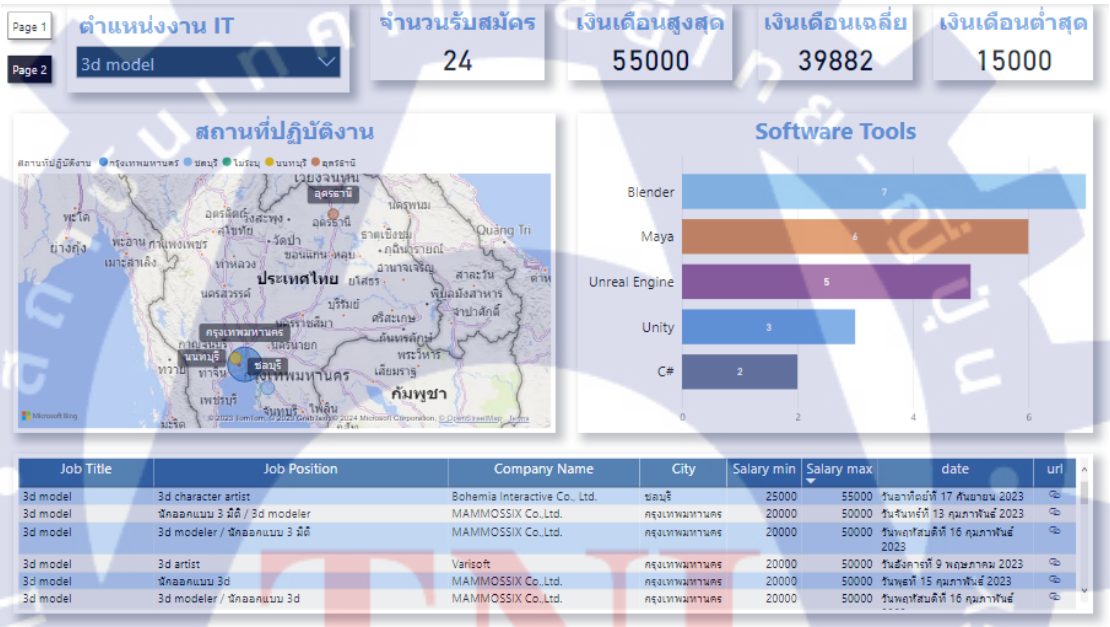


รูปที่ 4.15 Top 5 สถานที่ปฏิบัติงาน

5. Top 5 Software Tools แสดงถึง 5 อันดับเป็นจำนวนความต้องการทักษะในการใช้ภาษาโปรแกรมมิ่งหรือเครื่องมือที่ต้องการมากที่สุดจากข้อมูลการเปิดรับสมัครงานที่เก็บรวบรวมมาแสดงข้อมูลเป็นกราฟแท่งเพื่อให้ง่ายต่อการเปรียบเทียบ อันดับที่ 1 คือ SQL จำนวน 13,924 อันดับที่ 2 คือ Java จำนวน 5,977 อันดับที่ 3 คือ JavaScript จำนวน 5,607 อันดับที่ 4 คือ Python จำนวน 4,485 และอันดับที่ 5 คือ HTML จำนวน 3,857 ดังรูปที่ 4.16



รูปที่ 4.16 Top 5 Software Tools

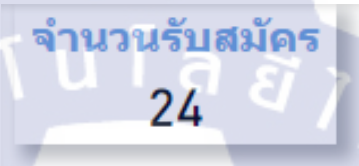


รูปที่ 4.17 Dashboard เทรนด์ตำแหน่งงาน IT หน้า 2

รูปที่ 4.18 Dropdown list ตำแหน่งงานที่ต้องการดู

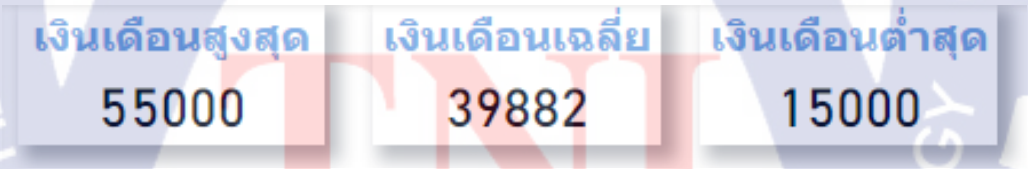
Dashboard เทรนด์ตำแหน่งงาน IT หน้าที่ 2 ดังรูปที่ 4.17 จะแสดงข้อมูลที่วิเคราะห์ออกมาเป็นเทรนด์ต่าง ๆ ของการรับสมัครงาน IT เฉพาะตำแหน่งงานที่ต้องการดูใช้วิธีเลือกเป็น drop down list ดังรูปที่ 4.18 Dropdown list ตำแหน่งงานที่ต้องการดู ทำให้สามารถบ่งบอกแนวโน้มลักษณะความต้องการแรงงานเฉพาะตำแหน่งที่ดูในตลาดแรงงาน IT มี 4 หัวข้อ พร้อมทั้งมีตารางข้อมูลการรับสมัครงานแต่ละตำแหน่งงานที่ลงในเว็บไซต์หางาน และมีลิ้งเข้าหน้าเว็บไซต์ในตำแหน่งงานนั้น ดังนี้

1. จำนวนรับสมัคร แสดงจำนวนรับสมัครงานตามตำแหน่งงาน IT ที่ได้เลือกดูจากข้อมูลที่เก็บรวบรวมมาจากเว็บไซต์หางานแสดงข้อมูลเป็น card ดังรูปที่ 4.19



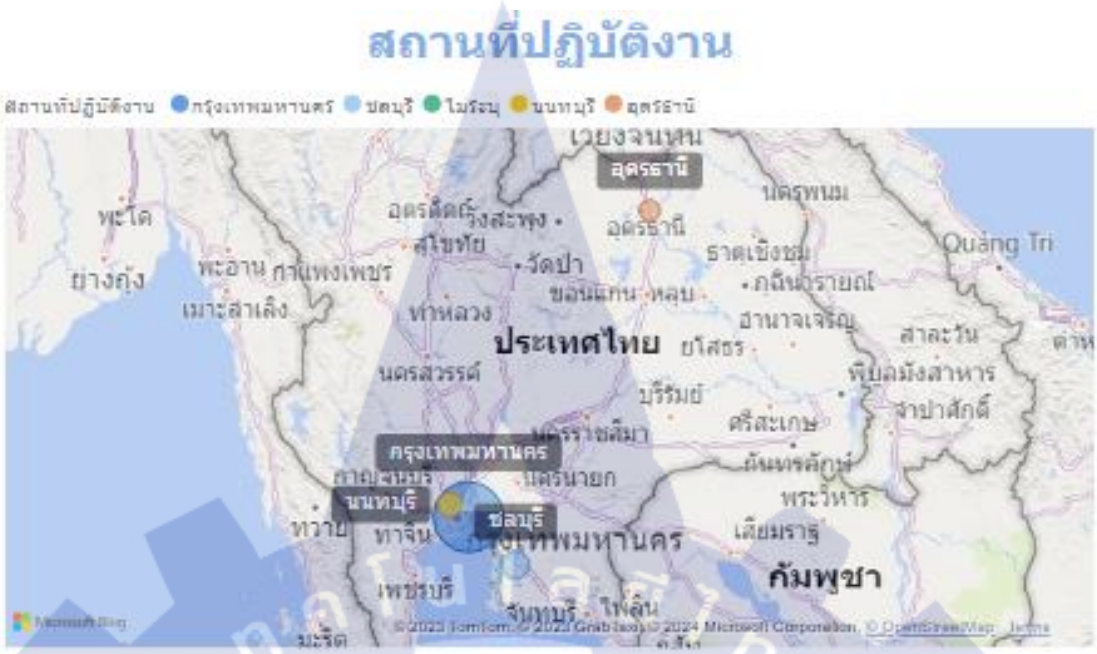
รูปที่ 4.19 จำนวนรับสมัคร Dashboard เทรนด์ตำแหน่งงาน IT หน้า 2

2. เงินเดือน แสดงเงินเดือนตามตำแหน่งงาน IT ที่เลือกดู ซึ่งแบ่งย่อยเป็น 3 หัวข้อ เป็นเงินเดือนสูงสุดเท่าที่มีในการรับสมัครงานตำแหน่งงาน IT เงินเดือนเฉลี่ย และเงินเดือนต่ำสุดเท่าที่มีในการรับสมัครงานตำแหน่งงาน IT แสดงข้อมูลโดยในรูปแบบ card ดังรูปที่ 4.20



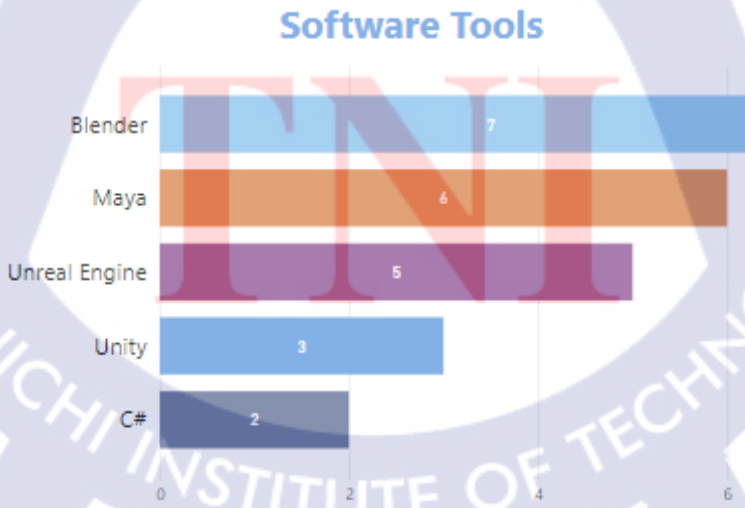
รูปที่ 4.20 ข้อมูลสถิติเงินเดือน Dashboard เทรนด์ตำแหน่งงาน IT หน้า 2

3. สถานที่ปฏิบัติงาน แสดงสถานที่ปฏิบัติงานที่เลือกดูเป็นจำนวนการเปิดรับสมัครงานในแต่ละจังหวัดจากข้อมูลที่เก็บรวบรวมมาแสดงข้อมูลเป็น bubble map ซึ่งแสดงทั้งตำแหน่งจังหวัดบนแผนที่จากจุด bubble และแสดงปริมานจากขนาด bubble ทำให้สามารถวิเคราะห์การกระจายตัวของงานในแต่ละภาค ดังรูปที่ 4.21



รูปที่ 4.21 สถานที่ปฏิบัติงาน

4. Software Tools แสดงถึงจำนวนความต้องการทักษะในการใช้ภาษาโปรแกรมมิ่งหรือเครื่องมือจากข้อมูลการเปิดรับสมัครงานที่เก็บรวบรวมข้อมูลมาตามตำแหน่งงาน IT ที่เลือกดูแสดงข้อมูลเป็นกราฟแท่งเพื่อให้ง่ายต่อการเปรียบเทียบ ดังรูปที่ 4.22

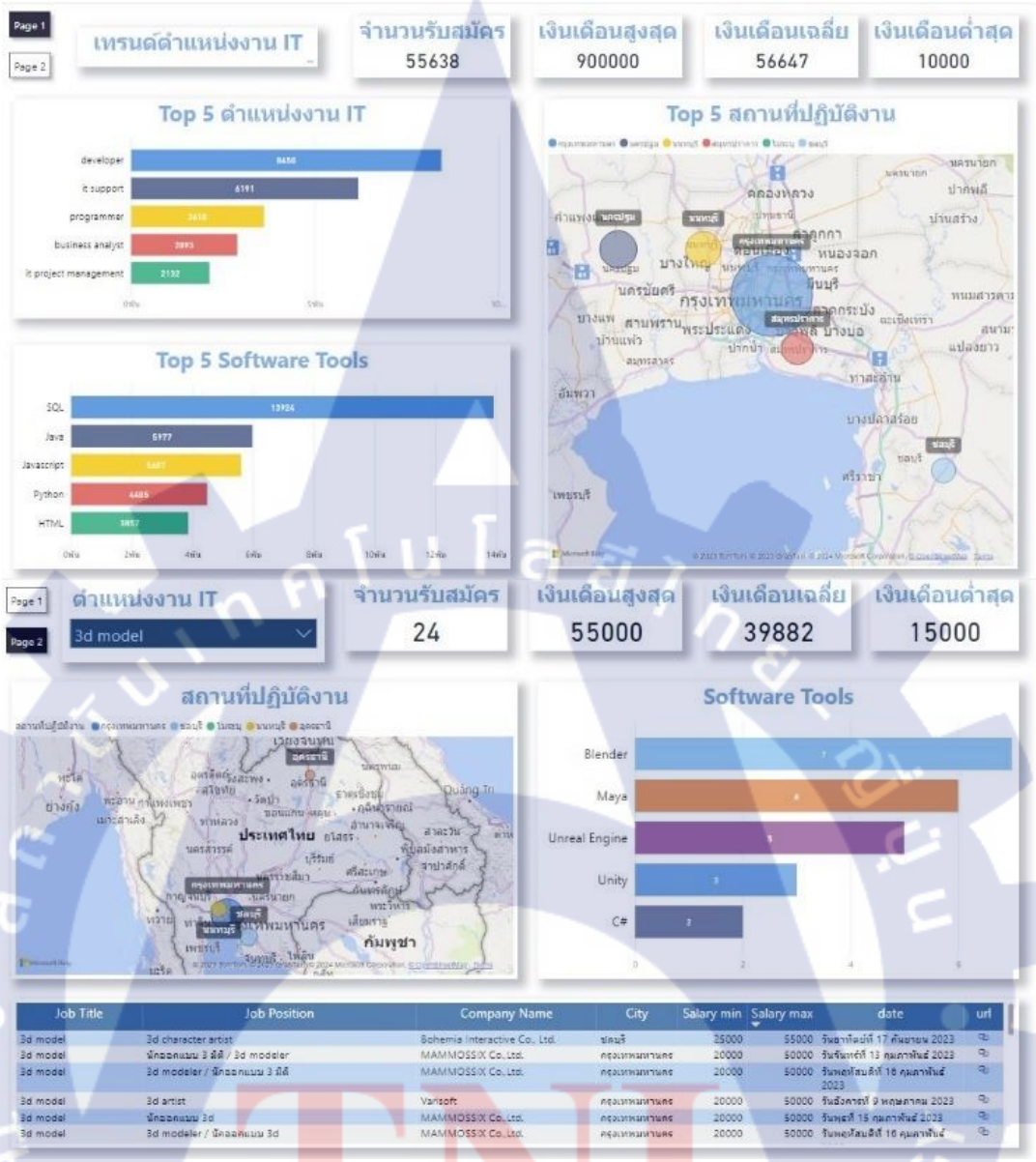


รูปที่ 4.22 Software Tools

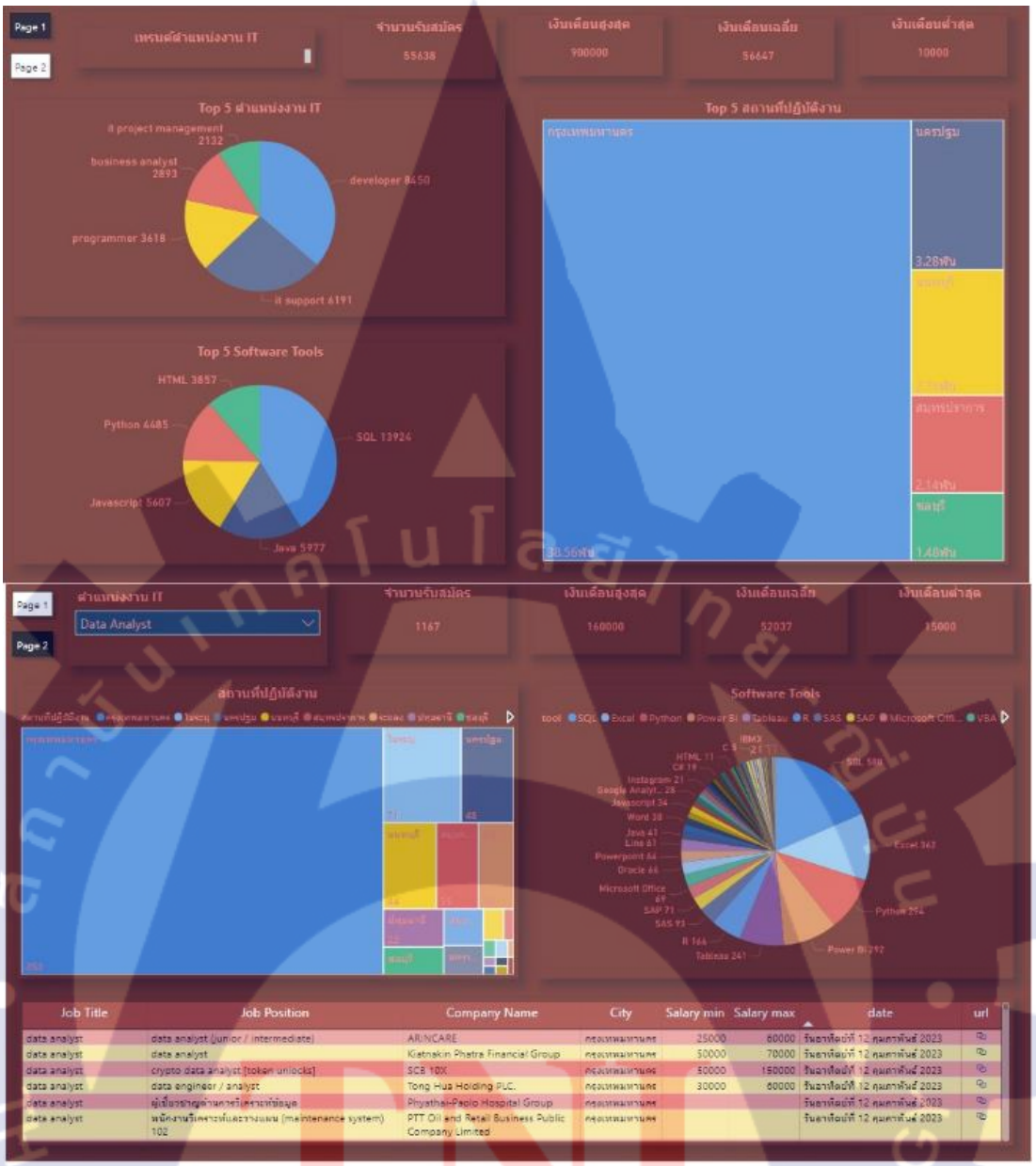
ผู้วิจัยได้สร้าง Dashboard เทรนด์ตำแหน่งงาน IT มี 2 หน้า 2 ชุด เพื่อใช้สำหรับประเมินการออกแบบ Dashboard ด้วยแบบสอบถาม ชุดที่ 1 แบบ A ตามรูปที่ 4.23 และชุดที่ 2 แบบ B ตามรูปที่ 4.24

