



การวิเคราะห์ข้อมูลเชิงลึก โดยใช้ข้อมูลรายงานปัญหาเส้นเลือดฟอยใน
กรุงเทพมหานคร

INSIGHT ANALYSIS USING CAPILLARIES (MICRO ISSUES) REPORT DATA IN
BANGKOK

นางสาวแสนมุก นันตะสุคนธ์

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย – ญี่ปุ่น
พ.ศ. 2565

การวิเคราะห์ข้อมูลเชิงลึก โดยใช้ข้อมูลรายงานปัญหาเส้นเลือดฟอยในกรุงเทพมหานคร
INSIGHT ANALYSIS USING CAPILLARIES (MICRO ISSUES) REPORT DATA IN
BANGKOK

นางสาวแสณมุก นันตะสุคนธ์

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย – ญี่ปุ่น

พ.ศ. 2565

คณะกรรมการสอบ

..... ประธานกรรมการสอบ
(อาจารย์.....)

..... กรรมการสอบ
(อาจารย์.....)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.ภาสกร อภิรักษ์รพินิต)

..... ประธานสหกิจศึกษาสาขาวิชา
(อาจารย์ สลิลดา ชีวกิติการ)

ลิขสิทธิ์ของสถาบันเทคโนโลยีไทย – ญี่ปุ่น

ชื่อโครงการ	การวิเคราะห์ข้อมูลเชิงลึก โดยใช้ข้อมูลรายงานปัญหาเส้นเลือด ฝอยในกรุงเทพมหานคร
	Insight Analysis Using Capillaries (Micro Issues) Report Data in Bangkok
ชื่อผู้เขียน	นางสาวเสน่ห์นันตะสุคนธ์
คณะวิชา	เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา	อาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต
พนักงานที่ปรึกษา	คุณอิติมา โภคสุพัฒน์
ชื่อบริษัท	บริษัท แอดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน)
ประเภทธุรกิจ/สินค้า	ให้บริการด้านดิจิทัลไลฟ์ มี 4 ธุรกิจคือ บริการโทรศัพท์เคลื่อนที่, อินเทอร์เน็ตความเร็วสูง, บริการลูกค้าองค์กรและดิจิทัลเซอร์วิส

บทสรุป

จากการปฏิบัติสหกิจศึกษา ณ บริษัท แอดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน) ในตำแหน่ง Business Data Analyst โดยภาระงานที่ได้รับมอบหมาย ได้แก่ การทำความสะอาดข้อมูล, การสร้างกราฟและ Dashboard, การทำงานกับข้อมูลขนาดใหญ่ (Big Data) รวมไปถึงการทำ Presentation สำหรับนำเสนอ จึงสามารถนำประสบการณ์เหล่านั้นมาประยุกต์ใช้กับการทำโครงการ สหกิจศึกษาวิเคราะห์ข้อมูลเชิงลึกนี้

จากการวิเคราะห์ข้อมูลเชิงลึก โดยใช้ข้อมูลรายงานปัญหาเส้นเลือดฝอยในกรุงเทพมหานคร ซึ่งเป็นข้อมูลที่เปิดให้ประชาชนแจ้งปัญหาขนาดย่ออย่าง ๆ ในกรุงเทพฯ เดิมพบว่า ข้อมูลประเภท “อื่น ๆ” มีจำนวนมากผิดปกติ จึงได้ทำการแก้ปัญหาโดยการใช้ Natural Language Processing (NLP) เข้ามาจัดการ จากนั้นจึงวิเคราะห์พบว่า ปัญหาที่ถูกแจ้งเข้ามามากเป็นอันดับที่ 1 คือ ปัญหาประเภท “ทางเท้า”

Project's name	Insight Analysis Using Capillaries (Micro Issues) Report Data in Bangkok
Writer	Sanmook Nantasukon
Faculty	Information Technology, Major Information Technology
Faculty Advisor	Paskorn Apirukvorapinit, Ph.D.
Job Supervisor	Thitima Poksupat
Company's name	Advanced Info Service PLC.
Business Type / Product	Provide digital life services, with 4 businesses: mobile phone service, high speed internet, corporate customer service and digital service.

Summary

Internship at Advanced Info Service PLC., in the position of Business Data Analyst. This role is responsible for cleaning data, preparing data, analyzing big data to find insights, visualizing them, and preparing presentations. Moreover, learning with real world of big data is to gain experience and knowledge which can be applied to this project.

From the insight analysis using Capillaries (Micro Issues) Report Data in Bangkok. This is public information for reporting various small-scale problems in Bangkok. Originally, there was an unusually large number of "Other" types of problem. Therefore, the problem was solved by using Natural Language Processing (NLP). Then the analysis found that the most reported problem is the type of "Footpath" problem.

กิตติกรรมประกาศ

จากการออกปฏิบัติงานสหกิจศึกษา ณ บริษัท แอดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน) ตั้งแต่วันที่ 1 มิถุนายน พ.ศ. 2565 ถึงวันที่ 30 กันยายน พ.ศ. 2565 เป็นเวลารวมทั้งสิ้น 4 เดือน ส่งผลให้ข้าพเจ้าได้รับความรู้ และเพิ่มพูนประสบการณ์นอกเหนือจากในตำราเรียน สิ่งเหล่านั้น ช่วยให้ข้าพเจ้าสามารถนำมาประยุกต์ใช้กับการทำโครงการสหกิจศึกษาฉบับนี้ได้สำเร็จ

เห็นอีสิ่งอื่นใด โครงการสหกิจศึกษาฉบับนี้และการปฏิบัติงานสหกิจศึกษาสำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือและสนับสนุนจากหลายฝ่าย ดังนี้ คุณธิติมา โภคสุพัฒน์ (Business Data Intelligence Analyst Manager), คุณวิชญุ ศรีแก้ว (Business Data Analyst), คุณวีรพงษ์ อรรฆย์ภูมิ (Business Data Analyst), คุณธนาวินท์ สร้อยอุป (Business Data Analyst), คุณจุมพล พงษ์สุวรรณ (Business Data Analyst) และบุคลากรท่านอื่นที่ไม่ได้กล่าวถึง ที่ได้ให้ความรู้ คำแนะนำ รวมไปถึง อาจารย์ ดร.ภาสกร อภิรักษ์รพินิต อาจารย์ที่ปรึกษา และอาจารย์ประจำคณะเทคโนโลยีสารสนเทศทุกท่าน ที่ได้มอบวิชาความรู้พื้นฐานอันเป็นทักษะในการออกปฏิบัติงานสหกิจศึกษาในครั้งนี้ ผ่านไปได้ด้วยดี

สุดท้ายนี้ ขอขอบพระคุณผู้สนับสนุนทุกท่านที่ก่อร่วมมาข้างต้น และบริษัท แอดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน) ที่ได้มอบโอกาสให้ข้าพเจ้าเข้ามาเป็นส่วนหนึ่งในการปฏิบัติงาน และปฏิบัติสหกิจศึกษา ตลอดจนเข้าร่วมโครงการต่าง ๆ ทำให้โครงการสหกิจศึกษาฉบับนี้และการปฏิบัติงานสหกิจศึกษาสำเร็จลุล่วงไปได้ด้วยดี หากโครงการสหกิจศึกษาฉบับนี้มีข้อผิดพลาดประการใด ข้าพเจ้าขอรับไว้ปรับปรุง และแก้ไขในโอกาสครั้งต่อไป

แสนมุก นันตะสุคนธ์
ผู้จัดทำโครงการ

สารบัญ

หน้า

บทสรุป	ก
Summary	ก
กิตติกรรมประกาศ.....	ค
สารบัญ	ง
สารบัญภาพ	ช
สารบัญตาราง	ภ
บทที่ 1 บหน้า	1
1.1 ชื่อและสถานที่ของผู้ประกอบการ	1
1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร	2
1.3 รูปแบบการจัดการองค์กรและการบริหารองค์กร	4
1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย.....	5
1.5 พนักงานที่ปรึกษาและตำแหน่งของพนักงานที่ปรึกษา.....	5
1.6 ระยะเวลาที่ปฏิบัติงาน	5
1.7 ที่มาและความสำคัญของปัญหา.....	5
1.8 วัตถุประสงค์หรือจุดมุ่งหมายของโครงการ	6
1.9 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย.....	6
1.10 นิยามศัพท์เฉพาะ	6
1.10.1 Data Cleaning	6
1.10.2 Data Pipeline	6
1.10.3 Web Scraping.....	6
1.10.4 Data Visualization	6
1.10.5 Data Dashboard	7

สารบัญ(ต่อ)

	หน้า
1.10.6 API (Application Programming Interface)	7
1.10.7 Parameter.....	7
บทที่ 2 ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน	8
2.1 ทฤษฎีที่ใช้ในการปฏิบัติงาน	8
2.1.1 Data Analytics Project Life Cycle.....	8
2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน	9
2.2.1 Databricks	9
2.2.2 Python.....	10
2.2.3 Pandas.....	11
2.2.4 Numpy	12
2.2.5 Pyspark.....	13
2.2.6 BeautifulSoup	14
2.2.7 PyThaiNLP.....	15
2.2.8 Wordcloud	16
2.2.9 Tableau.....	17
2.2.10 Visual Studio Code	18
2.2.11 HTML	19
2.2.12 Bootstrap.....	20
2.2.13 Github	21
2.2.14 Netlify.....	22

สารบัญ(ต่อ)

	หน้า
บทที่ 3 แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน.....	23
3.1 แผนงานการฝึกงาน	23
3.2 รายละเอียดที่นักศึกษาปฏิบัติในการฝึกงาน.....	24
3.2.1 Insurance Project	24
3.2.2 โครงการอื่น ๆ	27
3.3 ขั้นตอนการดำเนินโครงการสหกิจศึกษา	28
3.3.1 กำหนดหัวข้อโครงการและหาข้อมูลที่จะนำมาวิเคราะห์	28
3.3.2 รายละเอียดข้อมูลที่เลือก.....	29
3.3.3 ปัญหาที่พบจากข้อมูลที่เลือก.....	30
3.3.4 ดึงข้อมูลจาก Website (Web Scraping : API)	31
3.3.5 ทำความสะอาดข้อมูล (Data Cleaning)	34
3.3.6 แก้ไขปัญหา	35
3.3.7 วิเคราะห์ข้อมูลเชิงลึก	41
3.3.8 สร้าง Visualization และ Dashboard.....	43
3.3.9 นำ Dashboard เผยแพร่บน Website	44
บทที่ 4 สรุปผลการดำเนินงาน.....	46
4.1 ผลการดำเนินงานและผลการวิเคราะห์ข้อมูล.....	46
4.1.1 Dashboard 1 แสดงสรุปสถิติการแจ้งปัญหาเส้นเลือดฟ้อย.....	46
4.1.2 Dashboard 2 เทียบจำนวนสถานะในแต่ละเขต	48
4.1.3 Dashboard 3 จำนวนเพศที่แจ้งปัญหาทั้งหมด.....	49
4.1.4 Dashboard 4 จำนวนเพศที่แจ้งปัญหาในแต่ละเขต	50
4.1.5 Website https://sanmook-ntsk.netlify.app	51