โดยแบบ A เป็นการออกแบบที่ผู้วิจัยต้องการโดยเน้นความโล่งมองแล้วดูสบายตาจึงเลือกพื้นหลังสีขาว เลือกสีตัวอักษรเป็นสีฟ้าเข้ม เลือกโทนสีของ Dropdown list และตารางเป็นสีฟ้าเพื่อให้ Dashboard ดูออกโทนเย็น ส่วนตัวอักษรให้หัวข้อมีขนาดใหญ่ และรายละเอียดอย่างชื่อตำแหน่ง ทักษะ และจังหวัด มีขนาดใหญ่เท่าที่แสดงได้พอดี ส่วนการเลือกชนิด Data Visualization ให้เป็นกราฟแท่งในหัวข้อ Top 5 ตำแหน่งงาน IT และ Top 5 Software Tools ในหน้าแรก และ Software Tools ในหน้าที่ 2 เนื่องจากข้อมูลเป็นการเปรียบเทียบปริมาณเพื่อจัดอันดับกราฟแท่งแสดงให้เห็นได้ชัดเจน เลือก bubble map ในหัวข้อ Top 5 สถานที่ปฏิบัติงานในหน้าแรก และสถานที่ปฏิบัติงานในหน้าที่ 2 ซึ่ง bubble map แสดงทั้งตำแหน่งจังหวัดบนแผนที่จากจุด bubble และแสดงปริมานจากขนาด bubble ทำให้สามารถวิเคราะห์การกระจายตัวของงานในแต่ละภาค เลือกใช้ card ในการแสดงเงินเดือน และจำนวนการรับสมัคร เลือกใช้ filter รูปแบบ Dropdown list ที่เลือกได้ตัวเลือกเดียว เลือกตารางสำหรับแสดงข้อมูลการรับสมัครงานอย่างชื่อบริษัท ชื่อตำแหน่งงาน เงินเดือน วันที่เก็บข้อมูลมาในแต่ละตำแหน่งงานที่ลงในเว็บไซต์หางานรวมทั้งมีลิ้งเข้าหน้าเว็บไซต์ในตำแหน่งงานนั้น ๆ

ส่วน Dashboard แบบ B ผู้วิจัยจะออกแบบให้ตรงกันข้ามกับแบบ A เพื่อให้เห็นข้อแตกต่างอย่างชัดเจนจึงเลือกพื้นหลังสีแดงเข้ม เลือกสีตัวอักษรเป็นสีแดง เลือกโทนสีของ Dropdown list และตารางเป็นสีแดงเพื่อให้ Dashboard ดูออกโทนร้อน ส่วนตัวอักษรให้หัวข้อมีขนาดเล็ก และรายละเอียดอย่างชื่อตำแหน่ง ทักษะ และจังหวัด มีขนาดเล็ก ส่วนการเลือกชนิด Data Visualization ให้เป็นกราฟวงกลมในหัวข้อ Top 5 ตำแหน่งงาน IT และ Top 5 Software Tools ในหน้าแรก และ Software Tools ในหน้าที่ 2 เลือก tree map ในหัวข้อ Top 5 สถานที่ปฏิบัติงานในหน้าแรก และสถานที่ปฏิบัติงานในหน้าที่ 2 เลือกใช้ card ในการแสดงเงินเดือน และจำนวนการรับสมัครเหมือนกับแบบ A เลือกใช้ filter รูปแบบ Dropdown list ที่เลือกได้ตัวเลือกเดียวเหมือนกับแบบ A เลือกตารางสำหรับแสดงข้อมูลการรับสมัครงานอย่างชื่อบริษัท ชื่อตำแหน่งงาน เงินเดือน วันที่เก็บข้อมูลมาในแต่ละตำแหน่งงานที่ลงในเว็บไซต์หางานรวมทั้งมีลิ้งเข้าหน้าเว็บไซต์ในตำแหน่งงานนั้น ๆ เหมือนกับแบบ A



รูปที่ 4.23 Dashboard เทรนด์ตำแหน่งงาน IT แบบ A



รูปที่ 4.24 Dashboard เทรนด์ตำแหน่งงาน IT แบบ B

4.4 ประสิทธิภาพของระบบ

งานวิจัยนี้ทำการออกแบบ และพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติ ซึ่งการทำงานตามรูปที่ และประสิทธิภาพจะวัดโดยจับระยะเวลาการทำงานแต่ละส่วน มีหัวข้อดังตารางที่ 4.7

ตารางที่ 4.7 ตารางระยะเวลาการทำงานแต่ละส่วน

ขั้นตอนการทำงาน	ระยะเวลาการทำงาน
สกัดข้อมูลเว็บไซต์ jobsblognone	2 นาที 10 วินาที
สกัดข้อมูลเว็บไซต์ jobBBK	36 นาที 49 วินาที
สกัดข้อมูลเว็บไซต์ jobthai	31 นาที 35 วินาที
สกัดข้อมูลเว็บไซต์ jobsDB	20 นาที 53 วินาที
หน่วงเวลา	10 วินาที
รวมตารางข้อมูล	2 วินาที
ทำความสะอาดข้อมูล	11 วินาที
อัพเดทฐานข้อมูลข้อมูล	24 วินาที
อัพเดทข้อมูลใน Dashboard	4 นาที
ระยะเวลาทั้งหมด	1 ชั่วโมง 36 นาที

จากผลการทดสอบแล้วระบบสามารถทำงานอัตโนมัติได้ และใช้เวลาทำงาน ประมาณ 1 ชั่วโมง 36 นาที ในการจัดการกับรายอะเอียดงานเว็บไซต์ jobsblognone 391 งาน เว็บไซต์ jobBBK 2,305 งาน เว็บไซต์ jobthai 1,811 งาน และเว็บไซต์ jobsDB 2,188 งาน รวม 6,695 งาน ทำงานในวันที่ 5/1/2567 ซึ่งคนทั่วไปจะหาเทรนด์ตำแหน่งงาน IT ด้วยตนเองจะต้องใช้เวลามากกว่า 1 วัน

4.5 ความพึงพอใจต่อการใช้แดชบอร์ด

หลังจากสร้าง Dashboard เทรนด์ตำแหน่งงาน IT ทั้ง 2 แบบแล้ว ผู้วิจัยได้ทำการให้ผู้ทดสอบทดลองใช้ Dashboard เทรนด์ตำแหน่งงาน IT ทั้ง 2 แบบ และทำการตอบแบบสอบถามความพึงพอใจต่อการใช้ Dashboard โดยเริ่มทดสอบตั้งแต่วันที่ 18/1/2567 ถึงวันที่ 23/1/2567 ผู้ทดสอบจะมาจาก Facebook กลุ่มหางาน และกลุ่มไอที ต่าง ๆ ทั้งหมด 3 กลุ่ม และกลุ่มไลน์อีก 2 กลุ่ม มีคนเข้ามาตอบแบบสอบถามทั้งหมด 40 คน โดยมีผลการวิเคราะห์ดังนี้

ตารางที่ 4.8 แสดงข้อมูลผู้ทดสอบว่ามีความสนใจในงานสายไอทีหรือทำงานที่เกี่ยวข้องกับไอที

มีความสนใจหรือทำงานที่เกี่ยวข้องกับไอที	ความถี่	ร้อยละ
ใช่	36	90
ไม่ใช่	4	10
รวม	40	100

จากตารางที่ 4.8 พบว่าผู้ทดสอบมีความสนใจในงานสายไอทีหรือทำงานที่เกี่ยวข้องกับไอที 36 คน คิดเป็นร้อยละ 90 และไม่ใช่ 4 คน คิดเป็นร้อยละ 10 ต่อจากนี้ผู้วิจัยจะใช้ข้อมูลเฉพาะผู้ทดสอบมีความสนใจในงานสายไอทีหรือทำงานที่เกี่ยวข้องกับไอที 36 คนเท่านั้น

ตารางที่ 4.9 แสดงข้อมูลทั่วไปของผู้ทดสอบด้านเพศ

เพศ	ความถี่	ร้อยละ
ชาย	27	75
หญิง	9	25
รวม	36	100

จากตารางที่ 4.9 พบว่าผู้ทดสอบเพศชาย 27 คน คิดเป็นร้อยละ 75 และเพศหญิง 9 คน คิดเป็นร้อยละ 25

ตารางที่ 4.10 แสดงข้อมูลทั่วไปของผู้ทดสอบด้านอายุ

อายุ	ความถี่	ร้อยละ
ต่ำกว่าเท่ากับ 18 ปี	1	2.78
19 - 24 ปี	2	5.56
25 - 30 ปี	15	41.67
30 - 40 ปี	10	27.78
มากกว่า 40 ปี	8	22.22
รวม	36	100

จากตารางที่ 4.10 พบว่าผู้ทดสอบมีอายุต่ำกว่าเท่ากับ 18 ปี 1 คน คิดเป็นร้อยละ 2.76 อายุ 19 - 24 ปี 2 คน คิดเป็นร้อยละ 5.56 อายุ 25 - 30 ปี 15 คน คิดเป็นร้อยละ 41.67 อายุ 30 - 40 ปี 10 คน คิดเป็นร้อยละ 27.78 และอายุมากกว่า 40 ปี 8 คน คิดเป็นร้อยละ 22.22

ตารางที่ 4.11 แสดงข้อมูลทั่วไปของผู้ทดสอบด้านระดับการศึกษา

ระดับการศึกษา	**ความถี่**	**ร้อยละ**
มัธยม	1	2.78
ปริญญาตรี	18	50
ปริญญาโท	14	38.89
ปริญญาเอก	3	8.33
รวม	36	100

จากตารางที่ 4.11 พบว่าผู้ทดสอบมีระดับการศึกษา ระดับมัธยม 1 คน คิดเป็นร้อยละ 2.78 ระดับปริญญาตรี 18 คน คิดเป็นร้อยละ 50 ระดับปริญญาโท 14 คน คิดเป็นร้อยละ 38.89 และระดับปริญญาเอก 3 คน คิดเป็นร้อยละ 8.33

ตารางที่ 4.12 แสดงข้อมูลทั่วไปของผู้ทดสอบด้านอาชีพ

อาชีพเกี่ยวข้องกับไอที	**ความถี่**	**ร้อยละ**
ใช่	19	52.78
ไม่ใช่	17	47.22
รวม	36	100

จากตารางที่ 4.12 พบว่าผู้ทดสอบประกอบอาชีพเกี่ยวข้องกับไอที 19 คน คิดเป็นร้อยละ 52.78 และไม่เกี่ยวข้องกับไอที 17 คน คิดเป็นร้อยละ 47.22

จากตารางที่ 4.13 พบว่าผู้ทดสอบมีความคิดเห็นต่อข้อมูลใน Dashboard ว่ามีความต้องการในทุกหัวข้อ เนื่องมาจากผู้ทดสอบมีความคิดเห็นมีความต้องการเกินร้อยละ 50 ของจำนวนทั้งหมดในทุกหัวข้อ

ตารางที่ 4.13 แสดงความคิดเห็นของผู้ทดสอบด้านความต้องการข้อมูลใน Dashboard

ความต้องการข้อมูลใน Dashboard	ความถี่		ร้อยละ	
	ต้องการ	ไม่ต้องการ	ต้องการ	ไม่ต้องการ
จำนวนผู้รับสมัคร	28	8	77.78	22.22
เงินเดือนสูงสุด	33	3	91.67	8.33
เงินเดือนเฉลี่ย	31	5	86.11	13.89
เงินเดือนต่ำสุด	31	5	86.11	13.89
Top 5 ตำแหน่งงาน IT	29	7	80.56	19.44
Top 5 สถานที่ปฏิบัติงาน	28	8	77.78	22.22
Top 5 ทักษะโปรแกรมมิ่งหรือเครื่องมือที่ใช้ในการทำงาน	34	2	94.44	5.56
ข้อมูล เงินเดือน ทักษะโปรแกรมมิ่งหรือเครื่องมือที่ใช้ในการทำงาน และ สถานที่ปฏิบัติงานในแต่ละตำแหน่งงาน	33	3	91.67	8.33

ตารางที่ 4.14 แสดงความคิดเห็นของผู้ทดสอบด้านความเพียงพอของข้อมูลใน Dashboard

ความเพียงพอของข้อมูลใน Dashboard	ความถี่	ร้อยละ
เพียงพอ	18	50
ไม่เพียงพอ	18	50
รวม	36	100

จากตารางที่ 4.14 พบว่าผู้ทดสอบแสดงความคิดเห็นของข้อมูลใน Dashboard ว่าเพียงพอมี 18 คน คิดเป็นร้อยละ 50 และแสดงความคิดเห็นของข้อมูลใน Dashboard ว่าไม่เพียงพอมี 18 คน คิดเป็นร้อยละ 50 เป็นค่าก้ำกึ่งแต่มองว่าไม่เพียงพอเนื่องมาจากผู้ทดสอบให้ข้อเสนอแนะควรเพิ่มเติมข้อมูล ดังนี้

1. ลักษณะ Field Business ที่เจ้าหน้าที่ IT เข้าไปทำงาน เช่น Finance, Bank, Education, Science, Pure Math, AI, Robotic และ System Integrator เป็นต้น

2. อยากได้ข้อมูลหรือสถิติเทรนด์ที่นำ IT มาใช้กับงานทางทันตกรรม เช่น 3D model for dental

3. ช่วงเวลาทำงาน วันทำงาน onsite และ online กี่วัน

4. อยากรู้ trend ของแต่ละสาขาว่าเพิ่มขึ้นหรือลดลงในช่วงเวลาไหนบ้าง period คนสมัครเยอะน้อย หรืออัตราการแข่งกันเยอะน้อย

5. ควรมีข้อมูลประเภทสถานประกอบการ เช่น ฟรีแลนซ์ บริษัทเอกชน และ ราชการ เป็นต้น

6. ต้องเพิ่มการอธิบายคุณสมบัติของผู้ที่จะเข้ารับการศึกษาด้านไอที เพื่อเป็นข้อมูลแก่ผู้ที่สนใจ

7. ควรบอกด้วยว่า skillset แต่ละตำแหน่งประมาณไหน

8. ในส่วนของ Dashboard ควรมีข้อมูลเปรียบเทียบกับช่วงเวลาก่อนหน้านี้ ว่ามีแนวโน้มเพิ่มขึ้น หรือ ลดลงอย่างไรบ้าง

9. ข้อมูลในการรับสมัครน่าจะต้องมีเรื่องประสบการณ์ในการทำงานว่าเงินเดือนเท่านี้โดยส่วนมากรับที่ประสบการณ์กี่ปี เป็นต้น

10. รายได้ต่อชม. / จำนวนวันหรือชม.การทำงานต่อสัปดาห์ / Skill set ที่จำเป็นต้องใช้ในแต่ละตำแหน่งงาน / ประสบการณ์การทำงาน(จำนวนปี) / ระดับชั้นของตำแหน่งงาน (senior/junior) อาจจะเป็นการนำเสนอข้อมูลในหน้าย่อยอีกที ไม่จำเป็นต้องแสดงบน Dashboard ทั้งหมด

11. นอกเหนือจากการแสดงผลในรูปแบบ Interactive overview จำเป็นจะต้องตอบคำถามโดยระบุถึงปัญหา เพื่อนำข้อมูลมาตอบคำถามที่เกิดขึ้นในแต่จุด โดยใช้ปัจจัยอื่น ๆ ในฐานข้อมูลเป็นตัวสนับสนุนปัญหาที่เกิดขึ้น แนวโน้มที่จะเกิดขึ้นในอนาคต (Predict) การสร้างการเปรีบเทียบ Percentage ระหว่างกลุ่มข้อมูล หรือการเปรียบเทียบระหว่าง 2 ปัจจัยว่ามีผลมากน้อยแค่ไหน (Correlation/Linear Regression)

12. จำนวนประสบการณ์ (ปี) ของแต่ละ Job Position

13. ควรบอกทักษะภาษาที่ควรมีด้วย

14. หากมีข้อมูลเพิ่มในส่วนของช่วงระยะเวลาที่ผ่านมา เช่น เดือน หรือ ปี แล้วทำเป็นกราฟให้เห็นแนวโน้มของปริมาณความต้องการของแต่ละตำแหน่ง ก็จะช่วยในการตัดสินใจเลือกงานสายไอทีได้ดียิ่งขึ้น

การประเมินความพึงพอใจ ใช้มาตรวัดเจตคติแบบลิเคิรท (Likert rating scales) 5 ระดับ โดยได้กำหนดคาน้ำหนักของคะแนนที่ใช้ในการตอบแบบสอบถามตามตารางที่ 4.15 และเกณฑ์ในการแปลความหมายของคาคะแนนเฉลี่ยความพึงพอใจของผูตอบแบบสอบถามตามตารางที่ 4.16

ตารางที่ 4.15 ค่าน้ำหนักของคะแนนที่ใช้ในการตอบแบบสอบถาม

ตัวเลือก	คะแนน
มากที่สุด	5
มาก	4
ปานกลาง	3
ค่อนข้างน้อย	2
น้อย	1

ตารางที่ 4.16 เกณฑ์ในการแปลความหมายของคาคะแนนเฉลี่ยความพึงพอใจ

ความหมาย	คะแนนเฉลี่ย
ดีมาก	4.21 - 5.00
ดี	3.41 - 4.20
พอใช้	2.61 - 3.40
ต้องปรับปรุง	1.81 - 2.60
ต้องปรับปรุงเร่งด่วน	ต่ำกว่า 1.80

จากตารางที่ 4.17 ซึ่งได้ทำการคำนวนค่าเฉลี่ยรวมทั้งทำการแปลผลจากเกณฑ์ในการแปลความหมายของคาคะแนนเฉลี่ยความพึงพอใจแล้ว พบว่าผู้ทดสอบส่วนใหญ่แสดงความคิดเห็นต่อการออกแบบใน Dashboard แบบ A เกือบทุกหัวข้ออยู่ในเกณฑ์ ดี ซึ่งเป็นเกณฑ์ที่เกือบมีค่าสูงสุด และมีเพียงหัวข้อ ความสวยงามของ Dashboard ที่อยู่ในเกณฑ์ พอใช้ ซึ่งเป็นเกณฑ์ที่มีค่าอยู่ตรงกลางในค่า 5 ส่วนโดยภาพรวมความพึงพอใจต่อ Dashboard แบบ A พบว่าผู้ทดสอบส่วนใหญ่แสดงความคิดเห็นให้อยู่ในเกณฑ์ ดี ซึ่งเป็นเกณฑ์ที่เกือบมีค่าสูงสุด

ตารางที่ 4.17 แสดงความคิดเห็นของผู้ทดสอบด้านการออกแบบใน Dashboard แบบ A

ความคิดเห็นด้านการออกแบบใน Dashboard แบบ A	**จำนวน**	**ค่าเฉลี่ย**	**ค่าเบี่ยงเบนมาตฐาน**	**แปลผล**
ความเหมาะสมของสีพื้นหลัง และสีตัวอักษรบน Dashboard	36	3.67	1.05	ดี
ความเหมาะสมของขนาดตัวอักษรบน Dashboard	36	3.53	1.04	ดี
ความเหมาะสมของฟอนต์ตัวอักษรบน Dashboard	36	3.56	1.09	ดี
ความเหมาะสมในการวางตำแหน่งของ Visualization (เช่น กราฟหรือแผนภูมิ) ต่าง ๆ บน Dashboard	36	3.72	1.02	ดี
ความเหมาะสมในการเลือกชนิดของ Visualization (เช่น กราฟหรือแผนภูมิ) เพื่อแสดงข้อมูลบน Dashboard	36	3.83	0.93	ดี
Dashboard มีความเข้าใจง่าย	36	3.89	0.97	ดี
ความสวยงามของ Dashboard	36	3.39	1.21	พอใช้
ความต้องการแนะนำให้ผู้อื่นใช้งาน Dashboard	36	3.56	1.07	ดี
ภาพรวมความพึงพอใจต่อ Dashboard	36	3.64	0.98	ดี