สารบัญ(ต่อ)

	หน้า
บทที่ 5 บหสรุปและข้อเสนอแนะ.....	53
5.1 สรุปผลดำเนินโครงการ.....	53
5.2 ปัญหาและแนวทางในการแก้ไข.....	54
5.2.1 Response (HTTP Response) 500	54
5.2.2 Font path บน Databrick.....	55
5.3 ข้อเสนอแนะจากการดำเนินงาน.....	56
เอกสารอ้างอิง	57
ภาคผนวก	58
ประวัติผู้จัดทำโครงการ	77

สารบัญภาพ

หน้า

ภาพที่ 1.1	โลโก้ของบริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)	1
ภาพที่ 1.2	แผนที่บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน).....	1
ภาพที่ 1.3	AIS Sustainability Highlights	3
ภาพที่ 1.4	โครงการสร้างการบริหารงานของ AIS	4
ภาพที่ 1.5	โครงการสร้างกลุ่มธุรกิจของ AIS	4
ภาพที่ 2.1	Data Analytics Project Life Cycle Diagram	8
ภาพที่ 2.2	การใช้งาน Databricks.....	9
ภาพที่ 2.3	ตัวอย่างภาษา Python และ Python Logo.....	10
ภาพที่ 2.4	ตัวอย่างการใช้งาน Pandas.....	11
ภาพที่ 2.5	ตัวอย่างการใช้งาน Numpy	12
ภาพที่ 2.6	ตัวอย่างการใช้งาน Pyspark.....	13
ภาพที่ 2.7	ตัวอย่างการใช้งาน BeautifulSoup.....	14
ภาพที่ 2.8	Pythainlp Logo.....	15
ภาพที่ 2.9	ตัวอย่างการใช้งาน Pythainlp	15
ภาพที่ 2.10	Wordcloud	16
ภาพที่ 2.11	ตัวอย่างการใช้งาน Wordcloud for Python	16
ภาพที่ 2.12	ตัวอย่างการใช้งานโปรแกรม Tableau	17
ภาพที่ 2.13	Visual Studio Code Logo.....	18
ภาพที่ 2.14	ตัวอย่างการใช้งานโปรแกรม Visual Studio Code.....	18
ภาพที่ 2.15	ตัวอย่างภาษา HTML	19
ภาพที่ 2.16	ตัวอย่างการใช้งาน Bootstrap.....	20
ภาพที่ 2.17	ตัวอย่างการใช้งาน Github	21
ภาพที่ 2.18	ตัวอย่างการใช้งาน Netlify.....	22
ภาพที่ 3.1	ตัวอย่าง Code Scraping Data เบอร์โทรศัพท์อุปกรณ์ของบริษัทประกันภัย	24
ภาพที่ 3.2	ตัวอย่าง Code Clean Data	25
ภาพที่ 3.3	ตัวอย่างการใช้ Pyspark Query ข้อมูลบน Databrick.....	25
ภาพที่ 3.4	Visualization ที่สร้างด้วยโปรแกรม Tableau	26

สารบัญภาพ(ต่อ)

	หน้า
ภาพที่ 3.5 ตัวอย่าง Code สร้าง Sankey Diagram.....	27
ภาพที่ 3.6 Sankey Diagram ที่สร้าง	27
ภาพที่ 3.7 ข้อมูลรายงานปัญหาเส้นเลือดฟ้อยภายในเขตพื้นที่จังหวัดกรุงเทพฯ.....	28
ภาพที่ 3.8 ภาพจาก Website Traffy โครงการ Traffy x TeamChadChart.....	29
ภาพที่ 3.9 QR code สำหรับการแจ้งปัญหา	29
ภาพที่ 3.10 ภาพปัญหาประเภท “อื่น ๆ” ที่มีจำนวนมาก.....	30
ภาพที่ 3.11 Parameters ที่ traffy.in.th ระบุ	31
ภาพที่ 3.12 ตัวอย่าง Code สำหรับวน loop เพื่อดึงข้อมูลผ่าน API	32
ภาพที่ 3.13 ตัวอย่าง Code เพื่อคัดแยกส่วนที่ไม่ต้องการ	34
ภาพที่ 3.14 ตัวอย่างข้อมูล.....	34
ภาพที่ 3.15 จำนวนของข้อมูลแต่ละประเภทที่มีอยู่เดิม	35
ภาพที่ 3.16 ตัวอย่างการใช้งานการตัดคำ (Word Tokenization).....	36
ภาพที่ 3.17 การตัดคำ (Word Tokenization)	37
ภาพที่ 3.18 Code สร้าง Wordcloud ด้วยภาษา Python	38
ภาพที่ 3.19 Wordcloud จากข้อมูลปัญหาประเภท “อื่น ๆ”	38
ภาพที่ 3.20 ตัวอย่าง code สำหรับแบ่งประเภทเดิม	39
ภาพที่ 3.21 ลำดับของคำเรียงตามจำนวนที่ซ้ำ	39
ภาพที่ 3.22 Wordcloud จากข้อมูลที่เหลือจากการแบ่งประเภทเดิม.....	40
ภาพที่ 3.23 การนำข้อมูลมาแสดงเป็นกราฟเพื่อวิเคราะห์ข้อมูล	41
ภาพที่ 3.24 การนำกราฟมาเทียบกันเพื่อวิเคราะห์ข้อมูล.....	41
ภาพที่ 3.25 ตัวอย่าง code จำแนกเพศโดยคร่าว.....	42
ภาพที่ 3.26 การสร้าง Visualization และ Dashboard ใน Tableau	43
ภาพที่ 3.27 Tableau Public ตัวช่วยในการแสดงผล Dashboard บน Website	44
ภาพที่ 3.28 เมย์แพร์ Website ผ่าน Netlify.....	44
ภาพที่ 3.29 ตัวอย่าง Code ของ Website	45
ภาพที่ 4.1 Dashboard แสดงสรุปสถิติการแจ้งปัญหาเส้นเลือดฟ้อย	46
ภาพที่ 4.2 Dashboard แสดงการเทียบจำนวนสถานะในแต่ละเขต	48

สารบัญภาพ(ต่อ)

	หน้า
ภาพที่ 4.3 Dashboard แสดงจำนวนเพชรที่แจ้งปัญหาทั้งหมด	49
ภาพที่ 4.4 Dashboard แสดงจำนวนเพชรที่แจ้งปัญหานอกแต่ละเขต	50
ภาพที่ 4.5 Website https://sanmook-ntsk.netlify.app หน้าแรก	51
ภาพที่ 4.6 Website https://sanmook-ntsk.netlify.app หน้า Dashboard	51
ภาพที่ 4.7 Website https://sanmook-ntsk.netlify.app หน้า About	52
ภาพที่ 4.8 Website https://sanmook-ntsk.netlify.app ส่วน Footer	52
ภาพที่ 5.1 ตัวอย่าง HTTP Response 500 ที่ได้รับ	54
ภาพที่ 5.2 การติดตั้งและใช้งาน fontTools	55

สารบัญตาราง

หน้า

ตารางที่ 3.1 แผนการปฏิบัติสหกิจศึกษา.....	23
ตารางที่ 3.2 โครงสร้างข้อมูลรายงานปัญหาเส้นเลือดฟอย.....	33

บทที่ 1

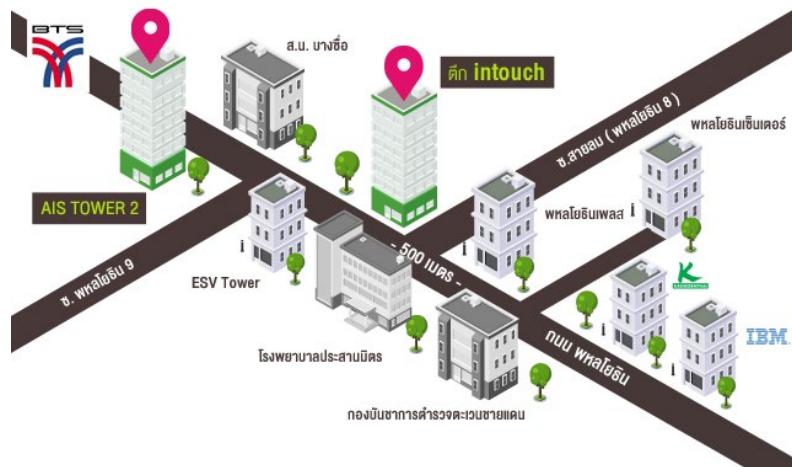
บทนำ

1.1 ชื่อและสถานที่ของผู้ประกอบการ

ชื่อสถานประกอบการ : บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)
ที่ตั้งสถานประกอบการ : อาคารเอไอเอส ทาวเวอร์ 2 เลขที่ 1291/1 ถ.พหลโยธิน
แขวงสามเสนใน เขตพญาไท กรุงเทพฯ 10400
โทรศัพท์ : 02-029-3247
โทรสาร : 02-299-5799
Email : enquiry@ais.co.th



ภาพที่ 1.1 โลโก้ของบริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)



ภาพที่ 1.2 แผนที่บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)

1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร

โลกกำลังเปลี่ยนแปลงอย่างรวดเร็ว เข้าสู่ยุคดิจิทัล ยุคที่ความเจริญด้านการสื่อสารและเทคโนโลยีต่างๆ เข้ามายึด主导ในชีวิตประจำวัน มีผลให้รูปแบบการใช้ชีวิต การเข้าถึงข้อมูล และความต้องการของผู้คน เปลี่ยนไป ชีวิตที่สอดคล้องกับสิ่งที่สามารถได้ทุกที่ทุกเวลา ทำให้ความสัมพันธ์ของผู้คนเข้าใกล้กันง่ายแค่เพียงปลายนิ้วสัมผัส เมื่อโลกถูกเชื่อมต่อให้เข้าใกล้กันมากขึ้น ได้เห็นและ อัพเดท ข้อมูลต่างๆ ของนานาประเทศเสมือนโลกที่ถูกย่อให้เล็กลงและเป็นโลกที่ทุกอย่าง รวดเร็ว จนเข้าใกล้คำว่า “ทันที” เป็นโลกที่เมื่อรู้ตัวอีกทีวันนี้ ก็กลายเป็นเมื่อวาน

เอไอเอส จึงก้าวไปอีกขั้นด้วยการนำการให้บริการทางดิจิทัลเต็มรูปแบบ มาตอบสนองความต้องการที่แตกต่างของทุกคนอย่างมีประสิทธิภาพทั้ง Mobile, Fixed Broadband และ Digital Service

Mobile

เครือข่ายมือถือคุณภาพที่ครอบคลุม ทุกพื้นที่พร้อมการพัฒนาอย่างต่อเนื่อง เพื่อคุณภาพการสื่อสารที่ดีที่สุดของ คนไทย เอไอเอสมุ่งมั่น ให้ทุกการสื่อสาร ไร้รอยต่อ (Seamless) ด้วยการเชื่อมโยง เทคโนโลยี 3G และ Super WiFi เพื่อ ประสิทธิภาพสูงสุดทั้งวันนี้และอนาคต

Fixed Broadband

เติมเต็มการใช้ชีวิตดิจิทัลอย่างเต็มรูปแบบ ตอบรับการใช้งานในบ้านด้วยบริการ AIS Fibre Broadband ความเร็วสูง บนเทคโนโลยีที่ดีที่สุด รายแรก รายเดียว (Fibre Optic) ให้คุณเชื่อมต่อโลกอินเทอร์เน็ตได้อย่างรวดเร็ว ไม่มีสะดุด ที่มาพร้อมกับ AIS PLAYBOX กล่องเดียว ครอบทุกความบันเทิง สนุกเต็มอิ่มกับ ความบันเทิง ชั้นนำระดับโลกทั้งภาพยนตร์ ซีรีส์ ทั่วไทย และต่างประเทศ กีฬา รายการทีวีสดและย้อนหลัง และ卡拉โอเกะ มากกว่า 14,000 เพลง พร้อม ความคม ชัด 4K ULTRA HD ความสูงที่คุณสนุก ได้ไม่รู้จบ โดยวางแผนการให้บริการ ครอบคลุม กว่า 10 ล้านครัวเรือน ทั่วประเทศไทย ในปี 2560

Digital Service

เอไอเอสให้มือถือของคุณเป็นได้มากกว่าแค่ โทรศัพท์ ครอบคลุม ตอบสนองต่อ ไลฟ์สไตล์ ยุคดิจิทัลเพื่อคุณ ด้วยการให้บริการ Digital Content หลากหลายรูปแบบ ทั้งที่ เป็น Mobile Application ที่เติมไปด้วย คุณภาพเพื่อตอบโจทย์ความต้องการ ทั้ง การทำงานและความบันเทิงของคุณ รวมถึง AIS Play Box ที่จะทำให้คุณสะดวก快捷 ได้และสนุกกับความบันเทิงชั้นนำระดับโลก ทั้ง ภาพยนตร์ ซีรีส์ กีฬา รายการทีวี และ卡拉โอเกะ เอไอเอสพร้อมแล้วที่จะ รวบรวม Application คุณภาพเหล่านี้มา เพื่อคุณ

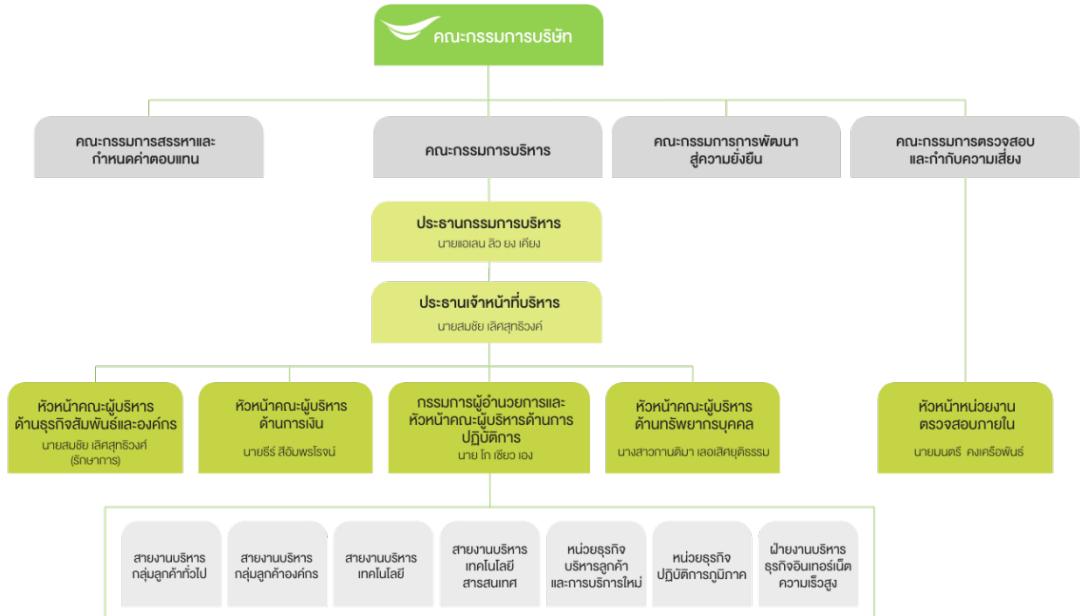
Quality Service and Lifestyle

มาตรฐานบริการที่มีคุณภาพ การดูแล เอาใจใส่ ทุกคนตามต้องการที่จะตอบทุก ไลฟ์สไตล์ และ รูปแบบการใช้ชีวิตที่ หลากหลาย เอไอเอสยกระดับการบริการ พร้อมเป็นผู้นำด้านการดูแลลูกค้า แห่ง ยุคดิจิทัล ที่จะทำให้คุณได้รับประสบการณ์ สุดพิเศษ กับ AIS Serenade และ AIS LIVE 360 ที่ จะเติมเต็มความสุขครบถ้วน องค์การใช้ชีวิต ตอบสนองความต้องการ ของคุณตลอด 24 ชั่วโมง ตลอดจนการ บริการที่ใส่ใจผ่านเทคโนโลยี พร้อม พนักงานที่มีคุณภาพผ่านศูนย์บริการ ที่ถูก พัฒนาขึ้นเพื่ออำนวยความสะดวก ให้กับลูกค้าและการบริการอย่างทั่วถึง [1]



ภาพที่ 1.3 AIS Sustainability Highlights

1.3 รูปแบบการจัดการองค์กรและการบริหารองค์กร



ภาพที่ 1.4 โครงสร้างการบริหารงานของ AIS



ภาพที่ 1.5 โครงสร้างกลุ่มธุรกิจของ AIS

1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย

แผนก	:	AIS Digital Life Co., Ltd. (ADVS)
ตำแหน่ง	:	Business Data Analyst
หน้าที่งานที่ได้รับมอบหมาย :		ทำความสะอาดข้อมูล เตรียมพร้อมข้อมูลก่อนการทำงาน วิเคราะห์ข้อมูล สร้างแผนภูมิหรือกราฟ รวมไปถึงสร้าง Dashboard สำหรับการนำเสนอ

1.5 พนักงานที่ปรึกษาและตำแหน่งของพนักงานที่ปรึกษา

พนักงานที่ปรึกษา	:	คุณธิติมา โภคสุพัฒน์
ตำแหน่ง	:	Business Data Intelligence Analyst Manager
แผนก	:	AIS Digital Life Co., Ltd. (ADVS)

1.6 ระยะเวลาที่ปฏิบัติงาน

ระยะเวลาปฏิบัติสหกิจศึกษา ตั้งแต่วันที่ 1 มิถุนายน พ.ศ. 2565 ถึงวันที่ 30 กันยายน พ.ศ. 2565 โดยมีการปฏิบัติในวันจันทร์ ถึง วันศุกร์ ตั้งแต่เวลา 8:30 น. ถึง 17:30 น.

1.7 ที่มาและความสำคัญของปัญหา

จากการที่ได้ออกปฏิบัติงานสหกิจศึกษา ณ บริษัท เออดวนซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) ตำแหน่ง Business Data Analyst ทำให้ข้าพเจ้าได้เรียนรู้กระบวนการทำงานในตำแหน่ง Business Data Analyst มีลักษณะงานตัวอย่างเช่น การทำความสะอาดข้อมูล การเตรียมพร้อมข้อมูลก่อนการทำงาน วิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลเชิงลึก การสร้างแผนภูมิหรือกราฟ รวมไปถึงการสร้าง Dashboard สำหรับการนำเสนอ ส่งผลให้ข้าพเจ้าสามารถนำกระบวนการเหล่านี้มาประยุกต์ใช้กับข้อมูลต่าง ๆ ได้

ข้อมูลในที่นี้ที่ข้าพเจ้าเลือกมาคือ ข้อมูลการรายงานปัญหาเส้นเลือดฟอย หรือ ข้อมูลการรายงานปัญหาขนาดย่อຍภัยในเขตพื้นที่กรุงเทพมหานคร ที่มีรายงานเข้ามาเป็นจำนวนมาก อีกทั้ง มีข้อมูลดังกล่าว�ังสามารถวิเคราะห์ในเชิงลึกและหลากหลายมุมมองได้มากขึ้นและยังเป็นประโยชน์ต่อส่วนรวมได้ จึงเหมาะสมแก่การนำกระบวนการที่ข้าพเจ้าได้เรียนรู้เหล่านั้นมาวิเคราะห์หาข้อมูลเชิงลึก

1.8 วัตถุประสงค์หรือจุดมุ่งหมายของโครงการ

1. เพื่อวิเคราะห์หาข้อมูลเชิงลึก จากข้อมูลรายงานปัญหาเส้นเลือดฟ้อยในกรุงเทพมหานคร
2. เพื่อสร้าง Data Visualization
3. เพื่อสร้าง Data Dashboard

1.9 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

1. สามารถวิเคราะห์หาข้อมูลเชิงลึกได้
2. สามารถสร้าง Data Visualization และ Data Dashboard ได้
3. ได้รับประสบการณ์จากการปฏิบัติงานและโครงการ

1.10 นิยามคัพท์เฉพาะ

1.10.1 Data Cleaning

คือการทำความสะอาดข้อมูล หรือการทำข้อมูลให้สมบูรณ์ เป็นกระบวนการตรวจสอบและการแก้ไข รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล เพื่อให้ข้อมูลมีคุณภาพ [2]

1.10.2 Data Pipeline

คือการย้าย ถ่ายโอนข้อมูลจากแหล่งข้อมูลใดๆ (Sources of Data) ซึ่งอาจจะมาจากแหล่งข้อมูลเดียวหรือจากหลายแหล่ง ไปยังอีกแหล่งข้อมูลหนึ่ง (Target/Destination Location) โดยอาจจะมีการประมวลผลข้อมูลหรือไม่มีการประมวลผลใดๆ ระหว่างการเคลื่อนย้ายข้อมูลก็ได้ [3]

1.10.3 Web Scraping

เป็นวิธีดึงข้อมูลต่าง ๆ จากหน้า Website ที่เปิดเผยแพร่ต่อสาธารณะ เช่น ราคา ข้อความ รูปภาพ ข้อมูลติดต่อ และอื่นๆ [4]

1.10.4 Data Visualization

คือการแสดงข้อมูลในรูปแบบรูปภาพนิ่ง เช่น รูปกราฟแท่ง รูปพายกราฟ เป็นต้น เพื่อให้มนุษย์สามารถเข้าใจข้อมูลได้ง่ายขึ้น มีจุดประสงค์หลัก ๆ ได้แก่ ถ่ายทอดสิ่งที่ข้อมูล ต้องการสื่อสารอย่างชัดเจน ช่วยอธิบายแนวโน้มทางสถิติ และแสดงรูปแบบเฉพาะในข้อมูลที่ยากจะเห็นด้วยตาเปล่า [5]