จากตารางที่ 4.18 ซึ่งได้ทำการคำนวนค่าเฉลี่ยรวมทั้งทำการแปลผลจากเกณฑ์ในการแปลความหมายของคาคะแนนเฉลี่ยความพึงพอใจแล้ว พบว่าผู้ทดสอบส่วนใหญ่แสดงความคิดเห็นต่อการออกแบบใน Dashboard แบบ B เกือบทุกหัวข้ออยู่ในเกณฑ์ พอใช้ ซึ่งเป็นเกณฑ์ที่มีค่าอยู่ตรงกลางในค่า 5 มีเพียงหัวข้อ ความเหมาะสมของสีพื้นหลังและสีตัวอักษรบน Dashboard ที่อยู่ในเกณฑ์ ต้องปรับปรุง ซึ่งเป็นเกณฑ์ที่มีค่าอยู่รองสุดท้ายในค่า 5 โดยภาพรวมความพึงพอใจต่อ Dashboard แบบ B พบว่าผู้ทดสอบส่วนใหญ่แสดงความคิดเห็นให้อยู่ในเกณฑ์ พอใช้ ซึ่งเป็นเกณฑ์ที่มีค่าอยู่ตรงกลาง

ตารางที่ 4.18 แสดงความคิดเห็นของผู้ทดสอบด้านการออกแบบใน Dashboard แบบ B

ความคิดเห็นด้านการออกแบบใน Dashboard แบบ B	**จำนวน**	**ค่าเฉลี่ย**	**ค่าเบี่ยงเบนมาตฐาน**	**แปลผล**
ความเหมาะสมของสีพื้นหลังและสีตัวอักษรบน Dashboard	36	2.28	1.10	ต้องปรับปรุง
ความเหมาะสมของขนาดตัวอักษรบน Dashboard	36	2.72	1.19	พอใช้
ความเหมาะสมของฟอนต์ตัวอักษรบน Dashboard	36	2.78	1.27	พอใช้
ความเหมาะสมในการวางตำแหน่งของ Visualization (เช่น กราฟหรือแผนภูมิ) ต่าง ๆ บน Dashboard	36	2.86	1.23	พอใช้
ความเหมาะสมในการเลือกชนิดของ Visualization (เช่น กราฟหรือแผนภูมิ) เพื่อแสดงข้อมูลบน Dashboard	36	2.83	1.28	พอใช้
Dashboard มีความเข้าใจง่าย	36	2.83	1.19	พอใช้
ความสวยงามของ Dashboard	36	2.72	1.17	พอใช้
ความต้องการแนะนำให้ผู้อื่นใช้งาน Dashboard	36	2.75	1.19	พอใช้
ภาพรวมความพึงพอใจต่อ Dashboard	36	2.72	1.12	พอใช้

จากการจัดทำแบบสอบถามความพึงพอใจต่อการใช้ Dashboard โดยเริ่มทดสอบตั้งแต่วันที่ 18/1/2567 ถึงวันที่ 23/1/2567 มีผู้เข้าร่วม 40 คน และตอบว่ามีความสนใจในงาน ไอที 36 คนโดยแบบสอบถามประกอบด้วย 2 ส่วน ส่วนแรก คือ ด้านความเพียงพอของข้อมูลใน Dashboard และความต้องการข้อมูลที่แสดงใน Dashboard ซึ่งผู้ทดสอบส่วนใหญ่นั้นต้องการข้อมูลที่แสดงใน Dashboard ส่วนความเพียงพอของข้อมูลคิดเป็นร้อยละ 50 เท่านั้น และมีผู้ทดสอบหลายท่านให้ความต้องการข้อมูลเพิ่มเติม ให้แสดงประเภทอุตสาหกรรมของบริษัทที่รับสมัครงาน ช่วงเวลาทำงาน การแบ่งกลุ่มงานว่าเป็นงานประจำ ราชการ เอกชน หรือ อิสระ ประสบการณ์ทำงาน ทักษะภาษา และต้องการข้อมูลเปรียบเทียบจำนวนที่รับสมัครในแต่ละช่วงเวลา ส่วนที่ 2 คือ ด้านการออกแบบ Dashboard ซึ่งมี 2 แบบ มีแบบ A และ B ทั้ง 2 แบบเป็นการออกแบบที่ตรงกันข้ามกันเพื่อดูว่าแบบ

ไหนคนชอบมากกว่ากัน ซึ่งแบบ A อยู่ในเกณฑ์ ดี แต่หัวข้อความสวยงามอยู่ในเกณฑ์ พอใช้ และแบบ B อยู่ในเกณฑ์ พอใช้ แต่หัวข้อสีพื้นหลัง และสีตัวอักษร อยู่ในเกณฑ์ ต้องปรับปรุง สรุปได้ว่าข้อมูลใน Dashboard เป็นที่ต้องการของผู้ทดสอบแต่ยังไม่เพียงพอ และผู้ทดสอบพึงพอใจในการออกแบบ Dashboard แบบ A มากกว่าแบบ B ซึ่งแบบ A มีตัวอักษรขนาดใหญ่ สีพื้นหลังสว่าง สีตัวอักษรเป็นสีฟ้าเข้ม สีของ Dropdown list และตารางเป็นสีฟ้าเพื่อให้ Dashboard ดูออกโทนเย็น การใช้กราฟแท่งในการแสดงข้อมูลที่มีการเปรียบเทียบ และการใช้ bubble map ในการแสดงข้อมูลสถานที่ปฏิบัติงานมีลักษณะข้อมูลเป็นความถี่ในแต่ละจังหวัดทำให้เห็นทั้งขนาด และการกระจายตัว

บทที่ 5
อภิปรายผลการวิจัย

งานวิจัยนี้เป็นการออกแบบ และพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติรวมทั้งสร้างรายงานการวิเคราะห์ข้อมูลโดยใช้ Data Visualization ต่าง ๆ ในรูปแบบรายงาน Dashboard เพื่อให้ผู้สนใจที่จะเปลี่ยนสายงานมาทำงานสาย IT ได้นำไปใช้ประกอบการตัดสินใจในการเลือกงาน และการตัดสินใจในการศึกษาต่อ โดยมีการสรุปผลวิจัยดังนี้

5.1 สรุปผลการวิจัย

การออกแบบ และพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติรวมทั้งสร้างรายงานการวิเคราะห์ข้อมูลโดยใช้ Data Visualization ต่าง ๆ ในรูปแบบรายงาน Dashboard มีขั้นตอนการทำการดังนี้

1. สกัดข้อมูลเว็บไซต์ jobsblognone
2. สกัดข้อมูลเว็บไซต์ jobBBK
3. สกัดข้อมูลเว็บไซต์ jobthai
4. สกัดข้อมูลเว็บไซต์ jobsDB
5. รวมตารางข้อมูล
6. ทำความสะอาดข้อมูล
7. อัพเดทฐานข้อมูลข้อมูล
8. อัพเดทข้อมูลใน Dashboard

จากการทดสอบระบบทำงานทั้ง 8 ขั้นตอนใช้เวลาประมาณ 1 ชั่วโมง 36 นาที โดยมีเงื่อนไขว่ามีโพสรับสมัครงานรวม 6,695 งาน ระบบเริ่มเก็บข้อมูลตั้งแต่ 12/2/2566 ถึงวันที่ 8/1/2567 โดยชุดข้อมูลที่เริ่มต้นใช้มาจากการสกัดข้อมูลทั้ง 4 เว็บไซต์หางาน ข้อมูลรวมกันมากกว่า 55,638 แถว หลังจากสร้าง Dashboard ตามปัจจัยในการเลือกงานที่ผู้วิจัยทำการศึกษาจากงานวิจัยอื่น ๆ แล้วมี 5 หัวข้อดังนี้

1. จำนวนรับสมัครงานตำแหน่งงาน IT ทั้งหมดที่ได้เก็บรวบรวมมาจากเว็บไซต์หางานซึ่งในตอนนี้มีจำนวน 55,638

2. เงินเดือน แสดงเงินเดือนตำแหน่งงาน IT ซึ่งแบ่งย่อยเป็น 3 หัวข้อ เป็นเงินเดือนสูงสุด 900,000 บาท เงินเดือนเฉลี่ย 56,647 บาท และเงินเดือนต่ำสุด 10,000 บาท

3. Top 5 ตำแหน่งงาน IT แสดงถึง 5 อันดับตำแหน่งงาน IT ที่เปิดรับสมัครมากที่สุดจากที่เก็บรวบรวมข้อมูลมา อันดับที่ 1 Developer อันดับที่ 2 IT Support อันดับที่ 3 Programmer อันดับที่ 4 Business Analyst และอันดับที่ 5 IT Project Management

4. Top 5 สถานที่ปฏิบัติงาน แสดงถึง 5 อันดับจังหวัดที่ปฏิบัติงาน IT นิยมที่สุดจากที่เก็บรวบรวมข้อมูลมา อันดับที่ 1 กรุงเทพ อันดับที่ 2 นครปฐม อันดับที่ 3 นนทบุรี อันดับที่ 4 สมุทรปราการ และอันดับที่ 5 IT ชลบุรี

5. Top 5 Software Tools แสดงถึง 5 อันดับทักษะในการใช้ภาษาโปรแกรมมิ่งหรือเครื่องมือที่ต้องการมากที่สุดจากที่เก็บรวบรวมข้อมูลมา อันดับที่ 1 SQL อันดับที่ 2 Java อันดับที่ 3 JavaScript อันดับที่ 4 Python และอันดับที่ 5 HTML

ความพึงพอใจต่อ Dashboard ผู้ทดสอบที่ได้เข้าร่วมทดลองการใช้งานมีความพึงพอใจต่อ Dashboard อยู่ในเกณฑ์ดี และข้อมูลใน Dashboard มีส่วนช่วยในการตัดสินใจในการเลือกงาน และศึกษาต่อ แต่ยังต้องการข้อมูล ประเภทอุตสาหกรรมของบริษัทที่รับสมัครงาน ช่วงเวลาทำงาน การแบ่งกลุ่มงานว่าเป็นงานประจำ ราชการ เอกชน หรือ อิสระ ประสบการณ์ทำงาน ทักษะภาษา และต้องการข้อมูลเปรียบเทียบจำนวนที่รับสมัครในแต่ละช่วงเวลา เพิ่มเติมเพื่อช่วยให้ตัดสินใจได้ดียิ่งขึ้น

5.2 อภิปรายผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานในประเทศไทยแบบอัตโนมัติ และแสดงผลข้อมูลด้วยภาพ ซึ่งระบบสามารถทำงานตามวัตถุประสงค์ได้เป็นอย่างดีจากการทดสอบระบบใช้เวลาเพียง 1 ชั่วโมง 36 นาที โดยมีเงื่อนไขว่ามีโพสรับสมัครงานรวม 6,695 งาน ระบบช่วยให้สามารถตัดสินใจเลือกงาน IT ได้อย่างรวดเร็วจากการดูข้อมูลใน Dashboard เร็วกว่าการนั่งดูตำแหน่งงานในเว็บไซต์หางานทีละตำแหน่งค่อยตัดสินใจมาก จากที่ได้พัฒนาระบบขั้นตอนสกัดข้อมูลการรับสมัครงานจากเว็บไซต์หางานนั้นไม่ยากที่สุดแต่ลำบากมากที่สุดเนื่องจากใช้เทคนิค web scaping เมื่อเว็บไซต์หางานเกิดมีการเปลี่ยนแปลงจะทำให้ระบบทำงานเกิดข้อผิดพลาด หรือถึงขั้นไม่สามารถทำงานต่อได้

จากการทำวิจัยได้รวบรวมข้อมูลปัจจัยที่ส่งผลในการตัดสินใจในการเลือกงานจาก 4 งานวิจัยสายสังคมมาเปรียบเทียบกับเนื้อหารับสมัครงานแล้วทำการออกแบบการสกัดข้อมูล การรับสมัครงานตำแหน่งงาน IT และวิเคราะห์ข้อมูลสร้าง Dashboard ปัจจัยที่เกี่ยวข้อง และสามารถสกัดข้อมูลออกมาได้ คือ เงินเดือน ลักษณะหน้าที่ที่ได้รับของงาน ชื่อเสียงองค์กร และสถานที่ปฏิบัติงานทั้งหมดเป็นที่ต้องการ ส่วนประสบการณ์การทำงานเป็นปัจจัยที่ไม่สามารถสกัดข้อมูลออกมาได้แต่เป็นที่ต้องการเช่นกัน ซึ่งปัจจัยดังกล่าวมีผลต่อการเลือกอาชีพเป็นไปตามงานวิจัยต่าง ๆ ที่ได้ศึกษามา รวมทั้งมีปัจจัยอื่น ๆ ที่ไม่มีการศึกษาในงานวิจัยที่ผู้วิจัยได้ศึกษามาแต่เป็นที่ต้องการ คือ ประเภทอุตสาหกรรมของบริษัทที่รับสมัครงาน ช่วงเวลาทำงาน การแบ่งกลุ่มงานว่าเป็นงานประจำ ราชการ เอกชน หรือ อิสระ และจำนวนความต้องการในตำแหน่งงานในแต่ละช่วงเวลา

ข้อมูลจากสำนักงานสภานโยบายการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรมแห่งชาติ (สอวช.) ได้แสดงข้อมูลแนวโน้มความต้องการบุคลากรในอุตสาหกรรม และทิศทางนโยบายการพัฒนา

กำลังคนของประเทศ พ.ศ. 2563-2567 มีการคาดการว่าอุตสาหกรรมดิจิทัลโดยภาพรวมจะมีความต้องการบุคลากรในอนาคตอยู่ที่ประมาณ 34,505 ตำแหน่ง แต่ข้อมูลที่ผู้วิจัยเก็บรวบรวมมาได้เป็นของปี พ.ศ. 2566 มีความต้องการ 55,638 ตำแหน่ง ซึ่งมากกว่าที่ สอวช. คาดการไว้เนื่องจากข้อมูลที่ใช้วิเคราะห์เป็นข้อมูลก่อนเกิดการระบาดของโควิด - 19 ซึ่งระบาดไปทั่วโลกตอนต้นปี พ.ศ. 2563 ซึ่งข้อมูลนี้ได้บ่งบอกแล้วว่าการระบาดของโควิด - 19 ทำให้องค์กรต่าง ๆ ได้นำเทคโนโลยีสารสนเทศมาช่วยเพิ่มช่องทางการทำงานในบริษัท และการบริการลูกค้าขึ้นอย่างมากจนทำให้ปัจจุบันปริมาณการขยายตัวของตลาดแรงงานอาชีพสาย IT มีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง จากข้อมูลที่ผู้วิจัยได้เก็บรวบรวมมา 5 อันดับทักษะในการใช้ภาษาโปรแกรมมิ่งหรือเครื่องมือที่ต้องการ มีความสอดคล้องกับ 5 อันดับตำแหน่งงานที่ต้องการ ตำแหน่งงาน คือ Developer และ Programmer ซึ่งใช้ SQL Java JavaScript Python และ HTML ในการพัฒนาเว็บไซต์ หรือ แอปพลิเคชัน ส่วน Business Analyst นั้นใช้ SQL และ Python ในการจัดการข้อมูล ส่วนจังหวัดที่ปฏิบัติงาน IT นิยมที่สุด คือ กรุงเทพ ซึ่งบริษัทที่ทำงานในออฟฟิศส่วนใหญ่กระจุกตัวในกรุงเทพเช่นกัน เงินเดือนมากที่สุดมากถึง 900,000 บาทเป็นตำแหน่ง Full-Stack Web Developer และ Network Engineer ซึ่งข้อมูลรับสมัครเป็นช่วง 50,000 – 900,000 บาทเมื่อสมัครจริงต้องมีการต่อรองเงินเดือนตามความสามารถเงินเดือนจริงอาจจะไม่ถึง 900,000 บาท

การออกแบบ Dashboard ได้ผลตอบรับจากผู้ทดสอบเป็นไปตามที่ผู้วิจัยศึกษาว่า Dashboard ที่ดีต้องตัวอักษรมีสีเข้มและใหญ่อ่านได้ชัดเจน สีพื้นหลังสว่างโทนเย็นมองสบายตา กราฟเปรียบเทียบข้อมูลได้รวดเร็ว และความสวยงามไม่ได้สำคัญมากนัก กราฟแท่งเหมาะกับข้อมูลที่มีการเปรียบเทียบ ส่วนกราฟวงกลม และ tree map ไม่เหมาะกับข้อมูลที่มีการเปรียบเทียบที่มีกลุ่มข้อมูลจำนวนมาก และค่าใกล้เคียงกัน bubble map ใช้แสดงข้อมูลทั้งตำแหน่งสถานที่ และปริมาณจากขนาด bubble ได้แต่ถ้าข้อมูลมีปรมาณใกล้เคียงกันเกินไปจะทำให้เปรียบเทียบข้อมูลลำบาก ซึ่งสามารถแก้ได้โดยใช้เฉดสีเข้มอ่อนในการช่วยแสดงปริมาณ

5.3 ข้อจำกัดของการวิจัย

งานวิจัยนี้มีข้อจำกัดในหลายอย่าง ดังนี้

1. ด้านข้อมูลเนื่องมาจากบริษัทรับสมัครบางแห่งกรอกรายละเอียดงานได้ไม่ครบถ้วนหรือถูกต้อง เช่น บางประกาศรับสมัครงานไม่มีข้อมูลเงินเดือน หรือลงข้อมูลเงินเดือนว่าตามที่ตกลงทำให้ข้อมูลเงินเดือนน้อยยากที่จะประเมิณเงินเดือนได้แม่นยำ

2. HR ที่ไม่เข้าใจตำแหน่งงานที่บริษัทต้องการจึงเหมารวมทุกตำแหน่งเป็นเจ้าหน้าที่ IT หรือตำแหน่งอื่น ซึ่งยังเป็นส่วนน้อยแต่ก็ทำให้ข้อมูลมีความผิดเพี้ยนได้เพราะระบบนี้ยึดชื่อตำแหน่งงานเป็นหลักทำให้ทักษะ และเครื่องมือดูไม่สอดคล้องกับตำแหน่งงาน

3. บางตำแหน่งงานไม่ใช่ตำแหน่ง IT แต่มีปนมากับระบบคัดกรองของเว็บไซต์หางาน

4. ข้อมูลวันประกาศรับสมัครเนื่องด้วยเว็บไซต์หางานแสดงข้อมูลวันที่รับสมัครเป็นวันที่กับเวลาที่ผ่านไปหลังจากประกาศถ้าเป็นวันที่ประกาศ ทำให้ยากต่อการเก็บข้อมูลวันประกาศ

5. เว็บไซต์หางานไม่มีรูปแบบ และการแยกหัวข้อในการประกาศ ทำให้ยากต่อการสกัดข้อมูลที่ต้องการ จึงยากต่อการพัฒนาให้มีข้อมูลที่ผู้ทดสอบให้คำแนะนำว่าต้องการ

6. ระบบอัพเดทข้อมูลใน Dashboard ใช้ RPA อย่าง Power Automate สั่งทำงานโดยให้เมาส์คลิก ทำให้ต้องระวังไม่ให้มีอะไรไปเลื่อนเมาส์ เนื่องจากใช้การหน่วงเวลาก่อนคลิกปุ่มถ้าเกิดโปรแกรมทำงานเสร็จล่าช้าจะทำให้ flow error ได้

7. Power Automate Desktop ใช้งานฟรีไม่สามารถตั้งเวลาให้ flow ทำงานเองได้จึงต้องใช้ทางอ้อมในการตั้งเวลา

8. เว็บไซต์หางานเกิดมีการเปลี่ยนแปลงจะทำให้ระบบทำงานเกิดข้อผิดพลาด หรือถึงขั้นไม่สามารถทำงานต่อได้

5.4 ข้อเสนอแนะ

งานวิจัยนี้ยังมีข้อมูลบางส่วนที่ผู้ทดสอบให้คำแนะนำว่าควรมีข้อมูล ประเภทอุตสาหกรรมของบริษัทที่รับสมัครงาน ช่วงเวลาทำงาน การแบ่งกลุ่มงานว่าเป็นงานประจำ ราชการ เอกชน หรืออิสระ ประสบการณ์ทำงาน ทักษะภาษา และต้องการข้อมูลเปรียบเทียบจำนวนที่รับสมัครในแต่ละช่วงเวลา เพิ่มเติมเพื่อช่วยให้ตัดสินใจได้ดียิ่งขึ้น พร้อมทั้งยังนำข้อมูลเหล่านี้ไปศึกษาปัจจัยที่ส่งผลต่อการเลือกงานเพิ่มเติม เพื่อใช้สร้าง Dashboard ที่มีคุณภาพดีกว่านี้ได้ ในการใช้ RPA สร้างให้ระบบทำงานอัตโนมัติโดยเฉพาะขั้นตอนอัพเดทข้อมูลให้กับ Dashboard ผู้วิจัยใช้การหน่วงเวลาเพื่อให้แต่ละขั้นตอนทำงานสำเร็จ ซึ่งถ้าปริมาณข้อมูลมีมากขึ้นการหน่วงเวลาจะไม่พอทำให้ระบบทำงานผิดพลาดได้จึงต้องพัฒนาแก้ไขส่วนนี้ให้ทำงานได้กับระบบ refresh ข้อมูลของ power bi service และงานวิจัยนี้สามารถขยายขอบเขตจากแค่งาน IT เป็นงานทุกประเภทได้ ซึ่งจะช่วยให้ทั้งนักเรียนสามารถเตรียมตัวศึกษาต่อได้ตรงความต้องการของตลาดแรงงาน และบางคนที่ต้องการเปลี่ยนงาน

บรรณานุกรม

บรรณานุกรม

[1] กองสถิติพยากรณ์ สำนักงานสถิติแห่งชาติ, *ตัวชี้วัดด้านเทคโนโลยีสารสนเทศ และการสื่อสารของ ประเทศไทย พ.ศ. 2562*, กรุงเทพฯ: สำนักงานสถิติแห่งชาติ, 2564.

[2] สำนักงานสภานโยบายการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรมแห่งชาติ, "*แนวโน้มความต้องการบุคลากรในอุตสาหกรรมเป้าหมายแห่งอนาคต (New S-Curve) และทิศทางนโยบายการพัฒนากำลังคนของประเทศ พ.ศ.2563 - 2567*," กรุงเทพฯ: พริ้นเอเบิ้ล, 2562.

[3] วิโรจน์ อรุณมานะกุล และคณะ, การสร้างคลังข้อมูลภาษาไทยแบบสั่งสม และการสกัดหาคำเขียนผิดอักขรวิธี และคำปรากฏใหม่, (รายงานวิจัยฉบับสมบูรณ์), กรุงเทพฯ: สำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) และสำนักงานกองทุนสนับสนุนการวิจัย (สกว.), 2559.

[4] อาทิตย์ สิทธิบรรเจิด, *การนำ Visualization ไปประยุกต์ใช้กับการแสดงข้อมูลบนระบบ Knowledge Management*, สารนิพนธ์ วท.บ (เทคโนโลยีสารสนเทศและวิทยาการข้อมูล), มหาวิทยาลัยกรุงเทพ, กรุงเทพฯ, 2552.

[5] พรทวี เถื่อนคำแสน และ บุญญรัตน์ สัมพันธ์วัฒนชัย, " ปัจจัยที่มีผลต่อแรงจูงใจในการตัดสินใจเลือกทำงานของพนักงานระดับปฏิบัติการในนิคมอุตสาหกรรมเหมราช อีสเทิร์นซีบอร์ด จังหวัดระยอง," *วารสารบริหารธุรกิจเทคโนโลยีมหานคร*, ปีที่ 13, ฉบับที่ 1, หน้า 1 - 23, มิถุนายน – กรกฎาคม 2559.

[6] พัชรินทร์ ชุติพงศ์รุ่งโรจน์, " *ปัจจัยที่มีผลต่อการตัดสินใจเลือกเข้าทำงานกับสำนักงานสอบบัญชีขนาดใหญ่ 4 แห่ง*,"สารนิพนธ์ บธ.ม. (พาณิชยศาสตร์ และการบัญชี), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2558.

[7] พรพิมล ชราศักดิ์, " *พฤติกรรมการเลือกประกอบอาชีพของคนว่างงาน กรมการจัดหางานกรุงเทพมหานครพื้นที่ 9*," [Online]. Available: https://mmm.ru.ac.th/MMM/IS/mlw12/6114961048.pdf [Accessed: 18 กุมภาพันธ์ 2565].

[8] ณัฐกร มั่นจิตต์ และ กตัญญู หิรัญญสมบูรณ์, "แรงจูงใจของแรงงานไทยที่ต้องการไปทำงานในสาธารณรัฐเกาหลีอย่างถูกกฎหมาย," *วารสารการบริหาร และจัดการ*, ปีที่ 10, ฉบับที่ 2, หน้า 14 - 27, เดือนกรกฎาคม – ธันวาคม 2563.

[9] กระทรวงแรงงาน, " *การตัดสินใจเลือกอาชีพ*," [Online]. Available: https://lb.mol.go.th/คนหางาน-คนว่างงาน/แนะแนวอาชีพ/การตัดสินใจเลือกอาชีพ [Accessed: 18 กุมภาพันธ์ 2565].

[10] ภูรดา จันทรัฐ, *การวัดระดับเนื้อหาเว็บไซต์การท่องเที่ยวตามความต้องการของนักท่องเที่ยวโดยการวิเคราะห์ข้อความในเว็บไซต์*, สารนิพนธ์ วท.บ. (เทคโนโลยีสารสนเทศ), สถาบันเทคโนโลยีไทย – ญี่ปุ่น, กรุงเทพฯ, 2561.

[11] สุพัฒนวรี ทิพย์เจริญ, "การใช้เทคนิคการระบุชื่อเฉพาะในการสกัดสารสนเทศเพื่อใช้สร้างฐานความรู้เฉพาะด้าน," FEU Academic Review, ปีที่ 2, ฉบับที่ 2, หน้า 5 - 15, ธันวาคม 2551 - พฤษภาคม 2552.

[12] R. B. Mbah, M. Rege, and B. Misra , "Discovering job market trends with text analytics," *2017 International Conference on Information Technology (ICIT),* Bhubaneswar, India, December 2017, pp. 137 - 142.

[13] Microsoft,"*ชนิดการแสดงภาพใน Power BI,*"[Online].Available:https://learn.microsoft.com/th-th/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a [Accessed: March 11, 2022].

[14] Microsoft Azure, "*2022 Gartner® Magic Quadrant™ for Analytics and Business Intelligence Platforms,*" [Online]. Available: https://info.microsoft.com/ww-landing-2022-gartner-mq-report-on-bi-and-analytics-platforms.html?LCID=EN-US [Accessed: February 15, 2022].

[15] 9expert, "*Power BI คืออะไร ?,*"[Online].Available:https://www.9experttraining.com/articles/power-bi-คืออะไร [Accessed: February 15, 2022].

[16] Microsoft,"*What is Power BI ?,*"[Online].Available:https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview[Accessed: February 15, 2022].

[17] วรัญญา แก้วช่วงศรี, "*การศึกษาปัจจัยที่ส่งผลกับผลสัมฤทธิ์จากการใช้งานเทคโนโลยีธุรกิจอัจฉริยะของธุรกิจ SMEs จากโครงการพัฒนาศักยภาพ SMEs ด้วยระบบ Digital ปี 2559,*" สารนิพนธ์ วท.ม.(นโยบายและการบริหารเทคโนโลยีสารสนเทศ), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2560.

[18] Microsoft,"*Data sources for the Power BI service,*"[Online].Available: https://learn.microsoft.com/en-us/power-bi/connect-data/service-get-data [Accessed: March 11, 2022].

[19] Microsoft," *What is a star schema?,*"[Online].Available: https://learn.microsoft.com/en-us/power-bi/guidance/star-schema [Accessed: March 11, 2022].

[20] Microsoft,“ *QuickStart: การเรียนรู้ข้อมูลพื้นฐานของ DAX ใน 30 นาที,”*[Online]. Available: https://support.microsoft.com/th-th/office/quickstart-การเรียนรู้ข้อมูลพื้นฐานของ-dax-ใน-30-นาที-51744643-c2a5-436a-bdf6-c895762bec1a [Accessed: 11 มีนาคม 2566].

[21] สุรจิต ภูมิคง, ธราอั่งสกุล, และจิติมนต์อั่งสกุล, “วิธีการสกัดข่าวด้านความปลอดภัย,” *Suranaree Journal of Social Science*, ปีที่ 2, ฉบับที่ 7,หน้า 79 - 97, ธันวาคม 2556.

[22] จักรินทร์ สันติรัตนภักดี, “การออกแบบ และพัฒนาระบบสกัดข้อมูลกึ่งมีโครงสร้างตามกระบวนการอีแอลที และค้นคืนสารสนเทศด้วยการนำเสนอภาพข้อมูล กรณีศึกษา: รายงานอุบัติเหตุทางถนนรายใหญ่บนเว็บไซต์,” *วารสารวิชาการสมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย (สสอท.)*, ปีที่ 9, ฉบับที่ 2,หน้า 17 - 34, กรกฎาคม – ธันวาคม 2563.

[23] A. Phaphuangwittayakul et al., “Analysis of skill demand in Thai labor market from online jobs recruitments websites,” *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Nakhonpathom, Thailand, July 11-13, 2018, pp. 1 - 5.

[24] D. M. Y. Lee et al., “A Social network analysis of jobs and skills,” *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, December 10-13, 2020, pp. 5747 - 5749.

ภาคผนวก

ภาคผนวก ก.