1.10.5 Data Dashboard

คือเครื่องมือในการจัดการข้อมูลแบบหนึ่งที่ติดตาม วิเคราะห์ และแสดงตัวชี้วัดความสำเร็จของงาน (KPIs – Key Performance Indicators) หรือ ดัชนีชี้วัดอื่นๆ รวมถึงจุดสำคัญของข้อมูลอุปกรณ์ในรูปแบบ Visual ทำให้สามารถเห็นความเป็นไปของธุรกิจหรือสิ่งที่สนใจได้แบบ Real-Time นอกจากนี้การที่มีข้อมูลขนาดใหญ่ และต้องการหา Insight การใช้ Data Dashboard จะทำให้เห็น Insight ของข้อมูลได้หลากหลายมุมมอง เพื่อนำไปใช้วิเคราะห์ต่ออย่างดียิ่งขึ้น [6]

1.10.6 API (Application Programming Interface)

แปลเป็นภาษาไทยตรงตัวได้ว่า “ส่วนต่อประสานโปรแกรมประยุกต์” หรือก็คือ “ตัวกลาง” ที่ให้ Website, Software, Application สามารถสื่อสารและทำงานร่วมกันได้โดยเชื่อมถึงข้อมูลกับฐานข้อมูล (Database) ของอีกฝ่าย ในรูปแบบของ Code ที่เป็นคำสั่งเรียกใช้ API [7]

1.10.7 Parameter

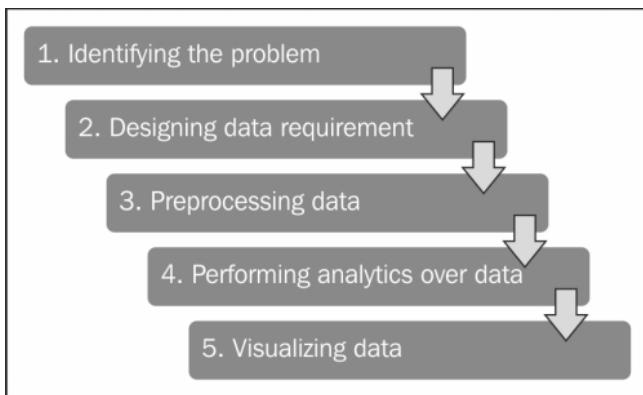
คือค่าหรือตัวแปรที่ส่งเข้าไปในฟังก์ชันของภาษา Python เพื่อให้การทำงานในฟังก์ชันนั้นสมบูรณ์ [8]

บทที่ 2

ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

2.1 ทฤษฎีที่ใช้ในการปฏิบัติงาน

2.1.1 Data Analytics Project Life Cycle



ภาพที่ 2.1 Data Analytics Project Life Cycle Diagram

ในการทำโครงการ Data Analytics มีขั้นตอนการดำเนินงานเฉพาะทางที่ควรปฏิบัติตาม เพื่อที่จะได้ผลลัพธ์ออกมาตามที่คาดหวัง ขั้นตอนดังกล่าวคือ Data Analytics Project Life Cycle ที่มีกระบวนการอยู่ 5 ขั้นตอนดังภาพที่ 2.1 ประกอบไปด้วย

1. Identifying The Problem: ระบุปัญหา เพื่อที่จะสามารถหาทางแก้ไขได้ตรงจุด
2. Designing Data Requirements: ออกแบบข้อมูลที่ต้องการ เพื่อที่จะได้ทราบว่าต้องใช้แหล่งข้อมูลใดมาแก้ปัญหา
3. Preprocessing Data: เตรียมความพร้อมของข้อมูล ก่อนการลงมือดำเนินการกับข้อมูล เช่น Data Cleaning, Data Aggregation เป็นต้น เพื่อให้ข้อมูลที่นำไปใช้ต่อเกิดประสิทธิภาพมากที่สุด
4. Performing Analytics Over Data: ลงมือวิเคราะห์ข้อมูล เช่น การหาข้อมูลเชิงลึกต่างๆ
5. Visualizing Data: นำข้อมูลออกมาแสดงผลในรูปของ Visualization หรือแสดงผลในรูปของแผนภูมิ กราฟ เพื่อประกอบการนำเสนอให้ผู้ฟังเข้าใจง่าย

2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน

2.2.1 Databricks

Databricks เป็นระบบ Unified Analytics Platform ที่ทำงานบน Apache Spark โดยออกแบบมาสำหรับระบบ Cloud โดยเฉพาะ มีฟีเจอร์ต่างๆมาให้ครบครัน เช่น One-click deployment, Auto-scaling และมีระบบ Optimized Databricks Runtime ที่ช่วยเร่งประสิทธิภาพของ Spark Job เมื่อทำงานบน Cloud ได้เร็วขึ้น 10-100 เท่า ในขณะเดียวกันยังมีเครื่องมือที่ช่วยอำนวยความสะดวกในการใช้งาน เช่น Interactive Notebook Environment, Monitoring Tools และ Security Control เพื่อช่วยตอบโจทย์การใช้งาน Apache Spark ในองค์กร ที่มีผู้ใช้งานจำนวนมากได้

ในครั้งนี้ Microsoft ได้ประกาศความร่วมมือกับ Databricks ไม่เพียงแค่เชื่อมต่อบริการระหว่าง Azure กับ Databricks เท่านั้น แต่เป็นการพัฒนา Databrick เป้าเป็น Service หนึ่งของ Microsoft Azure ซึ่งช่วยให้ผู้ใช้งานสามารถบริหารจัดการระบบผ่านทาง Console ของ Microsoft Azure ได้ทันที อีกทั้งยังสามารถเชื่อมต่อกับ Service อื่นๆของ Microsoft ได้อีกด้วย

```

1 pip install fontTools
Python interpreter will be restarted.
Collecting fontTools
  Downloading fonttools-4.35.0-py3-none-any.whl (947 kB)
Installing collected packages: fonttools
Successfully installed fontTools-4.35.0
Python interpreter will be restarted.

Command took 9.58 seconds -- by sammooks@ais.co.th at 8/17/2022, 11:16:02 AM on TRAINEE_DA

1 pip install Brotli
Python interpreter will be restarted.
Collecting Brotli
  Downloading Brotli-1.0.9-cp38-cp38-manylinux1_x86_64.whl (357 kB)
Installing collected packages: Brotli
Successfully installed Brotli-1.0.9
Python interpreter will be restarted.

Command took 5.55 seconds -- by sammooks@ais.co.th at 8/17/2022, 11:16:19 AM on TRAINEE_DA

1 import urllib.request
2 import fontTools.ttLib.woff2
3 import urllib3
4 import pandas as pd
5 import numpy as np

```

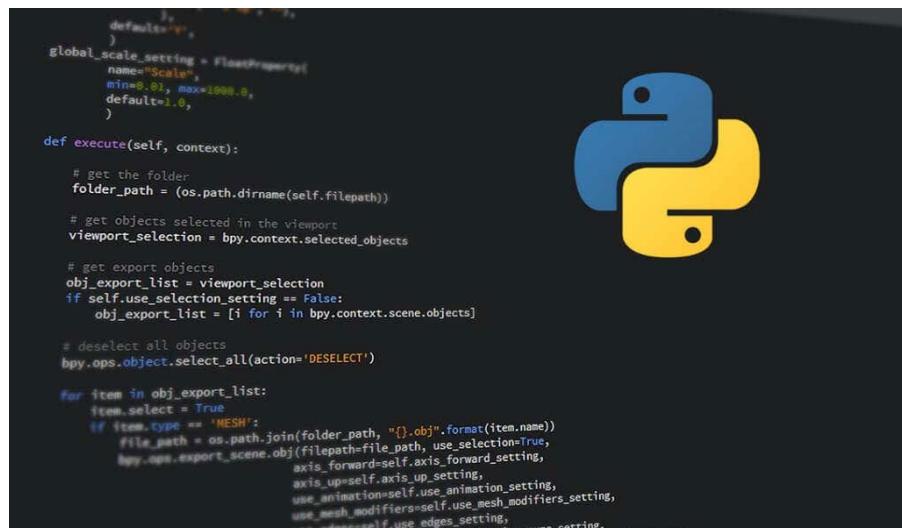
ภาพที่ 2.2 การใช้งาน Databricks

2.2.2 Python

ไพธอน หรือ Python เป็นภาษาเขียนโปรแกรมระดับสูง (High-Level Language) ที่ใช้กันอย่างกว้างขวางในการเขียนโปรแกรมเพื่อวัตถุประสงค์ทั่วไป ภาษา Python นั้นเป็นภาษาแบบการตีความ (Interpret) ที่ถูกออกแบบโดยมีปรัชญาที่จะทำให้ Code หรือรหัสทางภาษาหนึ่งอ่านได้ง่ายขึ้น และทำให้โปรแกรมเมอร์สามารถเข้าใจโครงสร้างของภาษาและแนวคิดการเขียน Code โดยใช้บรรทัดที่น้อยลงกว่าภาษาอื่น เช่น C, C++ และ Java

Python นั้นมีคุณสมบัติเป็นภาษาเขียนโปรแกรมแบบไดนามิกส์และมีระบบการจัดการหน่วยความจำอัตโนมัติและสนับสนุนการเขียนโปรแกรมหลายรูปแบบ ที่ประกอบไปด้วย การเขียนโปรแกรมเชิงวัตถุ Imperative การเขียนโปรแกรมแบบฟังก์ชัน และการเขียนโปรแกรมแบบขั้นตอน ภาษา Python มีไลบรารีที่ครอบคลุมการทำงานอย่างหลากหลาย ตัวแปรในภาษา Python นั้นมีให้ใช้ในหลายระบบปฏิบัติการ ทำให้ Code ของภาษา Python สามารถทำงานในระบบต่าง ๆ ได้อย่างกว้างขวาง ซึ่งแรกเริ่มนั้น Python ถูกพัฒนามาจาก CPython ซึ่งเป็นโปรแกรมแบบเปิด (Open Source) และมีชุมชนจำนวนมากเป็นต้นแบบในการพัฒนา เนื่องจากมันได้มีการนำไปพัฒนากระจายไปอย่างหลากหลาย CPython นั้นจึงถูกจัดการโดยองค์กรไม่แสวงหาผลกำไรซึ่งในเวลาต่อมาได้มีชื่ออย่างเป็นทางการว่า Python Software Foundation (PSF)

ภาษา Python ยังเป็นซอฟต์แวร์ประเภทโอเพนซอร์ส (Open source) หมายความว่าสามารถนำชอร์สโค้ด (Source Code) มาดัดแปลง แก้ไขได้ทั้งหมด โดยไม่จำเป็นต้องขออนุญาต และที่สำคัญสามารถนำไปใช้งานได้ฟรี โดยไม่มีค่าใช้จ่ายเรื่องค่าลิขสิทธิ์ใดๆ