Code ระบบ

Code ระบบ

```
def jobthai():

    from bs4 import BeautifulSoup
    import requests
    import pandas as pd

    print('Start Scraping jobthai')
    print('Start Load Web => jobthai')
    headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
                Chrome/56.0.2924.76 Safari/537.36'}
    url_job_it = 'https://www.jobthai.com/หางาน/งานคอมพิวเตอร์-ไอที/'
    response_url_job_it = requests.get(url_job_it,headers=headers)
    soup_url_job_it = BeautifulSoup(response_url_job_it.text, 'lxml')
    print('Finish Load Web => jobthai')

    print('Start Find Pagination_last_page')
    pagination_last_page = soup_url_job_it.find('a', class_='last-item')
    link = pagination_last_page.get('href')
    number_last_page = link[24:]
    number_last_page_int=int(number_last_page)
    print('Last Page = '+str(number_last_page_int))
    print('Finish Find Pagination_last_page')

    print('Start Creat Number all page')
    number_all_url_job_it = []
    for i in range (1,number_last_page_int+1):
      i = str(i)
      number_all_url_job_it.append(i)
    print('Finish Creat Number all page')

    print('Start Creat URL jobthai Category It')
    all_url_job_it = []
    for i in number_all_url_job_it:
      url = url_job_it+i
      all_url_job_it.append(url)
    print('Total URL jobthai Category It = '+str(len(all_url_job_it)))
    print('Finish Creat URL jobthai Category It')

    print('Start Load Web => URL jobthai Category It')
    soup_all_url_job_it = []
    for url in all_url_job_it:
```

```
    response = requests.get(url,headers=headers)
    soup = BeautifulSoup(response.text, 'lxml')
    soup_all_url_job_it.append(soup)
print('Total URL jobthai Category It = '+str(len(soup_all_url_job_it)))
print('Finish Load Web => URL jobthai Category It')

print('Start Find URL Position Work')
id_job = []
for soup in soup_all_url_job_it:
    page_all_job_it = soup.find_all('div', class_='ant-row msklqa-8 crLDlD')
    for page_all_job_it_0 in page_all_job_it:
        for page_all_job_it_1 in page_all_job_it_0:
            link = page_all_job_it_1.get('href')
            id = link[15:]
            id_job.append(id)
url_detail_job = 'https://www.jobthai.com/th/job'
url_all_detail_job = []
for id in id_job:
    url = url_detail_job + id
    url_all_detail_job.append(url)
print('Total URL Position Work = '+str(len(url_all_detail_job)))
print('Finish Find URL Position Work')

print('Start Scaping Text')
count = 1
detail_job = []
for url in url_all_detail_job:
    response = requests.get(url,headers=headers)
    soup = BeautifulSoup(response.text,'lxml')
    page = soup.find_all('div', class_='sc-1k7b98y-0 giRfNh')
    for i in page:
        try:
            Company_Name = i.find('div',class_='sc-10ooa77-0 DPLIo').text.strip('\n')
        except:
            try:
                Company_Name = i.find('div',class_='xagz98-2 kqNTeF').text.strip('\n')
            except:
                Company_Name = ''
        try:
            Job_Position = i.find('h1',class_='sc-1048v4y-6 hqeKVP').text.strip('\n')
        except:
            Job_Position = ''
        try:
            Company_Address = i.find('h3',class_='ohgq7e-0 bMVspY').text.strip('\n')
        except:
```

```
            Company_Address = ''
        try:
            Salary = i.find('div',class_='ant-col ant-col-xs-24 ant-col-xs-order-2 ant-col-sm-24 ant-col-sm-order-2 ant-col-md-
                    24 ant-col-md-order-3 ant-col-lg-24 ant-col-lg-order-3 ant-col-xl-24 ant-col-xl-order-3 ant-col-xxl-24 ant-
                    col-xxl-order-3').text.strip('\n')
        except:
            Salary = ''
        try:
            Job_Description = i.find('div',class_='o2xg4u-0 iSkyQt').text.replace('\n',' ').strip(' ')
        except:
            Job_Description = ''
        try:
            Candidate_Qualifications = i.find('div',class_='jltwsh-0 icDWk').text.strip('\n')
        except:
            Candidate_Qualifications = ''
        detail_job.append([url,Company_Name,Job_Position,Company_Address,Salary,Job_Description,Candidate_
                        Qualifications])
        count+=1
        print("Number of detail job downloaded = "+ str(count), end='\r')
    print('Total Scraping Position Work = '+str(len(detail_job)))
    print('Finish Scaping Text')

    import datetime
    currentdate=datetime.date.today()
    Currentdate=str(currentdate)

    print('Start Save To File Excel')
    column_names = ['url','Company Name','Job Position','Company Address','Salary','Job Description','Candidate
                    Qualifications']
    df = pd.DataFrame(detail_job, columns=column_names)
    df['date'] = Currentdate
    df.index = df.index + 1

    path_file = 'C:/Code_IS/code/scraping/'+Currentdate+'Jobthai.xlsx'
    writer = pd.ExcelWriter(path_file)
    df.to_excel(writer,'MySheet1',index=False)
    writer.save()

    print('Finish Save To File Excel')

def jobsdb():

    from bs4 import BeautifulSoup
    import requests
    import pandas as pd
```

```
print('Start Scraping jobsdb')
print('Start Load Web => jobsdb')
url_job_it = 'https://th.jobsdb.com/th/th/jobs/งานไอที/'
response_url_job_it = requests.get(url_job_it)
soup_url_job_it = BeautifulSoup(response_url_job_it.text, 'lxml')
print('Finish Load Web => jobsdb')

print('Start Find Pagination_last_page')
pagination_last_page = soup_url_job_it.find('option', attrs={'data-automation' : 'pagination-dropdown-option-
                                                null'}).next_sibling.text
number_last_page_int = int(pagination_last_page)
print('Last Page = '+str(number_last_page_int))
print('Finish Find Pagination_last_page')

print('Start Creat Number all page')
number_all_url_job_it = []
for i in range (1,number_last_page_int+1):
    i = str(i)
    number_all_url_job_it.append(i)
print('Finish Creat Number all page')

print('Start Creat URL jobsdb Category It')
all_url_job_it = []
for i in number_all_url_job_it:
    url = url_job_it+i
    all_url_job_it.append(url)
print('Total URL jobsdb Category It = '+str(len(all_url_job_it)))
print('Finish Creat URL jobsdb Category It')

print('Start Load Web => URL jobsdb Category It')
soup_all_url_job_it = []
for url in all_url_job_it:
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    soup_all_url_job_it.append(soup)
print('Total URL jobsdb Category It = '+str(len(soup_all_url_job_it)))
print('Finish Load Web => URL jobsdb Category It')

print('Start Find URL Position Work')
url_all_detail_job = []
for soup in soup_all_url_job_it:
    page_all_job_it = soup.find_all('a', attrs = {'rel' : 'nofollow noopener noreferrer'})
    for page_all_job_it_0 in page_all_job_it:
        link = 'https://th.jobsdb.com'+page_all_job_it_0.get('href')
```

```
        url_all_detail_job.append(link)
print('Total URL Position Work = '+str(len(url_all_detail_job)))
print('Finish Find URL Position Work')

print('Start Scaping Text')
count = 1
detail_job = []
for url in url_all_detail_job:
    response = requests.get(url)
    soup = BeautifulSoup(response.text,'lxml')
    page = soup.find_all('div', attrs= {'id':'contentContainer'})
    for i in page:
        try:
            Company_Name = i.find('span',class_='z1s6m00 _1hbhsw64y y44q7i0 y44q7i2 y44q7i21 _1d0g9qk4
                                y44q7ia').text.strip('\n')
        except:
            Company_Name = ''
        try:
            Job_Position = i.find('h1',class_='z1s6m00 _1hbhsw64y y44q7i0 y44q7il _1d0g9qk4 y44q7is
                                y44q7i21').text.strip('\n')
        except:
            Job_Position = ''
        try:
            Company_Address = i.find('span',class_='z1s6m00 _1hbhsw64y y44q7i0 y44q7i1 y44q7i21 y44q7ii').text
        except:
            Company_Address = ''
        try:
            Salary = i.find('div',class_='z1s6m00 _5135ge0 _5135ge3').text
        except:
            Salary = ''
        try:
            try:
                Job_Description_ = i.find('div',attrs={'data-automation':'job-details-job-highlights'}).text.strip('\n')
                Job_Description__ = i.find('div',attrs={'data-automation':'jobDescription'}).text.strip('\n')
                Job_Description = Job_Description_ +' รายละเอียดงาน '+ Job_Description__
            except:
                Job_Description = i.find('h4',string = 'รายละเอียดงาน').parent.parent.parent.text.strip('\n')

        except:
            Job_Description = ''
        try:
            Candidate_Qualifications = i.find('h4',string = 'รายละเอียดเพิ่มเติม').parent.parent.parent.text.strip('\n')
        except:
            Candidate_Qualifications = ''
        detail_job.append([url,Company_Name,Job_Position,Company_Address,Salary,Job_Description,Candidate_
```

```
                                Qualifications])
        count+=1
        print("Number of detail job downloaded = "+ str(count), end='\r')
    print('Total Scraping Position Work = '+str(len(detail_job)))
    print('Finish Scaping Text')

    import datetime
    currentdate=datetime.date.today()
    Currentdate=str(currentdate)

    print('Start Save To File Excel')
    column_names = ['url','Company Name','Job Position','Company Address','Salary','Job Description','Candidate
                        Qualifications']
    df = pd.DataFrame(detail_job, columns=column_names)
    df['date'] = Currentdate
    df.index = df.index + 1

    path_file = 'C:/Code_IS/code/scraping/'+Currentdate+'JobsDB.xlsx'
    writer = pd.ExcelWriter(path_file)
    df.to_excel(writer,'MySheet1',index=False)
    writer.save()

    print('Finish Save To File Excel')

def jobsblognone():

    from bs4 import BeautifulSoup
    import requests
    import pandas as pd
    import regex as re
    import math

    print('Start Scraping jobs.blognone')
    print('Start Load Web => jobs.blognone')
    url_job_it = 'https://jobs.blognone.com/search?page='
    response_url_job_it = requests.get(url_job_it)
    soup_url_job_it = BeautifulSoup(response_url_job_it.text, 'lxml')
    print('Finish Load Web => jobs.blognone')

    print('Start Find Pagination_last_page')
    pagination_last_page = soup_url_job_it.find('h3', class_='text-uppercase font-weight-normal mb-4 d-none d-md-
                                                block').text
    total_job = re.findall('\d',pagination_last_page)
    total_job_str = ''.join(total_job)
    total_job_int=int(total_job_str)
```

```
number_last_page_int = math.ceil(total_job_int/13)
print('Total Job = '+str(total_job_int))
print('Last Page = '+str(number_last_page_int))
print('Finish Find Pagination_last_page')

print('Start Creat Number all page')
number_all_url_job_it = []
for i in range (1,number_last_page_int+1):
    i = str(i)
    number_all_url_job_it.append(i)
print('Finish Creat Number all page')

print('Start Creat URL jobs.blognone Category It')
all_url_job_it = []
for i in number_all_url_job_it:
    url = url_job_it+i
    all_url_job_it.append(url)
print('Total URL jobs.blognone Category It = '+str(len(all_url_job_it)))
print('Finish Creat URL jobs.blognone Category It')

print('Start Load Web => URL jobs.blognone Category It')
soup_all_url_job_it = []
for url in all_url_job_it:
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    soup_all_url_job_it.append(soup)
print('Total URL jobs.blognone Category It = '+str(len(soup_all_url_job_it)))
print('Finish Load Web => URL jobs.blognone Category It')

print('Start Find URL Position Work')
url_all_detail_job = []
for soup in soup_all_url_job_it:
    page_all_job_it = soup.find_all('div',class_='position-relative d-flex')
    for page_all_job_it_0 in page_all_job_it:
        link = 'https://jobs.blognone.com'+page_all_job_it_0.parent.get('href')
        url_all_detail_job.append(link)
url_all_detail_job_set =set(url_all_detail_job)
print('Total URL Position Work = '+str(len(url_all_detail_job_set)))
print('Finish Find URL Position Work')

print('Start Scaping Text')
count = 1
detail_job = []
for url in url_all_detail_job_set:
    response = requests.get(url)
```

```
    soup = BeautifulSoup(response.text,'lxml')
    page = soup.find_all('div', class_='card mt-3')
    for i in page:
        try:
            Company_Name = i.find('span',class_='css-7sjuw3').text.strip('\n')
        except:
            Company_Name = ''
        try:
            Job_Position = i.find('h1',class_='css-htp84t').text.strip('\n')
        except:
            Job_Position = ''
        try:
            Company_Address = i.find(string=' - ').parent.text.strip('\n')
        except:
            Company_Address = ''
        try:
            Salary = i.find('div',class_='text-success css-1msjh1x').text.strip('\n')
        except:
            Salary = ''
        try:
            Job_Description = i.find('div',class_='col-md-8 css-1wekrze').text.strip('\n')
        except:
            Job_Description = ''
        try:
            Candidate_Qualifications = i.find('div',class_='col-md-8 css-1wekrze').text.strip('\n')
        except:
            Candidate_Qualifications = ''
        detail_job.append([url,Company_Name,Job_Position,Company_Address,Salary,Job_Description,Candidate_
                                Qualifications])
        count+=1
        print("Number of detail job downloaded = "+ str(count), end='\r')
print('Total Scraping Position Work = '+str(len(detail_job)))
print('Finish Scaping Text')

import datetime
currentdate=datetime.date.today()
Currentdate=str(currentdate)

print('Start Save To File Excel')
column_names = ['url','Company Name','Job Position','Company Address','Salary','Job Description','Candidate
                    Qualifications']
df = pd.DataFrame(detail_job, columns=column_names)
df['date'] = Currentdate
df.index = df.index + 1
```

```
    path_file = 'C:/Code_IS/code/scraping/'+Currentdate+'JobBlognone.xlsx'
    writer = pd.ExcelWriter(path_file)
    df.to_excel(writer,'MySheet1',index=False)
    writer.save()

    print('Finish Save To File Excel')

def jobBBK():

    from bs4 import BeautifulSoup
    import requests
    import regex as re
    import math
    import pandas as pd

    print('Start Scraping Jobbkk')
    print('Start Load Web => Jobbkk')
    url_job_it = 'https://www.jobbkk.com/หางาน/ไอที/'
    response_url_job_it = requests.get(url_job_it)
    soup_url_job_it = BeautifulSoup(response_url_job_it.text, 'lxml')
    print('Finish Load Web => Jobbkk')

    print('Start Find Pagination_last_page')
    pagination_last_page = soup_url_job_it.find('li', class_='col-md-4 col-sm-4 col-xs-6 search-job-number').text
    total_job = re.findall('\d',pagination_last_page)
    total_job_str = ''.join(total_job)
    total_job_int=int(total_job_str)
    number_last_page_int = math.ceil(total_job_int/25)
    print('Total Job = ' + str(total_job_int))
    print('Last Page = ' + str(number_last_page_int))
    print('Finish Find Pagination_last_page')

    print('Start Creat Number all page')
    number_all_url_job_it = []
    for i in range (1,number_last_page_int+1):
        i = str(i)
        number_all_url_job_it.append(i)
    print('Finish Creat Number all page')

    print('Start Creat URL Jobbkk Category It')
    all_url_job_it = []
    for i in number_all_url_job_it:
        url = url_job_it+i
        all_url_job_it.append(url)
    print('Total URL Jobkk Category It = '+str(len(all_url_job_it)))
```

```
print('Finish Creat URL Jobbkk Category It')

print('Start Load Web => URL Jobbkk Category It')
count = 1
soup_all_url_job_it = []
for url in all_url_job_it:
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    soup_all_url_job_it.append(soup)
    count+=1
    print("Number of Web Jobbkk Category It downloaded = "+ str(count), end='\r')
print('Total URL Jobbkk Category It = '+str(len(soup_all_url_job_it)))
print('Finish Load Web => URL Jobbkk Category It')

print('Start Find URL Position Work')
url_all_detail_job = []
for soup in soup_all_url_job_it:
    page_all_job_it = soup.find_all('a', class_='positon-work')
    for page_all_job_it_0 in page_all_job_it:
        link = page_all_job_it_0.get('href')
        url_all_detail_job.append(link)
print('Total URL Position Work = '+ str(len(url_all_detail_job)))
print('Finish Find URL Position Work')

print('Start Scaping Text')
count = 1
detail_job = []
for url in url_all_detail_job:
    response = requests.get(url)
    soup = BeautifulSoup(response.text,'lxml')
    page = soup.find_all('article', class_='row')
    for i in page:
        try:
            try:
                Company_Name = i.find('div',class_='col-md-12 col-sm-12 col-xs-12').text.strip('\n')
            except:
                Company_Name = i.find('p',class_='textRed fontSubHead').text.strip('\n')
        except:
            try:
                Company_Name = i.find('p',class_='textRed fontSubHead font-DB-HeaventRounded-Bold').text.strip('\n')
            except:
                Company_Name = ''
        try:
            try:
                Job_Position = i.find('div',class_='col-md-8 col-sm-8 col-xs-6').text.strip('\n')
```

```
        except:
            Job_Position = i.find('h1',class_='textRed fontSubHead').text.strip('\n')
    except:
        try:
            Job_Position = i.find('p',class_='textRed font-text-20 font-DB-HeaventRounded-Bold').text.strip('\n')
        except:
            Job_Position = ''
    try:
        try:
            Company_Address = i.find('b',string='สถานที่ปฏิบัติงาน : ').parent.text
        except:
            Company_Address = i.find('span',string='สถานที่ปฏิบัติงาน : ').parent.text
    except:
        Company_Address = ''
    try:
        try:
            Salary = i.find('b',string='เงินเดือน(บาท) : ').parent.text
        except:
            Salary = i.find('span',string='เงินเดือน(บาท) : ').parent.text
    except:
        Salary = ''
    try:
        Job_Description = i.find('p',string='หน้าที่ความรับผิดชอบ').parent.text.strip('\n')
    except:
        Job_Description = ''
    try:
        Candidate_Qualifications = i.find('p',string='คุณสมบัติ').parent.text.strip('\n')
    except:
        Candidate_Qualifications = ''
    detail_job.append([url,Company_Name,Job_Position,Company_Address,Salary,Job_Description,Candidate_
                        Qualifications])
    count+=1
    print("Number of detail job downloaded = "+ str(count), end='\r')
print('Total Scraping Position Work = '+str(len(detail_job)))
print('Finish Scaping Text')

import datetime
currentdate=datetime.date.today()
Currentdate=str(currentdate)

print('Start Save To File Excel')
column_names = ['url','Company Name','Job Position','Company Address','Salary','Job Description','Candidate
                    Qualifications']
df = pd.DataFrame(detail_job, columns=column_names)
df['date'] = Currentdate
```

```
    df.index = df.index + 1

    path_file = 'C:/Code_IS/code/scraping/'+Currentdate+'JobBKK.xlsx'
    writer = pd.ExcelWriter(path_file)
    df.to_excel(writer,'MySheet1',index=False)
    writer.save()

    print('Finish Save To File Excel')

def combine():

    import pandas as pd
    import datetime

    print('start combine')
    currentdate=datetime.date.today()