ภาพที่ 2.3 ตัวอย่างภาษา Python และ Python Logo

2.2.3 Pandas

Pandas คือ หนึ่งใน Library สำคัญของภาษา Python เริ่มพัฒนาโดย Wes McKinney นักพัฒนาซอฟต์แวร์ชาวอเมริกัน ปัจจุบัน Pandas เป็น Open Source ให้ทุกคนสามารถใช้ได้แบบฟรีๆ

Pandas มีความสามารถในการจัดการ และวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพตั้งแต่ ข้อมูลขนาดเล็กไปจนถึงข้อมูลขนาดใหญ่ ทำให้ Pandas ตอบโจทย์งานในยุคที่ข้อมูลมีขนาดใหญ่มาก ขึ้นได้ ไม่มีปัญหาติดขัดเหมือนกับ Spreadsheets อื่นๆ (เช่น Excel หรือ Google Sheets ซึ่งจะทำงานได้ช้าลงหากข้อมูลมีขนาดใหญ่ขึ้น)

นอกจากนี้ เมื่อเปรียบเทียบกับ Tools วิเคราะห์ข้อมูลอื่นๆ อย่าง Excel หรือ Google Sheets อาจไม่ตอบโจทย์เต็มที่ หากต้องการเชื่อมต่อกับแหล่งข้อมูลบางประเภท หรือทำ Automation (ระบบจัดการอัตโนมัติ) ในขณะที่ Pandas ซึ่งเป็นส่วนหนึ่งของ Python นั้นสามารถใช้การเขียนโค้ด เพื่อปรับแต่ง หรือเชื่อมต่อกับโปรแกรมอื่นๆ ได้สะดวกนั่นเอง

```
In [38]: import pandas as pd

one = pd.DataFrame({
    'Name': ['Amber', 'Jack', 'Brown', 'Smith', 'Young'],
    'subject_id':['sub1','sub2','sub4','sub6','sub5'],
    'Marks_scored':[93,90,82,64,71]},
    index=[1,2,3,4,5])

two = pd.DataFrame({
    'Name': ['Ben', 'Cole', 'Sam', 'Tom', 'Martial'],
    'subject_id':['sub2','sub4','sub3','sub6','sub5'],
    'Marks_scored':[96,80,73,77,81]},
    index=[1,2,3,4,5])
print (pd.concat([one,two]))
```

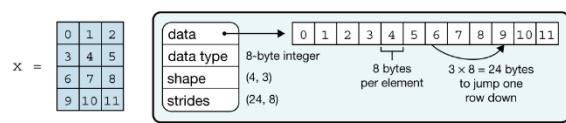
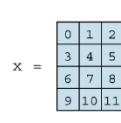
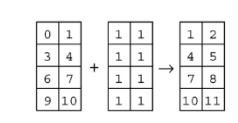
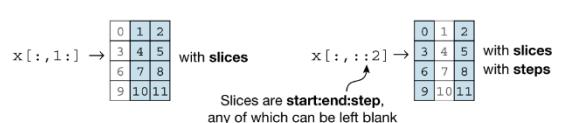
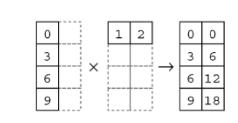
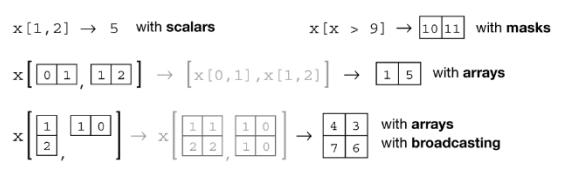
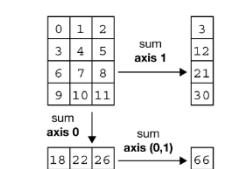
	Name	subject_id	Marks_scored
1	Amber	sub1	93
2	Jack	sub2	90
3	Brown	sub4	82
4	Smith	sub6	64
5	Young	sub5	71
1	Ben	sub2	96
2	Cole	sub4	80
3	Sam	sub3	73
4	Tom	sub6	77
5	Martial	sub5	81

ภาพที่ 2.4 ตัวอย่างการใช้งาน Pandas

2.2.4 Numpy

NumPy เป็น Library พื้นฐานที่ใช้คำนวณทางคณิตศาสตร์ด้วยภาษา Python สามารถคำนวณ หรือ ดำเนินการทางตรรกะใน Array หลายมิติ หรือ Matrix ได้อย่างรวดเร็ว เพราะ Library เขียนด้วยภาษา C ที่ Compile ไว้แล้ว

NumPy Arrays สามารถคำนวณและดำเนินการทางตรรกะใน Matrix, Array หลายมิติ และ Array ได้อย่างรวดเร็วมากกว่า Python Lists ในการใช้งาน NumPy Arrays จะประหยัด Memory ได้มากกว่าใช้ Python Lists NumPy Arrays มีขนาดคงที่เมื่อสร้าง ซึ่งแตกต่างจาก Python Lists (ซึ่งสามารถขยายได้แบบไม่มีกิจ) การเปลี่ยนขนาดของ ndarray จะสร้างอาร์เรย์ใหม่และลบต้นฉบับ การเปลี่ยนแปลงข้อมูลใน NumPy Arrays ก็สามารถทำได้เร็วกว่า Python Lists NumPy Arrays สามารถเข้าถึงข้อมูลภายใต้เร็วกว่า Python Lists

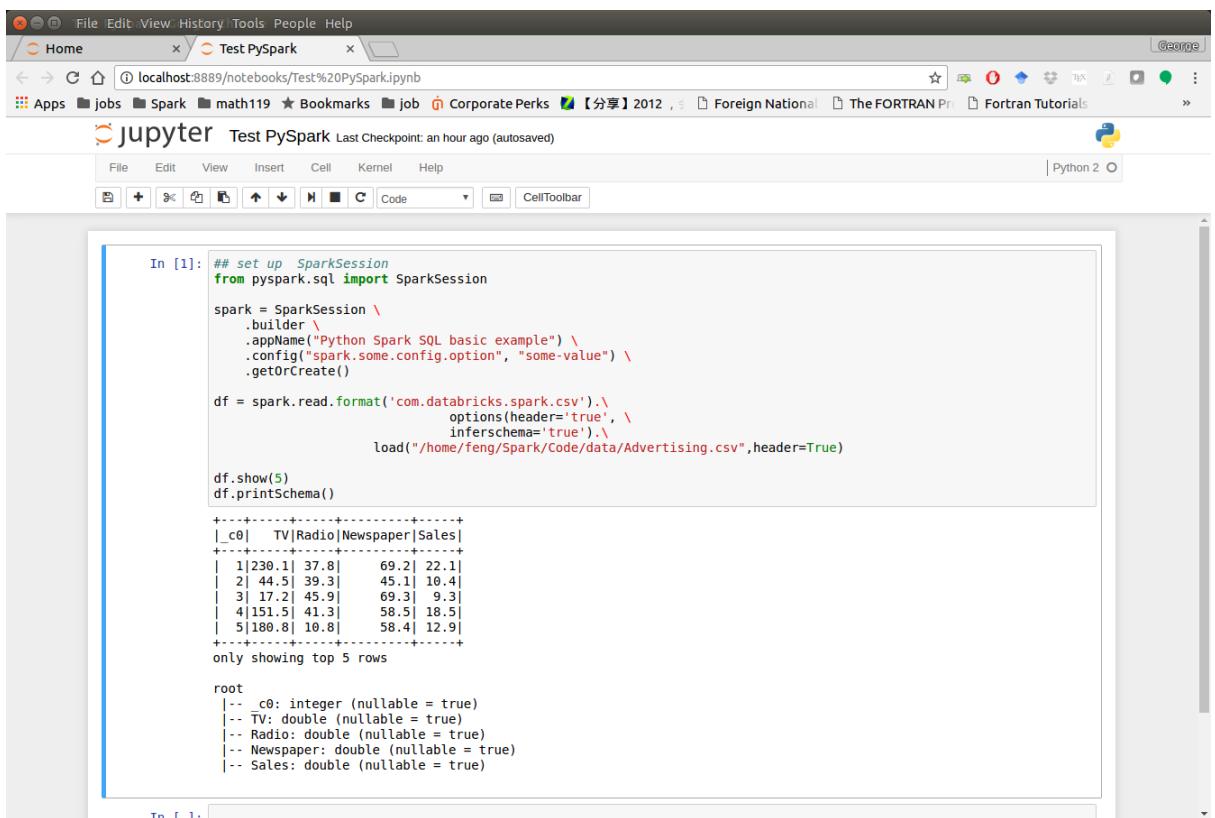
a Data structure  <p><code>x = </code> </p> <p>data → 0 1 2 3 4 5 6 7 8 9 10 11 data type 8-byte integer shape (4, 3) strides (24, 8)</p> <p>8 bytes per element $3 \times 8 = 24$ bytes to jump one row down</p>	d Vectorization  <p>$\begin{bmatrix} 0 & 1 \\ 3 & 4 \\ 6 & 7 \\ 9 & 10 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \\ 10 & 11 \end{bmatrix}$</p>	g Example <pre>In [1]: import numpy as np In [2]: x = np.arange(12) In [3]: x = x.reshape(4, 3)</pre>
b Indexing (view)  <p>$x[:, 1:] \rightarrow \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{bmatrix}$ with slices $x[:, ::2] \rightarrow \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{bmatrix}$ with slices with steps Slices are <code>start:end:step</code>, any of which can be left blank</p>	e Broadcasting  <p>$\begin{bmatrix} 0 \\ 3 \\ 6 \\ 9 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 \\ 3 & 6 \\ 6 & 12 \\ 9 & 18 \end{bmatrix}$</p>	<pre>In [4]: x Out[4]: array([[0, 1, 2], [3, 4, 5], [6, 7, 8], [9, 10, 11]]) In [5]: np.mean(x, axis=0) Out[5]: array([4.5, 5.5, 6.5]) In [6]: x = x - np.mean(x, axis=0) In [7]: x Out[7]: array([[-4.5, -4.5, -4.5], [-1.5, -1.5, -1.5], [1.5, 1.5, 1.5], [4.5, 4.5, 4.5]])</pre>
c Indexing (copy)  <p>$x[1, 2] \rightarrow 5$ with scalars $x[x > 9] \rightarrow [10, 11]$ with masks $x[[0, 1], [1, 2]] \rightarrow [[x[0, 1], x[1, 2]]]$ → [1, 5] with arrays $x[[1, 2], [1, 0]] \rightarrow x[[1, 1], [1, 0]]$ → [4, 3] with arrays with broadcasting</p>	f Reduction  <p>$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{bmatrix}$ sum axis 1 → [3, 12, 21, 30] sum axis 0 → [18, 22, 26] sum axis (0,1) → 66</p>	

ภาพที่ 2.5 ตัวอย่างการใช้งาน Numpy

2.2.5 Pyspark

Pyspark เป็นเครื่องมือหนึ่งที่เกิดจากการรวมตัวกันระหว่าง Apache Spark กับ Python ซึ่งทำให้สามารถเขียน Python ใน Spark ได้ หรือเรียกว่า易于其คือเป็น Python API ของ Apache Spark นอกจากนี้ Pyspark ยังช่วยให้ทำงานกับ RDDs (Resilient Distributed Datasets) ได้ง่ายขึ้น