    Currentdate= str(currentdate)
    dataJobbkk = pd.read_excel('C:/Code_IS/code/scraping/'+Currentdate+'JobBKK.xlsx')
    dataJoblognone = pd.read_excel('C:/Code_IS/code/scraping/'+Currentdate+'JobBlognone.xlsx')
    dataJobsdb = pd.read_excel('C:/Code_IS/code/scraping/'+Currentdate+'JobsDB.xlsx')
    dataJobthai = pd.read_excel('C:/Code_IS/code/scraping/'+Currentdate+'Jobthai.xlsx')
    combine_data = pd.concat([dataJobbkk,dataJoblognone,dataJobsdb,dataJobthai])
    print('finish combine')
    return combine_data

def tag(df):
    import pandas as pd
    import re
    import string

    print('start tag')

    def url_f(txt):
        key='token'
        try:
            str = re.sub(key+'.*', key, txt)
            return str
        except:
            return txt
    df['url']=df['url'].apply(url_f)

    df.dropna(subset=['Job Position'],inplace=True)
    df.drop_duplicates(subset=['url','Company Name','Job Position'],inplace=True)
    df['Job Position']=df['Job Position'].str.lower()
    df['Company Address']=df['Company Address'].str.lower()
```

```
df['Job Description']=df['Job Description'].str.lower()
df['Candidate Qualifications']=df['Candidate Qualifications'].str.lower()
df.fillna('NaN',inplace=True)

df2 = pd.DataFrame(df,columns=['ID_Job','date','url','Company Name','Job Position','tag_Jobname','tag_add',
                                'Salary min','Salary max'])

df2.loc[df['Job Position'].str.contains('it|technic')
  &df['Job Position'].str.contains('specialist'),'tag_Jobname']                    ='it specialist'
df2.loc[df['Job Position'].str.contains('data|information|ข้อมูล'),'tag_Jobname']      ='data management'
df2.loc[df['Job Position'].str.contains('analy|วิเคราะ'),'tag_Jobname']              ='data analyst'
df2.loc[df['Job Position'].str.contains('web|เว'),'tag_Jobname']                    ='web developer'
df2.loc[df['Job Position'].str.contains('automation'),'tag_Jobname']                ='automation engineer'
df2.loc[df['Job Position'].str.contains('qa'),'tag_Jobname']                        ='qa engineer'
df2.loc[df['Job Position'].str.contains('test|ทดสอบ'),'tag_Jobname']                ='tester'
df2.loc[df['Job Position'].str.contains('mobile|ios|android'),'tag_Jobname']        ='mobile developer'
df2.loc[df['Job Position'].str.contains('dev|พัฒนา'),'tag_Jobname']                 ='developer'
df2.loc[df['Job Position'].str.contains('sa'),'tag_Jobname']                        ='system analyst'
df2.loc[df['Job Position'].str.contains('business|bi'),'tag_Jobname']               ='business analyst'
df2.loc[df['Job Position'].str.contains('program'),'tag_Jobname']                   ='programmer'
df2.loc[df['Job Position'].str.contains('โปรแกรม'),'tag_Jobname']                   ='programmer'
df2.loc[df['Job Position'].str.contains('ux|ui|experience|interface'),'tag_Jobname'] ='ux ui'
df2.loc[df['Job Position'].str.contains('sale'),'tag_Jobname']                      ='sale'
df2.loc[df['Job Position'].str.contains('sap'),'tag_Jobname']                       ='sap developer'
df2.loc[df['Job Position'].str.contains('network|อินเตอ|เครือข่าย|infrastructure|infra|internet|server'),'tag_Jobname']='network
                                                                                                                    engineer'
df2.loc[df['Job Position'].str.contains('security'),'tag_Jobname']                  ='security'
df2.loc[df['Job Position'].str.contains('seo'),'tag_Jobname']                       ='seo'
df2.loc[df['Job Position'].str.contains('database|ฐานข้อมูล|บริหารข้อมูล|dba|ระบบข้อมูล|ระบบสาร|ระบบ'),'tag_Jobname']=
                                                                                                                    'database'
df2.loc[df['Job Position'].str.contains('erp'),'tag_Jobname']                       ='erp'
df2.loc[df['Job Position'].str.contains('cloud'),'tag_Jobname']                     ='cloud engineer'
df2.loc[df['Job Position'].str.contains('devops|devsecops'),'tag_Jobname']          ='devops engineer'
df2.loc[df['Job Position'].str.contains('account'),'tag_Jobname']                   ='account manager'
df2.loc[df['Job Position'].str.contains('scrum'),'tag_Jobname']                     ='scrum master'
df2.loc[df['Job Position'].str.contains('3d'),'tag_Jobname']                        ='3d model'
df2.loc[df['Job Position'].str.contains('iot'),'tag_Jobname']                       ='iot engineer'
df2.loc[df['Job Position'].str.contains('graphic|2d'),'tag_Jobname']                ='graphic design'
df2.loc[df['Job Position'].str.contains('noc'),'tag_Jobname']                       ='network operation center engineer'
df2.loc[df['Job Position'].str.contains('risk'),'tag_Jobname']                      ='risk manager'
df2.loc[df['Job Position'].str.contains('strategy'),'tag_Jobname']                  ='it strategy'
df2.loc[df['Job Position'].str.contains('hris'),'tag_Jobname']                      ='human resource information system'
df2.loc[df['Job Position'].str.contains('mis'),'tag_Jobname']                       ='mis'
df2.loc[df['Job Position'].str.contains('consul'),'tag_Jobname']                    ='consultant'
df2.loc[df['Job Position'].str.contains('train'),'tag_Jobname']                     ='technical training'
```

```
df2.loc[df['Job Position'].str.contains('ml'),'tag_Jobname']                    ='machine learning engineer'
df2.loc[df['Job Position'].str.contains('market'),'tag_Jobname']                ='digital marketing'
df2.loc[df['Job Position'].str.contains('special'),'tag_Jobname']               ='it specialist'
df2.loc[df['Job Position'].str.contains('game'),'tag_Jobname']                  ='game developer'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('manager'),'tag_Jobname']                  ='it manager'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('audit'),'tag_Jobname']                    ='it audit'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('admin'),'tag_Jobname']                    ='it administrator'
df2.loc[df['Job Position'].str.contains('service')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']                 ='it service engineer'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']                 ='it engineer'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('support|service'),'tag_Jobname']          ='it support'
df2.loc[df['Job Position'].str.contains('it|ไอที|สารสนเท|คอม')
    &df['Job Position'].str.contains('ผู้อำนวยการ|ฝ่าย|พนัก|หัวหน้า|แผนก|เจ้าหน้าที่|staff|offic|จัดการ|ผู้ช่วย|ช่าง'),'tag_Jobname']=
                                                                                                                    'it support'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('technician'),'tag_Jobname']               ='it technician'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('architect| archi'),'tag_Jobname']         ='it architect'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('supervis'),'tag_Jobname']                 ='it supervisor'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('management'),'tag_Jobname']               ='it management'
df2.loc[df['Job Position'].str.contains('it')
    &df['Job Position'].str.contains('compliance'),'tag_Jobname']               ='it compliance management'
df2.loc[df['Job Position'].str.contains('software|ซอฟ')
    &df['Job Position'].str.contains('support|สนับสนุน'),'tag_Jobname']         ='software support'
df2.loc[df['Job Position'].str.contains('data')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']                 ='data engineer'
df2.loc[df['Job Position'].str.contains('data|ข้อมูล')
    &df['Job Position'].str.contains('analy|วิเคราะ'),'tag_Jobname']            ='data analyst'
df2.loc[df['Job Position'].str.contains('data')
    &df['Job Position'].str.contains('scien'),'tag_Jobname']                    ='data science'
df2.loc[df['Job Position'].str.contains('mobile|ios|android')
    &df['Job Position'].str.contains('develop'),'tag_Jobname']                  ='mobile developer'
df2.loc[df['Job Position'].str.contains('develop|พัฒนา')
    &df['Job Position'].str.contains('application|แอพ'),'tag_Jobname']          ='application developer'
df2.loc[df['Job Position'].str.contains('develop|พัฒนา')
    &df['Job Position'].str.contains('web|php|เว'),'tag_Jobname']               ='web developer'
df2.loc[df['Job Position'].str.contains('develop|พัฒนา')
    &df['Job Position'].str.contains('software|ซอฟ'),'tag_Jobname']             ='software developer'
```

```
df2.loc[df['Job Position'].str.contains('full')
    &df['Job Position'].str.contains('stack'),'tag_Jobname']          ='full stack developer'
df2.loc[df['Job Position'].str.contains('develop')
    &df['Job Position'].str.contains('full')
    &df['Job Position'].str.contains('stack'),'tag_Jobname']          ='full stack developer'
df2.loc[df['Job Position'].str.contains('front')
    &df['Job Position'].str.contains('end'),'tag_Jobname']            ='front end developer'
df2.loc[df['Job Position'].str.contains('develop')
    &df['Job Position'].str.contains('front')
    &df['Job Position'].str.contains('end'),'tag_Jobname']            ='front end developer'
df2.loc[df['Job Position'].str.contains('develop')
    &df['Job Position'].str.contains('back')
    &df['Job Position'].str.contains('end'),'tag_Jobname']            ='back end developer'
df2.loc[df['Job Position'].str.contains('back')
    &df['Job Position'].str.contains('end'),'tag_Jobname']            ='back end developer'
df2.loc[df['Job Position'].str.contains('program')
    &df['Job Position'].str.contains('web'),'tag_Jobname']            ='web programmer'
df2.loc[df['Job Position'].str.contains('system')
    &df['Job Position'].str.contains('analy'),'tag_Jobname']          ='system analyst'
df2.loc[df['Job Position'].str.contains('ระบบ')
    &df['Job Position'].str.contains('วิเคราะ'),'tag_Jobname']        ='system analyst'
df2.loc[df['Job Position'].str.contains('system')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']       ='system engineer'
df2.loc[df['Job Position'].str.contains('business')
    &df['Job Position'].str.contains('analy'),'tag_Jobname']          ='business analyst'
df2.loc[df['Job Position'].str.contains('software')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']       ='software engineer'
df2.loc[df['Job Position'].str.contains('project')
    &df['Job Position'].str.contains('manager'),'tag_Jobname']        ='project manager'
df2.loc[df['Job Position'].str.contains('help')
    &df['Job Position'].str.contains('desk'),'tag_Jobname']           ='it help desk'
df2.loc[df['Job Position'].str.contains('quality')
    &df['Job Position'].str.contains('assurance'),'tag_Jobname']      ='qa engineer'
df2.loc[df['Job Position'].str.contains('quality')
    &df['Job Position'].str.contains('assurance')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']       ='qa engineer'
df2.loc[df['Job Position'].str.contains('qa')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']       ='qa engineer'
df2.loc[df['Job Position'].str.contains('project|it')
    &df['Job Position'].str.contains('coordinat'),'tag_Jobname']      ='project coordinator'
df2.loc[df['Job Position'].str.contains('product')
    &df['Job Position'].str.contains('manager'),'tag_Jobname']        ='product manager'
df2.loc[df['Job Position'].str.contains('system')
    &df['Job Position'].str.contains('admin'),'tag_Jobname']          ='system administrator'
df2.loc[df['Job Position'].str.contains('application')
```

```
&df['Job Position'].str.contains('consult'),'tag_Jobname']                ='application consultant'
df2.loc[df['Job Position'].str.contains('application')
&df['Job Position'].str.contains('support'),'tag_Jobname']                ='application support'
df2.loc[df['Job Position'].str.contains('system')
&df['Job Position'].str.contains('support'),'tag_Jobname']                ='system support'
df2.loc[df['Job Position'].str.contains('software')
&df['Job Position'].str.contains('consult'),'tag_Jobname']                ='software consultant'
df2.loc[df['Job Position'].str.contains('service')
&df['Job Position'].str.contains('desk'),'tag_Jobname']                   ='service desk'
df2.loc[df['Job Position'].str.contains('risk')
&df['Job Position'].str.contains('manager|management'),'tag_Jobname']        ='risk manager'
df2.loc[df['Job Position'].str.contains('technic')
&df['Job Position'].str.contains('support'),'tag_Jobname']                ='technical support'
df2.loc[df['Job Position'].str.contains('product')
&df['Job Position'].str.contains('owner'),'tag_Jobname']                  ='product owner'
df2.loc[df['Job Position'].str.contains('tech|it')
&df['Job Position'].str.contains('lead'),'tag_Jobname']                   ='technical lead'
df2.loc[df['Job Position'].str.contains('technic')
&df['Job Position'].str.contains('consul'),'tag_Jobname']                 ='technical consultant'
df2.loc[df['Job Position'].str.contains('solution')
&df['Job Position'].str.contains('consul'),'tag_Jobname']                 ='solution consultant'
df2.loc[df['Job Position'].str.contains('human')
&df['Job Position'].str.contains('resource')
&df['Job Position'].str.contains('information')
&df['Job Position'].str.contains('system'),'tag_Jobname']                 ='human resource information system'
df2.loc[df['Job Position'].str.contains('product')
&df['Job Position'].str.contains('support'),'tag_Jobname']                ='production support'
df2.loc[df['Job Position'].str.contains('product')
&df['Job Position'].str.contains('advisor'),'tag_Jobname']                ='product advisor'
df2.loc[df['Job Position'].str.contains('bpr')
&df['Job Position'].str.contains('consul'),'tag_Jobname']                 ='bpr consultant'
df2.loc[df['Job Position'].str.contains('solution')
&df['Job Position'].str.contains('design'),'tag_Jobname']                 ='solution design'
df2.loc[df['Job Position'].str.contains('technic')
&df['Job Position'].str.contains('implement'),'tag_Jobname']               ='technical implementation'
df2.loc[df['Job Position'].str.contains('client')
&df['Job Position'].str.contains('implement'),'tag_Jobname']               ='client implementation'
df2.loc[df['Job Position'].str.contains('machine')
&df['Job Position'].str.contains('learning'),'tag_Jobname']                ='machine learning engineer'
df2.loc[df['Job Position'].str.contains('project')
&df['Job Position'].str.contains('management|manager'),'tag_Jobname']        ='it project management'
df2.loc[df['Job Position'].str.contains('web')
&df['Job Position'].str.contains('master|admin'),'tag_Jobname']            ='web admin'
df2.loc[df['Job Position'].str.contains('digital')
&df['Job Position'].str.contains('market'),'tag_Jobname']                 ='digital marketing'
```

```
df2.loc[df['Job Position'].str.contains('data|information')
    &df['Job Position'].str.contains('management|manager|expert|executive'),'tag_Jobname'] ='data management'
df2.loc[df['Job Position'].str.contains('web')
    &df['Job Position'].str.contains('design'),'tag_Jobname']                    ='web design'
df2.loc[df['Job Position'].str.contains('appli')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']                  ='application engineer'
df2.loc[df['Job Position'].str.contains('appli')
    &df['Job Position'].str.contains('owner'),'tag_Jobname']                     ='application owner'
df2.loc[df['Job Position'].str.contains('solution')
    &df['Job Position'].str.contains('engineer'),'tag_Jobname']                  ='solution engineer'
df2.loc[df['Job Position'].str.contains('system')
    &df['Job Position'].str.contains('manager'),'tag_Jobname']                   ='system manager'

df_id_Jobname = pd.read_excel(r'C:\Code_IS\code\powerbi\id_Jobname.xlsx')
for idx, row in df_id_Jobname.iterrows():
    df2.loc[df2['tag_Jobname'].str.contains(str(row.tag_Jobname_name),na=False),'tag_Jobname']= str(row.tag_Jobname)

df['Salary'] = df['Salary'].replace(['k','000',regex=True)
df['Salary'] = df['Salary'].replace([',],'',regex=True)
df_salary = pd.DataFrame(columns=['tag_salary'])
def txt(txt):
    txt = re.findall('\d\d\d\d\d\d|\d\d\d\d\d',txt)
    txt = "-".join(u for u in txt )
    return txt
df_salary['tag_salary'] = df['Salary'].apply(txt)
df_salary=df_salary.tag_salary.str.split("-", expand = True)
df2['Salary min']=df_salary[0]
df2['Salary max']=df_salary[1]

df['Job Description'] = df['Job Description'].replace(['/'],' ',regex=True)
df['Candidate Qualifications'] = df['Candidate Qualifications'].replace(['/'],' ',regex=True)
df['Job Description'] = df['Job Description'].replace([','],' ',regex=True)
df['Candidate Qualifications'] = df['Candidate Qualifications'].replace([','],' ',regex=True)
lists = [' abap ',' .net ',' kotlin ',' c ',' c+ ',' c# ',' objective-c ',' swift ',' java ',' js ',' javascript '
        ,' html ',' css ',' vba ',' vb ',' python ',' ruby ',' php ',' sql ',' perl ',' go ',' r ',' golang '
        ,' typescript ',' solidity ',' x ',' ms office ',' microsoft office ','microsoft 365','ms 365'
        ,' facebook ',' instagram ',' line ',' twitter ',' tiktok ',' youtube ',' unreal engine ',' akeytsu '
        ,' mixamo ',' daz studio ',' maya ',' 3ds Max ',' cinema 4d ',' motionbuilder ',' deepmotion ',' move.ai '
        ,' blender ',' unity ',' hubspot ',' salesforce ',' zoho ',' pipedrive ',' asana ',' trello ',' basecamp '
        ,' wrike ',' surveymonkey ',' typeform ',' qualtrics ',' google forms ',' google analytics ',' tableau '
        ,' power bi ',' clickup ',' jira ',' airtable ',' google data studio ',' looker studio ',' word ',' excel '
        ,' microsoft team ',' powerpoint ',' visio ',' microsoft power platform ',' power apps ',' power automate '
        ,' virtual agents ',' power pages ',' react ',' vue ',' angular ',' svelte ',' jquery ',' ember ',' backbone '
        ,' semantic ui ',' foundation ',' preact ',' github ',' marvel ',' node.js ',' bootstrap ',' postman '
        ,' docker ',' kubernetes ',' flutter ',' laravel ',' django ',' mongodb ',' mysql ',' apache ',' nginx '
```

```
        ,' asp.net ',' express js ',' nestjs ',' rest api ',' flask ',' android studio ',' react native ',' xamarin '
        ,' firebase ',' xcode ',' robot framework ',' webdriver ',' selenium ',' jmeter ',' jenkin ',' appium '
        ,' katalon ',' sas ',' sap ',' oracle ',' odoo ',' sharepoint ',' ssms ',' sql server ',' rdbms ',' mariadb '
        ,' postgresql ',' ibm ',' db2 ',' microsoft access ',' amazon ',' redis ',' sqlite ',' microsoft azure '
        ,' gcp ',' google cloud platform ',' alibaba cloud ',' huawei cloud ',' amazon web services ',' aws ',' spark '
        ,' hive ',' airflow ',' kafka ',' snowflake ',' redshift ',' athena ',' databrick ',' data factory ',' pandas '
        ,' seaborn ',' scikit learn ',' pytorch ',' mlflow ',' rapidminer ',' chatgpt ',' tensorflow ',' keras '
        ,' matplotlib ',' numpy ',' hadoop ',' packet tracer ',' wireshark ',' solarwinds ',' nmap ',' prtg ',' nagios '