Apache Spark เป็นเครื่องมือหนึ่งที่ถูกดีไซน์มาให้ใช้งานแบบทำงานกลุ่มได้ โดยที่เชื่อมต่อระบบการทำงานของคอมพิวเตอร์เข้าด้วยกัน หรือเรียกว่า Cluster Computing Platform ซึ่งสามารถกระจายงานที่ต้องทำไปยังเครื่องอื่นๆภายในระบบ ทำให้สามารถประมวลผลข้อมูลขนาดใหญ่แบบเต็มประสิทธิภาพ หรือแบบ Real-Time ไปพร้อมๆกันได้ จุดเด่นของ Apache Spark คือ Fast และ General-Purpose



The screenshot shows a Jupyter Notebook interface with the title "Test PySpark". The code cell (In [1]) contains Python code to set up a SparkSession and read a CSV file. The output shows the schema and top 5 rows of the DataFrame.

```

In [1]: ## set up SparkSession
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .appName("Python Spark SQL basic example") \
    .config("spark.some.config.option", "some-value") \
    .getOrCreate()

df = spark.read.format('com.databricks.spark.csv') \
    .options(header='true', \
            inferSchema='true') \
    .load("/home/feng/Spark/Code/data/Advertising.csv", header=True)

df.show(5)
df.printSchema()

```

c0	TV	Radio	Newspaper	Sales
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9

only showing top 5 rows

```

root
|-- c0: integer (nullable = true)
|-- TV: double (nullable = true)
|-- Radio: double (nullable = true)
|-- Newspaper: double (nullable = true)
|-- Sales: double (nullable = true)

```

ภาพที่ 2.6 ตัวอย่างการใช้งาน Pyspark

2.2.6 BeautifulSoup

BeautifulSoup คือ Python Module ที่ใช้สำหรับการดึงข้อมูลออกมาจาก HTML และ XML หรืออีกทางหนึ่งที่มักเรียกว่าคือการสกัด (Extract) ข้อมูลออกมา นับเป็น Library ภาษา Python ที่ใช้สำหรับงาน Web Scraping ที่ได้รับความนิยมเป็นอย่างมาก เพราะใช้งานง่าย

```

1 from bs4 import BeautifulSoup
2 import requests
3 import csv
4 import pandas as pd
5 import json
6
7 url = "https://publicapi.traffy.in.th/share/teamchadchart/search?&offset=0"
8
9 res = requests.get(url)
10 res.encoding = "utf-8"
11 print(res)

```

<Response [201]>

Command took 18.92 seconds -- by sanmooks@ais.co.th at 7/19/2022, 2:34:18 PM on TRAINEE_DA

Cmd 5

```

1 url = "https://publicapi.traffy.in.th/share/teamchadchart/search?&offset=0"
2 res = requests.get(url)
3 soup = BeautifulSoup(res.text, 'html.parser')
4 text = soup.get_text()
5
6 df_sum = pd.json_normalize(json.loads(text)['results'])

```

Command took 0.44 seconds -- by sanmooks@ais.co.th at 7/19/2022, 2:34:40 PM on TRAINEE_DA

Cmd 6

```

1 n = 1000
2 while n <= 90000:
3     url = f"https://publicapi.traffy.in.th/share/teamchadchart/search?&offset={n}"
4
5     res = requests.get(url)
6     print(res)
7     print(n)
8     if res.status_code != 201:
9         continue
10    else:
11        soup = BeautifulSoup(res.text, 'html.parser')
12        text = soup.get_text()
13        df = pd.json_normalize(json.loads(text)['results'])
14        df_sum = df_sum.append(df, ignore_index=True)
15        n += 1000

```

ภาพที่ 2.7 ตัวอย่างการใช้งาน BeautifulSoup

2.2.7 PyThaiNLP

PyThaiNLP คือไลบรารี Python สำหรับงานด้านการประมวลผลข้อมูลภาษาไทย พัฒนาขึ้นมาโดยคนไทย มีฟังก์ชันที่มีประโยชน์มาก many สำหรับการประมวลผลภาษาไทย

NLP หรือ Natural Language Processing แปลเป็นไทยคือการประมวลผลภาษาธรรมชาติ เป็นวิธีการที่ทำให้คอมพิวเตอร์สามารถเข้าใจคำ ประโยค ข้อความภาษาที่มนุษย์ใช้สื่อสาร



ภาพที่ 2.8 Pythainlp Logo

```
[25]: from pythainlp import word_tokenize

text = "ก็จะรู้ความชั่วร้ายที่ทำไว้ และคงจะไม่ยอมให้ท่านบนหลังคน"

print("default (newmm):")
print(word_tokenize(text))
print("\nnewmm and keep_whitespace=False:")
print(word_tokenize(text, keep_whitespace=False))

default (newmm):
['ก็', 'จะ', 'รู้', 'ความ', 'ชั่ว', 'ร้าย', 'ที่', 'ทำ', 'ไว้', 'และ', 'คง', 'จะ', 'ไม่', 'ยอม', 'ให้', 'ท่าน', 'บน', 'หลัง', 'คน']

newmm and keep_whitespace=False:
['ก็', 'จะ', 'รู้', 'ความ', 'ชั่ว', 'ร้าย', 'ที่', 'ทำ', 'ไว้', 'และ', 'คง', 'จะ', 'ไม่', 'ยอม', 'ให้', 'ท่าน', 'บน', 'หลัง', 'คน']
```

ภาพที่ 2.9 ตัวอย่างการใช้งาน Pythainlp

2.2.8 Wordcloud



ภาพที่ 2.10 Wordcloud

แต่ในที่นี้จะใช้ Wordcloud for Python ที่เป็น Library ภาษา Python ซึ่งมีความสามารถในการแสดง Wordcloud ได้ทันที



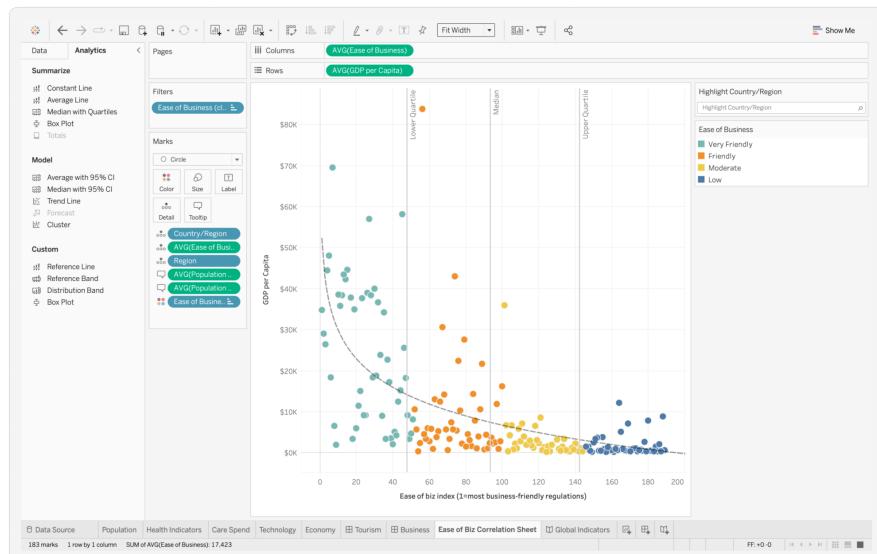
ภาพที่ 2.11 ตัวอย่างการใช้งาน Wordcloud for Python

2.2.9 Tableau

Tableau เป็นโปรแกรมสำหรับการวิเคราะห์ข้อมูล และหาคำตอบในเชิงธุรกิจได้อย่างมีประสิทธิภาพ โดยการใช้ Tableau จะเป็นการช่วยให้ผู้ใช้งานมีความเข้าใจในข้อมูลมากขึ้น รวมถึงผู้ใช้งาน สามารถนำข้อมูลที่มีอยู่มาวิเคราะห์และแสดงผลอย่างมีประสิทธิภาพ

Tableau สามารถนำข้อมูลที่มีอยู่ มาสร้างเป็น Visualization, Dashboard และสามารถ Share Content ได้อย่างรวดเร็วด้วย Tableau Server

โปรแกรม Tableau เป็นโปรแกรมที่ใช้งานง่าย ถูกสร้างขึ้นมาโดยผู้เชี่ยวชาญด้าน Computer Graphic, Database และ Human-Computer Interaction จึงทำให้โปรแกรมเป็นมิตร กับทั้งทาง Business User และทาง IT

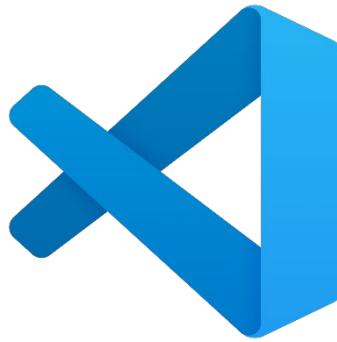


ภาพที่ 2.12 ตัวอย่างการใช้งานโปรแกรม Tableau

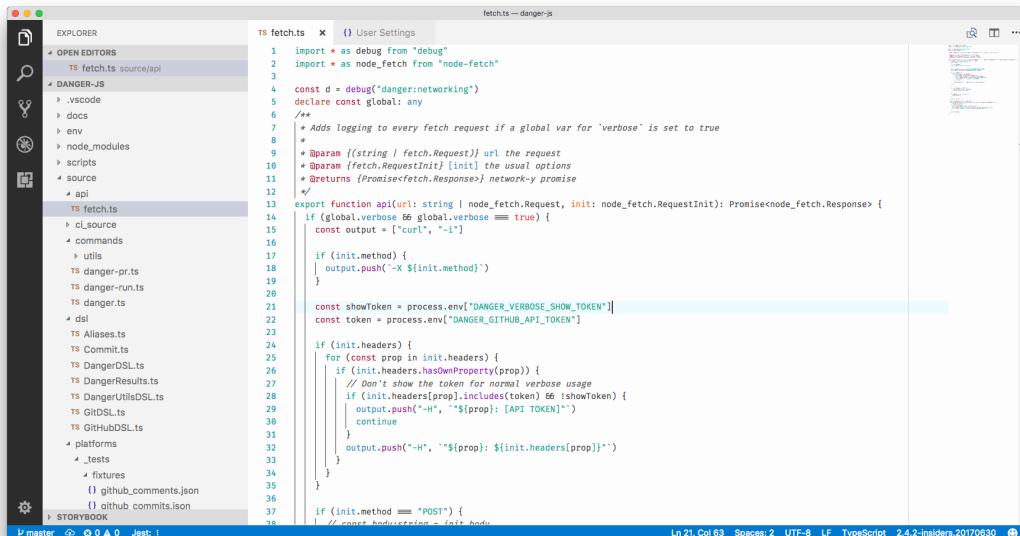
2.2.10 Visual Studio Code

Visual Studio Code หรือ VSCode เป็นโปรแกรม Code Editor ที่ใช้ในการแก้ไขและปรับแต่งโค้ด จากค่ายไมโครซอฟท์ มีการพัฒนาอุปกรณ์ในรูปแบบของ Open Source จึงสามารถนำมาใช้งานได้แบบฟรี ๆ ที่ต้องการความเป็นมืออาชีพ

ซึ่ง Visual Studio Code นั้น เหมาะสำหรับนักพัฒนาโปรแกรมที่ต้องการใช้งานข้ามแพลตฟอร์ม รองรับการใช้งานทั้งบน Windows, macOS และ Linux สนับสนุนทั้งภาษา JavaScript, TypeScript และ Node.js สามารถเชื่อมต่อกับ Git ได้ นำมาใช้งานได้ง่ายไม่ซับซ้อน มีเครื่องมือส่วนขยายต่าง ๆ ให้เลือกใช้อย่างมากมาก ไม่ว่าจะเป็น 1. การเปิดใช้งานภาษาอื่น ๆ ทั้งภาษา C++, C#, Java, Python, PHP หรือ Go 2. Themes 3. Debugger 4. Commands เป็นต้น



ภาพที่ 2.13 Visual Studio Code Logo



ภาพที่ 2.14 ตัวอย่างการใช้งานโปรแกรม Visual Studio Code

2.2.11 HTML

HTML ย่อมาจาก Hyper Text Markup Language คือภาษาคอมพิวเตอร์ที่ใช้ในการแสดงผลของเอกสารบน Website หรือที่เรียกว่า Webpage ถูกพัฒนาและกำหนดมาตรฐานโดยองค์กร World Wide Web Consortium (W3C) และจากการพัฒนาทางด้าน Software ของ Microsoft ทำให้ภาษา HTML เป็นอีกภาษาหนึ่งที่ใช้เขียนโปรแกรมได้ หรือที่เรียกว่า HTML Application HTML เป็นภาษาประเภท Markup สำหรับการสร้าง Webpage โดยใช้ภาษา HTML สามารถทำโดยใช้โปรแกรม Text Editor ต่างๆ เช่น Notepad, Editplus หรือจะอาศัยโปรแกรมที่เป็นเครื่องมือช่วยสร้าง Webpage เช่น Microsoft FrontPage, Dream Weaver ซึ่งอำนวยความสะดวกในการสร้างหน้า HTML ส่วนการเรียกใช้งานหรือทดสอบการทำงานของเอกสาร HTML จะใช้โปรแกรม Web Browser เช่น IE Microsoft Internet Explorer (IE), Mozilla Firefox, Safari, Opera, และ Netscape Navigator เป็นต้น

```

1  <!DOCTYPE HTML>
2  <html lang="en-US">
3  <head>
4      <meta charset="UTF-8">
5      <title></title>
6  </head>
7  <body>
8
9  </body>
10 </html>

```

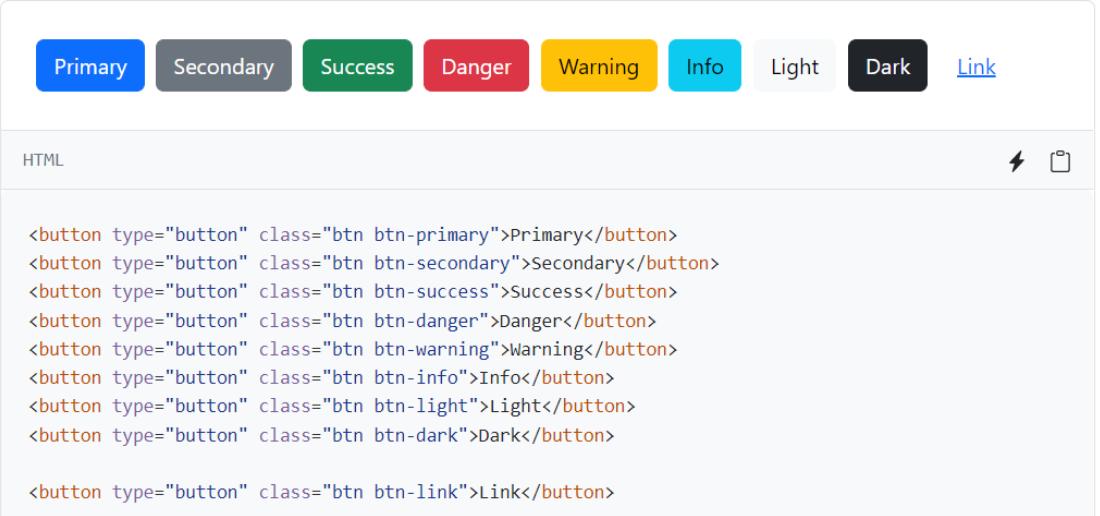
ภาพที่ 2.15 ตัวอย่างภาษา HTML

2.2.12 Bootstrap

Bootstrap คือชุดคำสั่งที่ประกอบด้วยภาษา CSS, HTML และ Javascript เป็นชุดคำสั่งที่ถูกพัฒนาขึ้นมาเพื่อกำหนดรอบหรือรูปแบบการพัฒนา Website ในส่วนของการปฏิสัมพันธ์กับผู้ใช้งาน Website (User Interface) จึงสามารถเรียก Bootstrap ว่าเป็น Front-end Framework คือใช้สำหรับ พัฒนา Website สำหรับการแสดงผล ซึ่งแตกต่างจากภาษาประเภท Server Side Script อย่าง PHP, Python หรือภาษาอื่น ๆ

Bootstrap ถูกพัฒนาขึ้นโดย Mark Otto และ Jacob Thornton ทีมพัฒนาของ Twitter Inc. ก่อนหน้านี้ใช้ชื่อว่า Twitter Blueprint และเปิดให้นักพัฒนาสามารถนำไปใช้งานพัฒนา Website ได้แบบฟรี (Open Source) ในชื่อว่า Bootstrap Framework

จุดเด่นของของ Bootstrap Framework มี UI เริ่มต้นแบบที่สวยงามและใช้งานง่าย, มีการปรับปรุง และพัฒนาอย่างต่อเนื่อง, ปัจจุบันเป็นเวอร์ชัน 3.3.0 เป็นที่นิยมของนักพัฒนาทั่วโลก ทำให้สามารถเรียนรู้และแก้ปัญหาได้ง่าย, Code หรือชุดคำสั่งต่าง ๆ ค่อนข้างสะอาดมี Folder ต้นแบบแค่ 3 ส่วน คือ JS, CSS, Fonts, ประหยัดเวลาในการพัฒนา Website และนำไปพัฒนาต่อได้ง่าย และเป็น Responsive Framework พัฒนา Website ที่รองรับการแสดงผลได้หลากหลาย Device



```

Primary Secondary Success Danger Warning Info Light Dark Link
HTML
<button type="button" class="btn btn-primary">Primary</button>
<button type="button" class="btn btn-secondary">Secondary</button>
<button type="button" class="btn btn-success">Success</button>
<button type="button" class="btn btn-danger">Danger</button>
<button type="button" class="btn btn-warning">Warning</button>
<button type="button" class="btn btn-info">Info</button>
<button type="button" class="btn btn-light">Light</button>
<button type="button" class="btn btn-dark">Dark</button>

<button type="button" class="btn btn-link">Link</button>

```

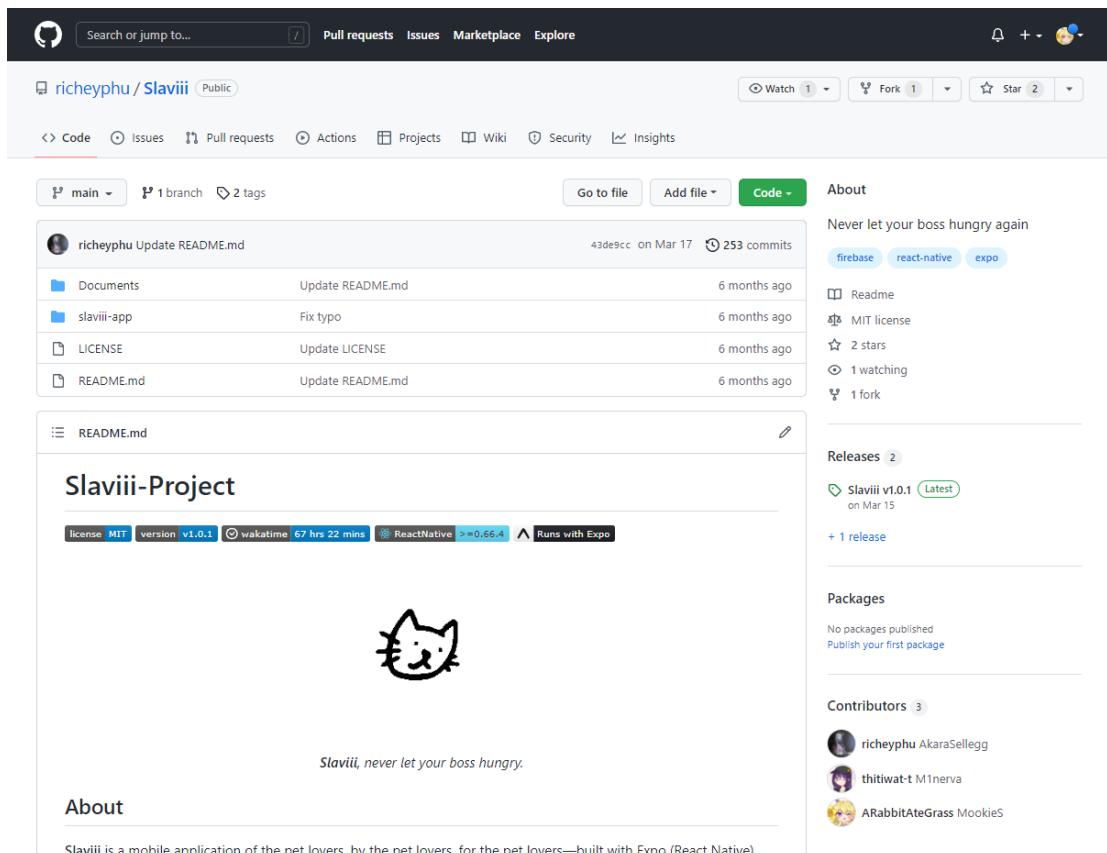
ภาพที่ 2.16 ตัวอย่างการใช้งาน Bootstrap

2.2.13 Github

เริ่มจาก “Git” คือเครื่องมือ Version control ที่มีไว้สำหรับจัดเก็บความเปลี่ยนแปลงของไฟล์ในโปรเจคส่วนตัวหรือทีม ไม่ว่าจะเพิ่ม ลบ เปลี่ยน ระบบก็จะบันทึกไว้ ทำให้สามารถรู้ประวัติการเปลี่ยนแปลงและความเคลื่อนไหวต่าง ๆ ของโปรเจคที่ถืออยู่ได้