        ,' spring ','restful api',' google keyword planner ',' ahrefs ',' ubersuggest ',' google search console '
        ,' se ranking ',' keyword hero ',' hotjar ',' google optimize ',' uipath ',' ansible ',' terraform ',' rundeck '
        ,' napalm ',' libcloud ',' prometheus ',' grafana ',' zabbix ',' sensu ',' elk ']
for word in lists:
    df2.loc[df['Job Description'].str.contains(word,na=False),word]='1'
    df2.loc[df['Candidate Qualifications'].str.contains(word,na=False),word]='1'
    df2[word].fillna('0',inplace=True)

address = ['นครราชสีมา', 'เชียงใหม่', 'กาญจนบุรี', 'ตาก', 'อุบลราชธานี', 'สุราษฎร์ธานี', 'ชัยภูมิ', 'แม่ฮ่องสอน', 'เพชรบูรณ์',
        'ลำปาง', 'อุดรธานี', 'เชียงราย', 'น่าน', 'เลย', 'ขอนแก่น', 'พิษณุโลก', 'บุรีรัมย์', 'นครศรีธรรมราช', 'สกลนคร',
        'นครสวรรค์', 'ศรีสะเกษ', 'กำแพงเพชร', 'ร้อยเอ็ด', 'สุรินทร์', 'อุตรดิตถ์', 'สงขลา', 'สระแก้ว', 'กาฬสินธุ์', 'อุทัยธานี',
        'สุโขทัย', 'แพร่', 'ประจวบคีรีขันธ์', 'จันทบุรี', 'พะเยา', 'เพชรบุรี', 'ลพบุรี', 'ชุมพร', 'นครพนม', 'สุพรรณบุรี', 'ฉะเชิงเทรา',
        'มหาสารคาม', 'ราชบุรี', 'ตรัง', 'ปราจีนบุรี', 'กระบี่', 'พิจิตร', 'ยะลา', 'ลำพูน', 'นราธิวาส', 'ชลบุรี', 'มุกดาหาร', 'บึงกาฬ',
        'พังงา', 'ยโสธร', 'หนองบัวลำภู', 'สระบุรี', 'ระยอง', 'พัทลุง', 'ระนอง', 'อำนาจเจริญ', 'หนองคาย', 'ตราด', 'พระนครศรีอยุธยา',
        'สตูล', 'ชัยนาท', 'นครปฐม', 'นครนายก', 'ปัตตานี', 'กรุงเทพมหานคร', 'ปทุมธานี', 'สมุทรปราการ', 'อ่างทอง', 'สมุทรสาคร',
        'สิงห์บุรี', 'นนทบุรี', 'ภูเก็ต', 'สมุทรสงคราม']
for i in address:
    df2.loc[df['Company Address'].str.contains(i),'tag_add']=i
data_Bankok_into_list = ['คลองเตย', 'คลองเตย', 'คลองตัน', 'พระโขนง', 'คลองสาน', 'คลองต้นไทร', 'คลองสาน', 'บางลำภูล่าง',
                'สมเด็จเจ้าพระยา', 'คลองสามวา', 'ทรายกองดิน', 'ทรายกองดินใต้', 'บางชัน', 'สามวาตะวันตก', 'สามวาตะวันออก',
                'คันนายาว', 'คันนายาว', 'จตุจักร', 'จตุจักร', 'จอมพล', 'จันทรเกษม', 'ลาดยาว', 'เสนานิคม', 'จอมทอง',
                'จอมทอง', 'บางขุนเทียน', 'บางค้อ', 'บางมด', 'ดอนเมือง', 'สีกัน', 'ดินแดง', 'ดินแดง', 'ดุสิต', 'ดุสิต',
                'ถนนนครไชยศรี', 'วชิรพยาบาล', 'สวนจิตรลดา', 'สี่แยกมหานาค', 'ตลิ่งชัน', 'คลองชักพระ', 'ฉิมพลี', 'ตลิ่งชัน',
                'บางเชือกหนัง', 'บางพรม', 'บางระมาด', 'ทวีวัฒนา', 'ทวีวัฒนา', 'ศาลาธรรมสพน์', 'ทุ่งครุ', 'ทุ่งครุ', 'บางมด',
                'ธนบุรี', 'ดาวคะนอง', 'ตลาดพลู', 'บางยี่เรือ', 'บุคคโล', 'วัดกัลยาณ์', 'สำเหร่', 'หิรัญรูจี', 'บางเขน', 'ท่าแร้ง',
                'อนุสาวรีย์', 'บางแค', 'บางแค', 'บางแคเหนือ', 'บางไผ่', 'หลักสอง', 'บางกอกใหญ่', 'วัดท่าพระ', 'วัดอรุณ',
                'บางกอกน้อย', 'บางขุนนนท์', 'บางขุนศรี', 'บ้านช่างหล่อ', 'ศิริราช', 'อรุณอมรินทร์', 'บางกะปิ', 'คลองจั่น', 'หัวหมาก',
                'บางขุนเทียน', 'ท่าข้าม', 'แสมดำ', 'บางคอแหลม', 'บางโคล่', 'บางคอแหลม', 'วัดพระยาไกร', 'บางซื่อ', 'บางซื่อ',
                'บางนา', 'บางนา', 'บางบอน', 'บางบอน', 'บางพลัด', 'บางบำหรุ', 'บางพลัด', 'บางยี่ขัน', 'บางอ้อ', 'บางรัก',
                'บางรัก', 'มหาพฤฒาราม', 'สี่พระยา', 'สีลม', 'สุริยวงศ์', 'บึงกุ่ม', 'คลองกุ่ม', 'ปทุมวัน', 'ปทุมวัน', 'รองเมือง',
                'ลุมพินี', 'วังใหม่', 'ประเวศ', 'ดอกไม้', 'ประเวศ', 'หนองบอน', 'ป้อมปราบศัตรูพ่าย', 'คลองมหานาค', 'บ้านบาตร',
                'ป้อมปราบ', 'วัดเทพศิรินทร์', 'วัดโสมนัส', 'พญาไท', 'สามเสนใน', 'พระโขนง', 'บางจาก', 'พระนคร', 'ชนะสงคราม',
                'ตลาดยอด', 'บวรนิเวศ', 'บางขุนพรหม', 'บ้านพานถม', 'พระบรมมหาราชวัง', 'วังบูรพาภิรมย์', 'วัดราชบพิธ', 'วัดสามพระยา',
                'ศาลเจ้าพ่อเสือ', 'สำราญราษฎร์', 'เสาชิงช้า', 'ภาษีเจริญ', 'คลองขวาง', 'คูหาสวรรค์', 'บางจาก', 'บางด้วน', 'บางแวก',
                'บางหว้า', 'ปากคลองภาษีเจริญ', 'มีนบุรี', 'มีนบุรี', 'แสนแสบ', 'ยานนาวา', 'ช่องนนทรี', 'บางโพงพาง', 'ราชเทวี',
                'ถนนเพชรบุรี', 'ถนนพญาไท', 'ทุ่งพญาไท', 'มักกะสัน', 'ราษฎร์บูรณะ', 'บางปะกอก', 'ราษฎร์บูรณะ', 'ลาดกระบัง',
                'ขุมทอง', 'คลองสองต้นนุ่น', 'คลองสามประเวศ', 'ทับยาว', 'ลาดกระบัง', 'ลำปลาทิว', 'ลาดพร้าว', 'จรเข้บัว', 'ลาดพร้าว',
```

```
            'วังทองหลาง', 'วังทองหลาง', 'วัฒนา', 'คลองเตยเหนือ', 'คลองตันเหนือ', 'พระโขนงเหนือ', 'สวนหลวง', 'สวนหลวง',
            'สะพานสูง', 'สะพานสูง', 'สัมพันธวงศ์', 'จักรวรรดิ', 'ตลาดน้อย', 'สัมพันธวงศ์', 'สาทร', 'ทุ่งมหาเมฆ', 'ทุ่งวัดดอน',
            'ยานนาวา', 'สายไหม', 'คลองถนน', 'สายไหม', 'ออเงิน', 'หนองแขม', 'หนองแขม', 'หนองค้างพลู', 'หนองจอก',
            'กระทุ่มราย', 'คลองสิบ', 'คลองสิบสอง', 'คู้ฝั่งเหนือ', 'โคกแฝด', 'ลำต้อยติ่ง', 'ลำผักชี', 'หนองจอก', 'หลักสี่',
            'ตลาดบางเขน', 'ทุ่งสองห้อง', 'ห้วยขวาง', 'บางกะปิ', 'สามเสนนอก', 'ห้วยขวาง', 'bts', 'mrt', 'brt', 'srt']
address = set(data_Bankok_into_list)
for i in address:
    df2.loc[df['Company Address'].str.contains(i),'tag_add']='กรุงเทพมหานคร'
data_nakon_into_list=['เมืองนครปฐม', 'พระปฐมเจดีย์', 'บางแขม', 'พระประโทน', 'ธรรมศาลา', 'ตาก้อง', 'มาบแค', 'สนามจันทร์',
            'ดอนยายหอม',
            'ถนนขาด', 'บ่อพลับ', 'นครปฐม', 'วังตะกู', 'หนองปากโลง', 'สามควายเผือก', 'ทุ่งน้อย', 'หนองดินแดง', 'วังเย็น', 'โพรงมะเดื่อ',
            'ลำพยา', 'สระกะเทียม', 'สวนป่าน', 'ห้วยจรเข้', 'ทัพหลวง', 'หนองงูเหลือม', 'บ้านยาง', 'กำแพงแสน', 'ทุ่งกระพังโหม', 'กระตีบ',
            'ทุ่งลูกนก', 'ห้วยขวาง', 'ทุ่งขวาง', 'สระสี่มุม', 'ทุ่งบัว', 'ดอนข่อย', 'สระพัฒนา', 'ห้วยหมอนทอง', 'ห้วยม่วง', 'กำแพงแสน',
            'รางพิกุล', 'หนองกระทุ่ม', 'วังน้ำเขียว', 'นครชัยศรี', 'นครชัยศรี', 'บางกระเบา', 'วัดแค', 'ท่าตำหนัก', 'บางแก้ว', 'ท่ากระชับ',
            'ขุนแก้ว', 'ท่าพระยา', 'พะเนียด', 'บางระกำ', 'โคกพระเจดีย์', 'ศรีษะทอง', 'แหลมบัว', 'ศรีมหาโพธิ์', 'สัมปทวน', 'วัดสำโรง',
            'ดอนแฝก', 'ห้วยพลู', 'วัดละมุด', 'บางพระ', 'บางแก้วฟ้า', 'ลานตากฟ้า', 'งิ้วราย', 'ไทยาวาส', 'ดอนตูม', 'สามง่าม', 'ห้วยพระ',
            'ลำเหย', 'ดอนพุทรา', 'บ้านหลวง', 'ดอนรวก', 'ห้วยด้วน', 'ลำลูกบัว', 'บางเลน', 'บางเลน', 'บางปลา', 'บางหลวง', 'บางภาษี',
            'บางระกำ', 'บางไทรป่า', 'หินมูล', 'ไทรงาม', 'ดอนตูม', 'นิลเพชร', 'บัวปากท่า', 'คลองนกกระทุง', 'นราภิรมย์', 'ลำพญา',
            'ไผ่หูช้าง', 'สามพราน', 'ท่าข้าม', 'ทรงคนอง', 'หอมเกร็ด', 'บางกระทึก', 'บางเตย', 'สามพราน', 'บางช้าง', 'ไร่ขิง', 'ท่าตลาด',
            'กระทุ่มล้ม', 'คลองใหม่', 'ตลาดจินดา', 'คลองจินดา', 'ยายชา', 'บ้านใหม่', 'อ้อมใหญ่', 'พุทธมณฑล', 'ศาลายา', 'คลองโยง',
            'มหาสวัสดิ์']
address = set(data_nakon_into_list)
for i in address:
    df2.loc[df['Company Address'].str.contains(i),'tag_add']='นครปฐม'
data_non_into_list=['เมืองนนทบุรี', 'สวนใหญ่', 'ตลาดขวัญ', 'บางเขน', 'บางกระสอ', 'ท่าทราย', 'บางไผ่', 'บางศรีเมือง', 'บางกร่าง', 'ไทรม้า',
            'บางรักน้อย', 'บางกรวย', 'วัดชลอ', 'บางกรวย', 'บางสีทอง', 'บางขนุน', 'บางขุนกอง', 'บางคูเวียง', 'มหาสวัสดิ์', 'ปลายบาง',
            'ศาลากลาง', 'บางใหญ่', 'บางม่วง', 'บางแม่นาง', 'บางเลน', 'เสาธงหิน', 'บางใหญ่', 'บ้านใหม่', 'บางบัวทอง', 'โสนลอย',
            'บางบัวทอง', 'บางรักใหญ่', 'บางคูรัด', 'ละหาร', 'ลำโพ', 'พิมลราช', 'บางรักพัฒนา', 'ไทรน้อย', 'ไทรน้อย', 'ราษฎร์นิยม',
            'หนองเพรางาย', 'ไทรใหญ่', 'ขุนศรี', 'คลองขวาง', 'ทวีวัฒนา', 'ปากเกร็ด', 'ปากเกร็ด', 'บางตลาด', 'บ้านใหม่', 'บางพูด',
            'บางตะไนย์', 'คลองพระอุดม', 'ท่าอิฐ', 'เกาะเกร็ด', 'อ้อมเกร็ด', 'คลองข่อย', 'บางพลับ', 'คลองเกลือ']
address = set(data_non_into_list)
for i in address:
    df2.loc[df['Company Address'].str.contains(i),'tag_add']='นนทบุรี'
data_pathum_into_list=['เมืองปทุมธานี', 'บางปรอก', 'บ้านใหม่', 'บ้านกลาง', 'บ้านฉาง', 'บ้านกระแชง', 'บางขะแยง', 'บางคูวัด', 'บางหลวง',
            'บางเดื่อ', 'บางพูด', 'บางพูน', 'บางกะดี', 'สวนพริกไทย', 'หลักหก', 'คลองหลวง', 'คลองหนึ่ง', 'คลองสอง', 'คลองสาม',
            'คลองสี่', 'คลองห้า', 'คลองหก', 'คลองเจ็ด', 'ธัญบุรี', 'ประชาธิปัตย์', 'บึงยี่โถ', 'รังสิต', 'ลำผักกูด', 'บึงสนั่น',
            'บึงน้ำรักษ์', 'หนองเสือ', 'บึงบา', 'บึงบอน', 'บึงกาสาม', 'บึงชำอ้อ', 'หนองสามวัง', 'ศาลาครุ', 'นพรัตน์', 'ลาดหลุมแก้ว',
            'ระแหง', 'ลาดหลุมแก้ว', 'คูบางหลวง', 'คูขวาง', 'คลองพระอุดม', 'บ่อเงิน', 'หน้าไม้', 'ลำลูกกา', 'คูคต', 'ลาดสวาย',
            'บึงคำพร้อย', 'ลำลูกกา', 'บึงทองหลาง', 'ลำไทร', 'บึงคอไห', 'พืชอุดม', 'สามโคก', 'บางเตย', 'คลองควาย', 'สามโคก',
            'กระแชง', 'บางโพธิ์เหนือ', 'เชียงรากใหญ่', 'บ้านปทุม', 'บ้านงิ้ว', 'เชียงรากน้อย', 'บางกระบือ', 'ท้ายเกาะ']
address = set(data_pathum_into_list)
for i in address:
    df2.loc[df['Company Address'].str.contains(i),'tag_add']='ปทุมธานี'
data_samutprakan_into_list=['เมืองสมุทรปราการ', 'ปากน้ำ', 'สำโรงเหนือ', 'บางเมือง', 'ท้ายบ้าน', 'บางปูใหม่', 'แพรกษา', 'บางโปรง',
            'บางปู',
```

```
                    'บางด้วน', 'บางเมืองใหม่', 'เทพารักษ์', 'ท้ายบ้านใหม่', 'แพรกษาใหม่', 'บางบ่อ', 'บางบ่อ', 'บ้านระกาศ',
                    'บางพลีน้อย', 'บางเพรียง', 'คลองด่าน', 'คลองสวน', 'เปร็ง', 'คลองนิยมยาตรา', 'บางพลี', 'บางพลีใหญ่', 'บางแก้ว',
                    'บางปลา', 'บางโฉลง', 'ราชาเทวะ', 'หนองปรือ', 'พระประแดง', 'ตลาด', 'บางพึ่ง', 'บางจาก', 'บางครุ',
                    'บางหญ้าแพรก', 'บางหัวเสือ', 'สำโรงใต้', 'บางยอ', 'บางกะเจ้า', 'บางน้ำผึ้ง', 'บางกระสอบ', 'บางกอบัว', 'ทรงคนอง',
                    'สำโรง', 'สำโรงกลาง', 'พระสมุทรเจดีย์', 'นาเกลือ', 'บ้านคลองสวน', 'แหลมฟ้าผ่า', 'ปากคลองบางปลากด',
                    'ในคลองบางปลากด', 'บางเสาธง', 'บางเสาธง', 'ศีรษะจรเข้น้อย', 'ศีรษะจรเข้ใหญ่']
    address = set(data_samutprakan_into_list)
    for i in address:
        df2.loc[df['Company Address'].str.contains(i),'tag_add']='สมุทรปราการ'
    data_samutsakon_into_list=['เมืองสมุทรสาคร', 'มหาชัย', 'ท่าฉลอม', 'โกรกกราก', 'บ้านบ่อ', 'บางโทรัด', 'กาหลง', 'นาโคก', 'ท่าจีน', 'นาดี',
                    'ท่าทราย', 'คอกกระบือ', 'บางน้ำจืด', 'พันท้ายนรสิงห์', 'โคกขาม', 'บ้านเกาะ', 'บางกระเจ้า', 'บางหญ้าแพรก',
                    'ชัยมงคล', 'กระทุ่มแบน', 'ตลาดกระทุ่มแบน', 'อ้อมน้อย', 'ท่าไม้', 'สวนหลวง', 'บางยาง', 'คลองมะเดื่อ',
                    'หนองนกไข่', 'ดอนไก่ดี', 'แคราย', 'ท่าเสา', 'บ้านแพ้ว', 'บ้านแพ้ว', 'หลักสาม', 'ยกกระบัตร', 'โรงเข้',
                    'หนองสองห้อง', 'หนองบัว', 'หลักสอง', 'เจ็ดริ้ว', 'คลองตัน', 'อำแพง', 'สวนส้ม', 'เกษตรพัฒนา']
    address = set(data_samutsakon_into_list)
    for i in address:
        df2.loc[df['Company Address'].str.contains(i),'tag_add']='สมุทรสาคร'
    df2.loc[df['Company Address'].str.contains('อยุธยา'),'tag_add']='พระนครศรีอยุธยา'
    df2.loc[df['Company Address'].str.contains('ไม่ระบุ'),'tag_add']='ไม่ระบุ'
    df2['tag_add'].fillna('ไม่ระบุ',inplace=True)

    df_id_add = pd.read_excel(r'C:\Code_IS\code\powerbi\id_add.xlsx')
    for idx, row in df_id_add.iterrows():
        df2.loc[df2['tag_add'].str.contains(str(row.tag_add_name)),'tag_add']= str(row.tag_add)

    print('finish tag')
    return df2

def skill_table(df):
    import pandas as pd

    df2 = pd.DataFrame(df,columns=['ID_Job'
        ,' abap ',' .net ',' kotlin ',' c ',' c+ ',' c# ',' objective-c ',' swift ',' java ',' js ',' javascript '
        ,' html ',' css ',' vba ',' vb ',' python ',' ruby ',' php ',' sql ',' perl ',' go ',' r ',' golang '
        ,' typescript ',' solidity ',' x ',' ms office ',' microsoft office ','microsoft 365','ms 365'
        ,' facebook ',' instagram ',' line ',' twitter ',' tiktok ',' youtube ',' unreal engine ',' akeytsu '
        ,' mixamo ',' daz studio ',' maya ',' 3ds Max ',' cinema 4d ',' motionbuilder ',' deepmotion ',' move.ai '
        ,' blender ',' unity ',' hubspot ',' salesforce ',' zoho ',' pipedrive ',' asana ',' trello ',' basecamp '
        ,' wrike ',' surveymonkey ',' typeform ',' qualtrics ',' google forms ',' google analytics ',' tableau '
        ,' power bi ',' clickup ',' jira ',' airtable ',' google data studio ',' looker studio ',' word ',' excel '
        ,' microsoft team ',' powerpoint ',' visio ',' microsoft power platform ',' power apps ',' power automate '
        ,' virtual agents ',' power pages ',' react ',' vue ',' angular ',' svelte ',' jquery ',' ember ',' backbone '
        ,' semantic ui ',' foundation ',' preact ',' github ',' marvel ',' node.js ',' bootstrap ',' postman '
        ,' docker ',' kubernetes ',' flutter ',' laravel ',' django ',' mongodb ',' mysql ',' apache ',' nginx '
        ,' asp.net ',' express js ',' nestjs ',' rest api ',' flask ',' android studio ',' react native ',' xamarin '
```

```
    ,' firebase ',' xcode ',' robot framework ',' webdriver ',' selenium ',' jmeter ',' jenkin ',' appium '
    ,' katalon ',' sas ',' sap ',' oracle ',' odoo ',' sharepoint ',' ssms ',' sql server ',' rdbms ',' mariadb '
    ,' postgresql ',' ibm ',' db2 ',' microsoft access ',' amazon ',' redis ',' sqlite ',' microsoft azure '
    ,' gcp ',' google cloud platform ',' alibaba cloud ',' huawei cloud ',' amazon web services ',' aws ',' spark '
    ,' hive ',' airflow ',' kafka ',' snowflake ',' redshift ',' athena ',' databrick ',' data factory ',' pandas '
    ,' seaborn ',' scikit learn ',' pytorch ',' mlflow ',' rapidminer ',' chatgpt ',' tensorflow ',' keras '
    ,' matplotlib ',' numpy ',' hadoop ',' packet tracer ',' wireshark ',' solarwinds ',' nmap ',' prtg ',' nagios '
    ,' spring ','restful api',' google keyword planner ',' ahrefs ',' ubersuggest ',' google search console '
    ,' se ranking ',' keyword hero ',' hotjar ',' google optimize ',' uipath ',' ansible ',' terraform ',' rundeck '
    ,' napalm ',' libcloud ',' prometheus ',' grafana ',' zabbix ',' sensu ',' elk ' ])

skill_table_unpivot = pd.melt(df2, id_vars='ID_Job', value_vars=[
    ' abap ',' .net ',' kotlin ',' c ',' c+ ',' c# ',' objective-c ',' swift ',' java ',' js ',' javascript '
    ,' html ',' css ',' vba ',' vb ',' python ',' ruby ',' php ',' sql ',' perl ',' go ',' r ',' golang '
    ,' typescript ',' solidity ',' x ',' ms office ',' microsoft office ','microsoft 365','ms 365'
    ,' facebook ',' instagram ',' line ',' twitter ',' tiktok ',' youtube ',' unreal engine ',' akeytsu '
    ,' mixamo ',' daz studio ',' maya ',' 3ds Max ',' cinema 4d ',' motionbuilder ',' deepmotion ',' move.ai '
    ,' blender ',' unity ',' hubspot ',' salesforce ',' zoho ',' pipedrive ',' asana ',' trello ',' basecamp '
    ,' wrike ',' surveymonkey ',' typeform ',' qualtrics ',' google forms ',' google analytics ',' tableau '
    ,' power bi ',' clickup ',' jira ',' airtable ',' google data studio ',' looker studio ',' word ',' excel '
    ,' microsoft team ',' powerpoint ',' visio ',' microsoft power platform ',' power apps ',' power automate '
    ,' virtual agents ',' power pages ',' react ',' vue ',' angular ',' svelte ',' jquery ',' ember ',' backbone '
    ,' semantic ui ',' foundation ',' preact ',' github ',' marvel ',' node.js ',' bootstrap ',' postman '
    ,' docker ',' kubernetes ',' flutter ',' laravel ',' django ',' mongodb ',' mysql ',' apache ',' nginx '
    ,' asp.net ',' express js ',' nestjs ',' rest api ',' flask ',' android studio ',' react native ',' xamarin '
    ,' firebase ',' xcode ',' robot framework ',' webdriver ',' selenium ',' jmeter ',' jenkin ',' appium '
    ,' katalon ',' sas ',' sap ',' oracle ',' odoo ',' sharepoint ',' ssms ',' sql server ',' rdbms ',' mariadb '
    ,' postgresql ',' ibm ',' db2 ',' microsoft access ',' amazon ',' redis ',' sqlite ',' microsoft azure '
    ,' gcp ',' google cloud platform ',' alibaba cloud ',' huawei cloud ',' amazon web services ',' aws ',' spark '
    ,' hive ',' airflow ',' kafka ',' snowflake ',' redshift ',' athena ',' databrick ',' data factory ',' pandas '
    ,' seaborn ',' scikit learn ',' pytorch ',' mlflow ',' rapidminer ',' chatgpt ',' tensorflow ',' keras '
    ,' matplotlib ',' numpy ',' hadoop ',' packet tracer ',' wireshark ',' solarwinds ',' nmap ',' prtg ',' nagios '
    ,' spring ','restful api',' google keyword planner ',' ahrefs ',' ubersuggest ',' google search console '
    ,' se ranking ',' keyword hero ',' hotjar ',' google optimize ',' uipath ',' ansible ',' terraform ',' rundeck '
    ,' napalm ',' libcloud ',' prometheus ',' grafana ',' zabbix ',' sensu ',' elk ',var_name='tool',value_name='value')
skill_table_unpivot=skill_table_unpivot.astype('str')
skill_table = skill_table_unpivot[skill_table_unpivot.value=='1']
df_id_Toolname = pd.read_excel(r'C:\Code_IS\code\powerbi\id_tool.xlsx')
for idx, row in df_id_Toolname.iterrows():
    skill_table.loc[skill_table['tool'].str.contains(str(row.tool_tag),na=False,regex=False),'tool']= str(row.tool)
skill_table.to_csv('C:/Code_IS/code/powerbi/skill_table.csv',index=False)

def update_data(df_update):

    import pandas as pd
```

```
import datetime
import os
print('start update_data')

file_directory = 'C:/Code_IS/code/powerbi'
files = os.listdir(file_directory)
if 'JobPost_tag.csv' in files:
    data = pd.read_csv('C:/Code_IS/code/powerbi/JobPost_tag.csv')
    combine_data = pd.concat([data,df_update],ignore_index=True)
    combine_data.dropna(subset=['tag_Jobname'],inplace=True)
    combine_data.drop_duplicates(subset=['url','Company Name','Job Position'],inplace=True,ignore_index=True)
    combine_data['ID_Job'] = combine_data.index + 1
    skill_table(combine_data)
    combine_data.to_csv('C:/Code_IS/code/powerbi/JobPost_tag.csv',index=False)
    combine_data.to_csv('C:/Code_IS/code/powerbi/JobPost.csv',index=False,columns=['date','ID_Job','url'
                        ,'Company Name','Job Position','tag_Jobname','tag_add','Salary min','Salary max'])
    print('finish update_data')
else:
    df_update.dropna(subset=['tag_Jobname'],inplace=True)
    df = pd.DataFrame(df_update)
    df['ID_Job'] = df.index + 1
    skill_table(df)
    df.to_csv('C:/Code_IS/code/powerbi/JobPost_tag.csv',index=False)
    df.to_csv('C:/Code_IS/code/powerbi/JobPost.csv',index=False,columns=['date','ID_Job','url'
                        ,'Company Name','Job Position','tag_Jobname','tag_add','Salary min','Salary max'])
    print('finish create_data')

import time
start_time_all = time.time()
start_time = time.time()
jobsblognone()
end_time = time.time()
elapsed_time = end_time - start_time
print('work time jobsblognone=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
jobBBK()
end_time = time.time()
elapsed_time = end_time - start_time
print('work time jobBBK=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
jobthai()
end_time = time.time()
elapsed_time = end_time - start_time
```

```
print('work time jobthai=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
jobsdb()
end_time = time.time()
elapsed_time = end_time - start_time
print('work time jobsdb=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
time.sleep(10)
combine = combine()
end_time = time.time()
elapsed_time = end_time - start_time
print('work time combine=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
df_tag = tag(combine)
end_time = time.time()
elapsed_time = end_time - start_time
print('work time tag=',elapsed_time,'seconds')
print('****************************************')
start_time = time.time()
update_data(df_tag)
end_time = time.time()
elapsed_time = end_time - start_time
print('work time tag=',elapsed_time,'seconds')
print('****************************************')
end_time_all = time.time()
elapsed_time_all = end_time_all - start_time_all
print('all work time =',elapsed_time_all,'seconds')
```

ภาคผนวก ข.

แบบสอบถามการวิจัย

แบบสอบถามความพึงพอใจของผู้ที่ได้เข้ามาเยี่ยมชมต่อ Dashboard

แบบประเมิน Dashboard เทรนด์ของอาชีพสายไอที

ตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 ซึ่งมีผลบังคับใช้เมื่อวันที่ 1 มิถุนายน 2565 ผู้ทำงานวิจัยให้ความสำคัญอย่างยิ่งต่อการคุ้มครองข้อมูลส่วนบุคคลและการปฏิบัติตามกฎหมายคุ้มครองข้อมูลส่วนบุคคล ดังนั้นข้อมูลในแบบประเมินนี้จะจัดเก็บข้อมูลของท่านเป็นความลับ และเปิดเผยเท่าที่จำเป็นสำหรับนำไปใช้ทำงานค้นคว้าอิสระเท่านั้น

sr.karn_st@tni.ac.th Switch account

Not shared

* Indicates required question

ท่านมีความสนใจในงานสายไอที หรือ ทำงานที่เกี่ยวข้องกับไอทีไหม ? *

- (●) ใช่
- ( ) ไม่ใช่

Next

Clear form

แบบประเมิน Dashboard เทรนด์ของอาชีพสายไอที ส่วนที่ 1

ข้อมูลพื้นฐานของผู้ทำแบบประเมิน

ท่านเพศอะไร ? *

- (●) ชาย
- ( ) หญิง

ท่านอายุเท่าไหร่? *

- ( ) ต่ำกว่าเท่ากับ 18 ปี
- ( ) 19-24 ปี
- (•) 25-30 ปี
- ( ) 30-40 ปี
- ( ) มากกว่า 40 ปี

ท่านมีระดับการศึกษาอะไร ? *

- ( ) มัธยม
- ( ) ปริญาตรี
- (•) ปริญาโท
- ( ) ปริญญาเอก

ปัจจุบันท่านประกอบอาชีพสายไอทีใช่หรือไม่ ? *

- ( ) ใช่
- (•) ไม่ใช่

Back Next Clear form

แบบประเมิน Dashboard เทรนด์ของอาชีพสายไอที ส่วนที่ 2 ประโยชน์ของ Dashboardและความเหมาะสมของข้อมูล

ลิ้ง Dashboard เทรนด์ของอาชีพสายไอที
https://app.powerbi.com/view?r=eyJrIjoiYmJiOWYzNWYtYzhkYi00ZjE5LWIxYmEtMmIxYWUwZDk5MDc0IiwidCI6IjFhNzg0NmJlLWUwMTktNDBkNS1iMDY0LTJjZjAxMDAyNzJhNSIsImMiOjEwfQ==&pageName=ReportSection

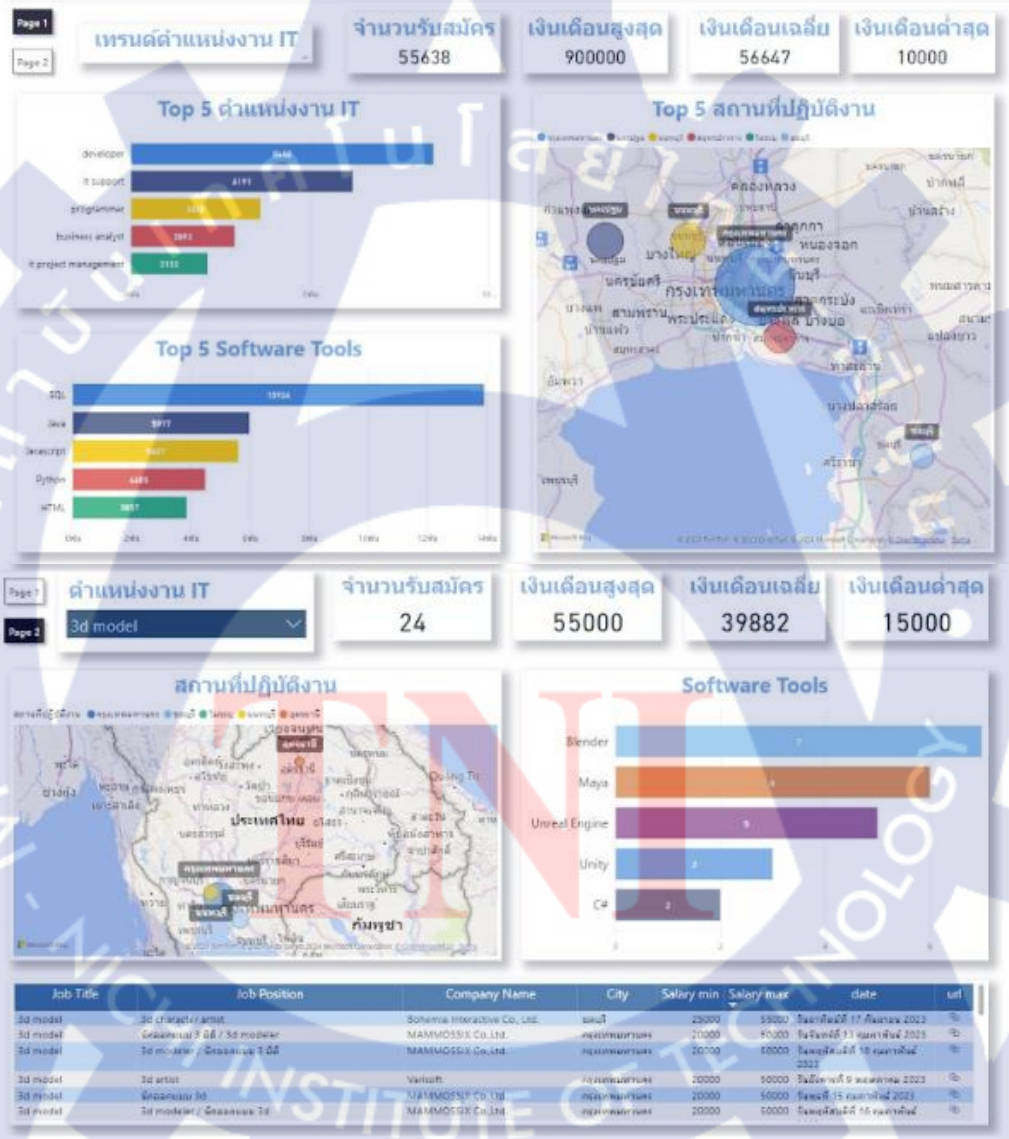
แบบประเมินความต้องการข้อมูลบน Dashboard เพื่อใช้ช่วยในการตัดสินใจใน การเลือกเปลี่ยนงานมาสายไอที , การหางานไอทีตำแหน่งใหม่ หรือ เลือกศึกษาต่อในหลักสูตรที่เกี่ยวข้องกับไอที *

	ต้องการ	ไม่ต้องการ
1. ท่านต้องการข้อมูล"จำนวนรับสมัคร"หรือไม่ ?	◉	○
2.ท่านต้องการข้อมูล"เงินเดือนสูงสุด"หรือไม่ ?	◉	○
3.ท่านต้องการข้อมูล"เงินเดือนเฉลี่ย"หรือไม่?	◉	○
4.ท่านต้องการข้อมูล"เงินเดือนต่ำสุด"หรือไม่?	◉	○
5.ท่านต้องการข้อมูล"Top 5 ตำแหน่งงาน IT"หรือไม่ ?	◉	○
6.ท่านต้องการข้อมูล"Top 5 สถานที่ปฏิบัติงาน"หรือไม่ ?	◉	○
7.ท่านต้องการข้อมูล"Top 5 ทักษะโปรแกรมมิ่งหรือเครื่องมือที่ใช้ในการทำงาน"หรือไม่ ?	◉	○
8.ท่านต้องการข้อมูล เงินเดือน ,ทักษะโปรแกรมมิ่งหรือเครื่องมือที่ใช้ในการทำงาน และ สถานที่ปฏิบัติงานในแต่ละตำแหน่งงานหรือไม่ ?	◉	○

ท่านมีความคิดเห็นว่า Dashboard เทรนด์ของอาชีพสายไอทีนี้เพียงพอต่อการตัดสินใจในการเลือกเปลี่ยนงานมาสายไอที , การหางานไอทีตำแหน่งใหม่ หรือ เลือกศึกษาต่อในหลักสูตรที่เกี่ยวข้องกับไอที หรือไม่ ? *

ลิ้ง Dashboard เทรนด์ของอาชีพสายไอที

https://app.powerbi.com/view?r=eyJrIjoiYmJiOWYzNWYtYzhkYi00ZjE5LWIxYmEtMmIxYWUwZDk5MDc0IiwidCI6IjFhNzg0NmJlLWUwMTktNDBkNS1iMDY0LTJjZjAxMDAyNzJhNSIsImMiOjEwfQ==&pageName=ReportSection



- (●) เพียงพอ
- ( ) ไม่เพียงพอ

ข้อเสนอแนะหากท่านมีความต้องการข้อมูลใดเพิ่มเติมบน Dashboard

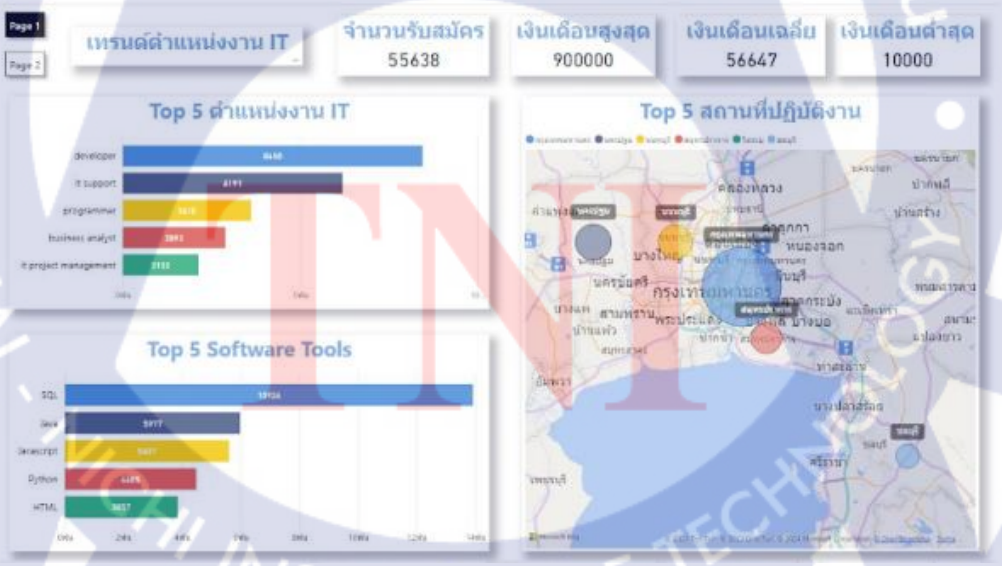
Your answer

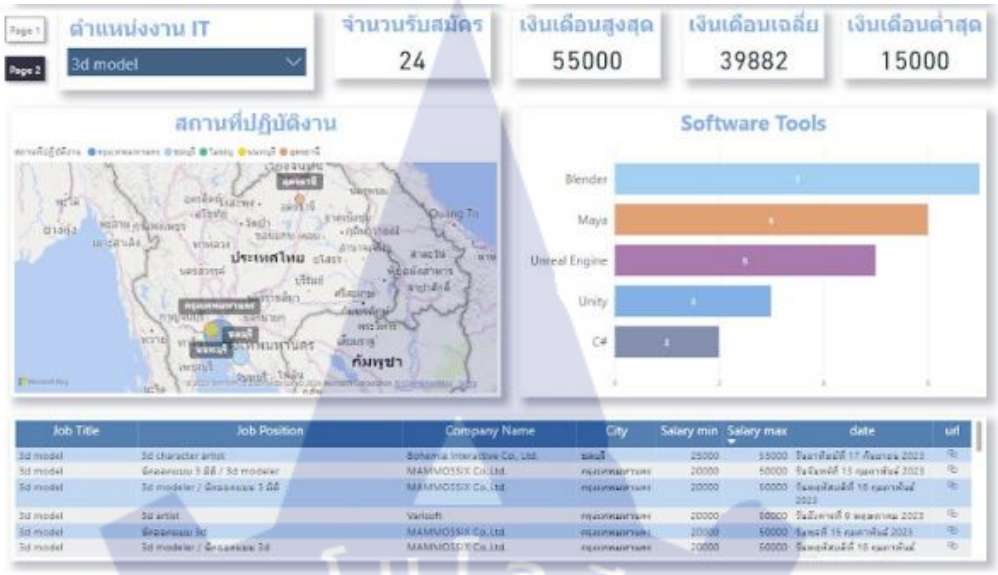
Back Next Clear form

แบบประเมิน Dashboard เทรนด์ของอาชีพสายไอที ส่วนที่ 3 ความเหมาะสมในการออกแบบ Dashboard แบบที่ A

ลิ้ง Dashboard แบบที่ A
https://app.powerbi.com/view?r=eyJrIjoiYmJiOWYzNWYtYzhkYi00ZjE5LWIxYmEtMmIxYWUwZDk5MDc0IiwidCI6IjFhNzg0NmJlLWUwMTktNDBkNS1iMDY0LTJjZjAxMDAyNzJhNSIsImMiOjEwfQ%3D%3D&pageName=ReportSection

ความเหมาะสมในการออกแบบ Dashboard แบบที่ A ? *





	มากที่สุด	มาก	ปานกลาง	ค่อนข้างน้อย	น้อย
ความเหมาะสมของสีพื้นหลังและสีตัวอักษรบน Dashboard ?	◉	○	○	○	○
ความเหมาะสมของขนาดตัวอักษรบน Dashboard ?	◉	○	○	○	○
ความเหมาะสมของฟอนต์ตัวอักษรบน Dashboard ?	◉	○	○	○	○
ความเหมาะสมในการวางตำแหน่งของ visualization (เช่น กราฟหรือแผนภูมิ) ต่างๆ บน Dashboard ?	◉	○	○	○	○

ความเหมาะสมในการเลือกชนิดของ visualization (เช่น กราฟหรือแผนภูมิ)เพื่อแสดงข้อมูลบน Dashboard ?	◉	○	○	○	○
Dashboard มีความเข้าใจง่าย ?	◉	○	○	○	○
ความสวยงามของ Dashboard ?	◉	○	○	○	○
ความต้องการแนะนำให้ผู้อื่นใช้งาน Dashboard ?	◉	○	○	○	○
ภาพรวมความพึงพอใจต่อ Dashboard ?	◉	○	○	○	○

ข้อเสนอแนะ

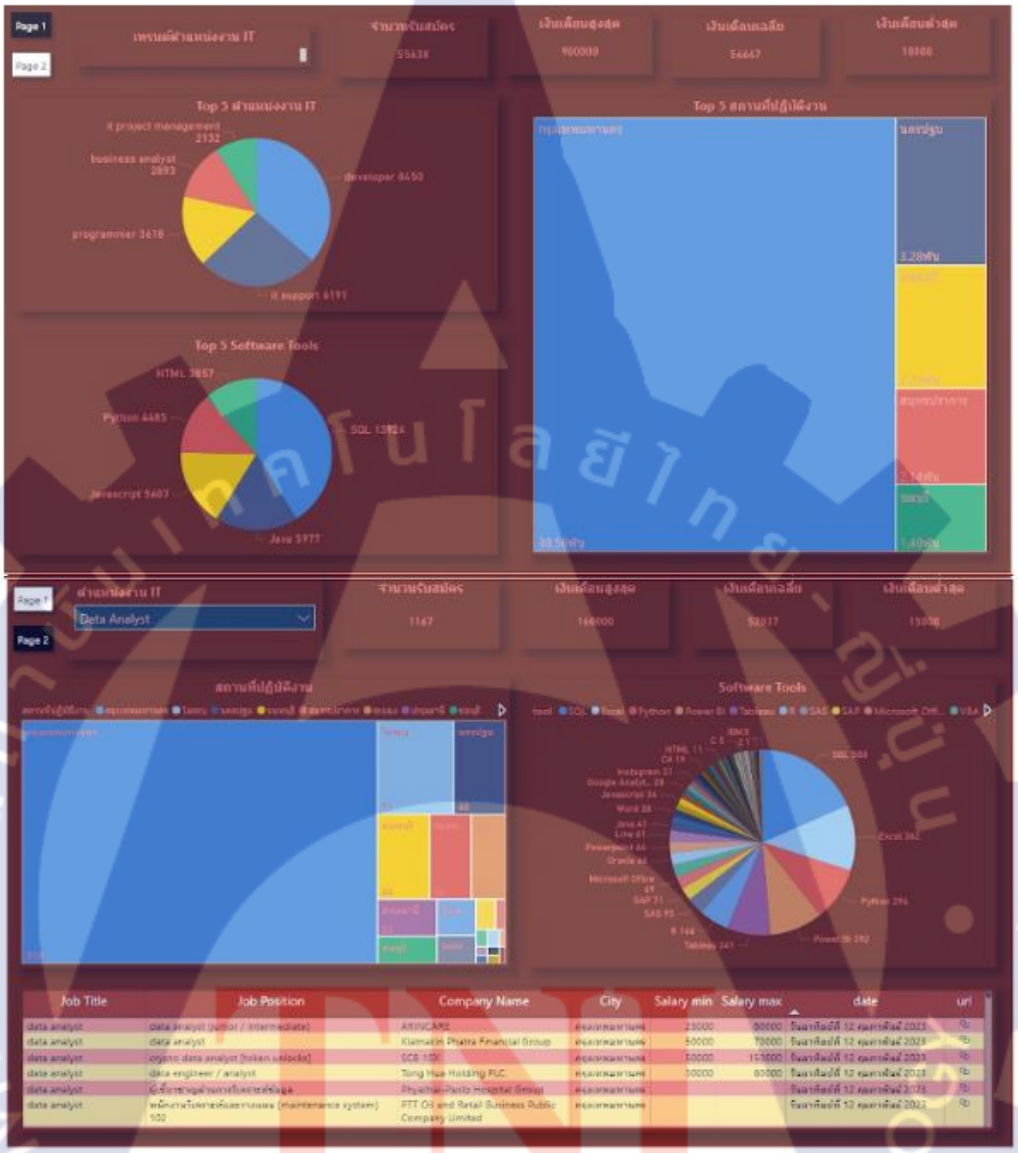
Your answer

Back Next Clear form

แบบประเมิน Dashboard เทรนด์ของอาชีพสายไอที ส่วนที่ 3 ความเหมาะสมในการออกแบบ Dashboard แบบที่ B

ลิ้งDashboard แบบที่ B
https://app.powerbi.com/view?r=eyJrIjoiODM4OWMyODMtYzgwNC00MTM0LTg4Y2MtZTU4NThlODYwODg0IiwidCI6IjFhNzg0NmJlLWUwMTktNDBkNS1iMDY0LTJjZjAxMDAyNzJhNSIsImMiOjEwfQ%3D%3D&pageName=ReportSection

ความเหมาะสมในการออกแบบ Dashboard แบบที่ B ? *



	มากที่สุด	มาก	ปานกลาง	ค่อนข้างน้อย	น้อย
ความเหมาะสมของสีพื้นหลังและสีตัวอักษรบน Dashboard ?	◉	○	○	○	○
ความเหมาะสมของขนาดตัวอักษรบน Dashboard ?	◉	○	○	○	○

ความเหมาะสมของฟอนต์ตัวอักษรบน Dashboard ?	◉	○	○	○	○
ความเหมาะสมในการวางตำแหน่งของ visualization (เช่น กราฟหรือแผนภูมิ) ต่างๆ บน Dashboard ?	◉	○	○	○	○
ความเหมาะสมในการเลือกชนิดของ visualization (เช่น กราฟหรือแผนภูมิ)เพื่อแสดงข้อมูลบน Dashboard ?	◉	○	○	○	○
Dashboard มีความเข้าใจง่าย ?	◉	○	○	○	○
ความสวยงามของ Dashboard ?	◉	○	○	○	○
ความต้องการแนะนำให้ผู้อื่นใช้งาน Dashboard ?	◉	○	○	○	○
ภาพรวมความพึงพอใจต่อ Dashboard ?	◉	○	○	○	○

ข้อเสนอแนะ

Your answer

Back Submit Clear form

ประวัติผู้เขียนสารนิพนธ์

ชื่อ – สกุล	นาย กานต์ สิริกานต์กุล
วัน - เดือน - ปี เกิด	19 สิงหาคม พ.ศ. 2537
ที่อยู่ปัจจุบัน	347 หมู่บ้านลาซาลเพลส ซอยลาซาล 24
	แขวงบางนา เขตบางนา กทม. 10260
E-mail	sr.karn_st@tni.ac.th

ประวัติการศึกษา

พ.ศ. 2566	วิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
	สถาบันเทคโนโลยีไทย-ญี่ปุ่น
พ.ศ. 2560	วิศวกรรมบัณฑิต สาขาเทคโนโลยียานยนต์
	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ประวัติการทำงาน

พ.ศ. 2560 - 2564	วิศวกรออกแบบรถดัดแปลง
	บริษัท ที เค ดี ไฟเบอร์ จำกัด