ส่วน “Github” คือ Website ที่ให้บริการ Git (Version Control Repository) โดย Github จะให้บริการบนออนไลน์แพลตฟอร์ม ทำให้คุณสามารถเข้าถึงข้อมูลผ่านหน้า Website ได้ทุกที่ทุกเวลา ในส่วนของการใช้บริการก็มีให้ใช้งานแบบเสียเงินและไม่เสียเงิน

GitHub ก่อตั้งขึ้นมาในวันที่ 24 กุมภาพันธ์ 2009 โดยใช้โลโก้เป็นรูปแมวมีหนวดหมักที่มีชื่อว่า “Octocat” หลังจากนั้นก็ Github ก็เติบโตอย่างรวดเร็ว จนในปัจจุบันบริษัท GitHub มีมูลค่ามากกว่า \$360 ล้านเหรียญสหรัฐและมีผู้ใช้งานมากกว่า 37 ล้านคนทั่วโลก และมี Repository รวมกันกว่า 100 ล้านไฟล์บนระบบ



ภาพที่ 2.17 ตัวอย่างการใช้งาน Github

2.2.14 Netlify

Netlify เป็นผู้ให้บริการโฮสต์ Static Website บน Cloud Server รองรับทั้ง Static HTML Website ธรรมดា หรือจะให้ Build จาก Static Site Generator อย่าง Hugo หรือ Jekyll และ Deploy บน Cloud ของ Netlify ก็ได้ (หรือจะใช้ Tool ก็ได้เช่นกัน รองรับ Node, Python, Ruby)

Netlify ดีอย่างไร นอกจากการที่ไม่ต้องเช่า Server เองแล้ว Netlify ก็ยังมีฟีเจอร์ให้โฮสต์ Website บน Cloud Server ที่ทำหน้าที่เป็น CDN ได้ด้วย, รองรับ Custom Domain พร้อมทั้งใช้ HTTPS ได้ฟรีจาก Let's Encrypt, มีระบบ Continuous Deployment มาให้ใช้ฟรีๆ (Deploy อัตโนมัติ, เก็บ Snapshot ของการ Deploy แต่ละครั้งไว้ได้, มี Log ให้ดูย้อนหลัง), มีระบบ Optimize Static Assets (JS/CSS/รูปภาพต่างๆ) ให้อัตโนมัติทุกครั้งที่ Deploy, ใช้งานฟีเจอร์ที่เคยต้องมี Backend อย่าง Sign up/Login หรือ Form ได้ และใช้งาน AWS Lambda Functions ได้ นอกจากนี้ยังมีฟีเจอร์อื่นๆ ที่กำลังพัฒนาอีกมากมาย ทั้งหมดนี้ใช้งานได้โดยที่ไม่ต้องมี Server เป็นของตัวเอง ยิ่งไปกว่านั้น ทั้งหมดนี้ สามารถใช้งานได้โดยไม่ต้องเสียค่าใช้จ่าย

The screenshot shows the Netlify Team Overview page for the 'ARabbitAteGrass's team'. At the top, it displays the team name and a 'Starter' plan badge. Below this, there are four key usage metrics: Bandwidth used (60 KB/100 GB), Build minutes used (0/300), Concurrent builds (0/1), and Team members (1). To the right of these metrics is a blue square icon with a white geometric pattern. The main area is divided into two sections: 'Sites' and 'Builds'. The 'Sites' section shows one site named 'sanmook-proj-chadchart' which 'Deploys from GitHub'. The 'Builds' section shows three completed builds for the same project, each with a timestamp: 'Production: main@5a@fea7' (Aug 30 at 10:47 AM), 'Production: main@3d287ef' (Aug 30 at 10:47 AM), and 'Production: main@339cf38' (Aug 30 at 10:47 AM).

ภาพที่ 2.18 ตัวอย่างการใช้งาน Netlify

บทที่ 3

แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน

การปฏิบัติสหกิจศึกษา ณ บริษัท แออดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน) ในตำแหน่ง Business Data Analyst ประกอบไปด้วยงานที่ได้รับมอบหมายจากพี่เลี้ยง และการทำโครงการ สหกิจศึกษา มีรายละเอียดต่าง ๆ ดังนี้

3.1 แผนงานการฝึกงาน

ระยะเวลาการปฏิบัติสหกิจศึกษาร่วมทั้งสิ้น 4 เดือน โดยส่วนใหญ่ 2 เดือนแรกจะเป็นการทำางานที่ได้รับมอบหมายจากพี่เลี้ยง เป็นการช่วยแบ่งเบาภาระโครงการหลักที่พี่เลี้ยงรับผิดชอบอยู่ ส่วน 2 เดือนหลัง จะเป็นการทำโครงการสหกิจศึกษาควบคู่ไปกับการรับงานที่ได้รับมอบหมายจากพี่เลี้ยง

ตารางที่ 3.1 แผนการปฏิบัติสหกิจศึกษา

หัวข้องาน	เดือนที่ 1	เดือนที่ 2	เดือนที่ 3	เดือนที่ 4
1. ศึกษาเรียนรู้ Process การทำงานของบริษัท	■ ■ ■			
2. ศึกษาเรียนรู้ tools ที่ต้องใช้ในการทำงาน		■ ■ ■ ■		
2.1. เรียนรู้การทำ Visualization		■ ■ ■ ■		
2.2. เรียนรู้การ Scraping Website		■ ■ ■ ■		
- Scraping website for insurance project		■ ■ ■ ■		
- Web scraping API for insurance project		■ ■ ■ ■		
2.3. เรียนรู้การทำ Data Cleaning			■ ■ ■ ■	
- Data Cleaning for insurance project			■ ■ ■ ■	
3. หาข้อมูลและคิดหัวข้อโครงการ			■ ■ ■ ■	
4. ลงมือทำโครงการ			■ ■ ■ ■ ■ ■ ■ ■	
4.1. เตรียมความพร้อมของข้อมูล			■ ■ ■ ■ ■ ■ ■ ■	
4.2. ลงมือวิเคราะห์ข้อมูล			■ ■ ■ ■ ■ ■ ■ ■	
4.3. สร้าง Visualization				■ ■ ■ ■ ■ ■ ■ ■
5. สรุปผล เรียบเรียงทำเล่มรายงาน				■ ■ ■ ■ ■ ■ ■ ■
5.1. เตรียมนำเสนอ				■ ■ ■ ■ ■ ■ ■ ■

3.2 รายละเอียดที่นักศึกษาปฏิบัติในการฝึกงาน

3.2.1 Insurance Project

Insurance Project คือโครงการเกี่ยวกับประกันภัย ที่เป็นงานหลักของพี่เลี้ยง คือการรับความต้องการของลูกค้า และนำข้อมูลที่มีอยู่มาวิเคราะห์ว่ามีทางใดที่จะได้มาซึ่งข้อมูลที่ลูกค้าต้องการ หรือวิเคราะห์เพื่อหาข้อมูลเชิงลึกที่ซ่อนอยู่ในความต้องการของลูกค้า เช่น ลูกค้าต้องการคนที่มีรถ เพื่อที่จะส่งโฆษณาประกันรถยนต์ไปได้ถูกกลุ่มลูกค้า ซึ่งการวิเคราะห์ข้อมูลและหาข้อมูลมาให้ตรงกับความต้องการของลูกค้า เป็นหนึ่งในหน้าที่หลักของ Business Data Analyst

โดยที่งานที่ได้รับมอบหมายจากพี่เลี้ยง ส่วนมากจะเป็นการช่วยแบ่งเบาภาระของพี่เลี้ยง ยกตัวอย่างงานที่ได้รับมอบหมายจากพี่เลี้ยง เช่น การ Scraping Data เเบร์โทรศัพท์อุ่รรถของบริษัทประกันภัยที่ถูกเผยแพร่ไว้บน Website ของบริษัทประกันภัยนั้น

```
[10] res = requests.get(url)
    res.encoding = "utf-8"
    print(res)

... <Response [200]>

[11]
soup = BeautifulSoup(res.text, 'html.parser')
t = soup.get_text()

data = soup.find_all('div',{'class':'table_serve'})[0]

> v
df = pd.DataFrame(columns = ['name','tel','car_brand','car_type'])
for i in data.find_all('tr')[1:]:
    name = i.find_all('a')[0].get_text().strip()

    tel = i.find_all('td',{'align':'left'})[0].get_text()
    junk = i.find_all('td',{'align':'left'})[0].span.get_text()
    tel = tel.replace(junk,'').strip()

    car_brand = i.find_all('td',{'align':'center'})[0].get_text()

    car_type = i.find_all('td',{'align':'center'})[1].get_text()
    junk = i.find_all('td',{'align':'center'})[1].span.get_text()
    car_type = car_type.replace(junk,'').strip()
    df = df.append({'name':name , 'tel':tel , 'car_brand':car_brand , 'car_type':car_type},ignore_index=True)

[12]
```

ภาพที่ 3.1 ตัวอย่าง Code Scraping Data เเบร์โทรศัพท์อุ่รรถของบริษัทประกันภัย

หลังจาก Scraping Data เรียบร้อยแล้ว ก็ต้องมีการทำ Data Cleaning หรือการทำความสะอาดข้อมูล เพราะข้อมูลที่ได้มาอาจมีข้อมูลที่ไม่ต้องการ ก็ต้องทำการทำความสะอาดข้อมูลให้เรียบร้อยก่อนนำไปใช้งานในขั้นตอนถัดไป

```

[23] df['first_number'] = df['tel_split'].str.replace('-', '')

[24] df['first_number'] = df['first_number'].str.replace(' ', '')

[25] df['first_number'] = [x[:2] for x in df['first_number']]

[26]
df['chk'] = np.select([
    (df['first_number'].isin(['02','03','04','05','07'])) & (df['tel_count'] == 9),
    ((df['first_number'].isin(['06','08','09'])) & (df['tel_count'] == 10)),
    True
], [
    'Y',
    'N'
])

[27] df_y = df[df['chk'] == 'Y']
df_n = df[df['chk'] == 'N']

```

ภาพที่ 3.2 ตัวอย่าง Code Clean Data

รวมไปถึงงานที่ได้รับมอบหมายจากพี่เลี้ยงให้ Query (คัดเลือกเฉพาะข้อมูลที่ต้องการ) เพื่อตรวจสอบข้อมูลที่มีอยู่บนฐานข้อมูลขนาดใหญ่ (Big Data) บน Databrick โดยพี่เลี้ยงแนะนำให้ใช้ Pyspark ในการ Query เนื่องจากเป็นภาษาที่สร้างขึ้นมาเพื่อใช้ในการจัดการกับข้อมูลขนาดใหญ่ (Big Data) โดยเฉพาะ สามารถแสดงผลข้อมูลที่ต้องการได้อย่างรวดเร็วกว่า SQL และ Pandas ดังนั้นการใช้ Pyspark จัดการกับข้อมูลขนาดใหญ่จึงเหมาะสมกว่า

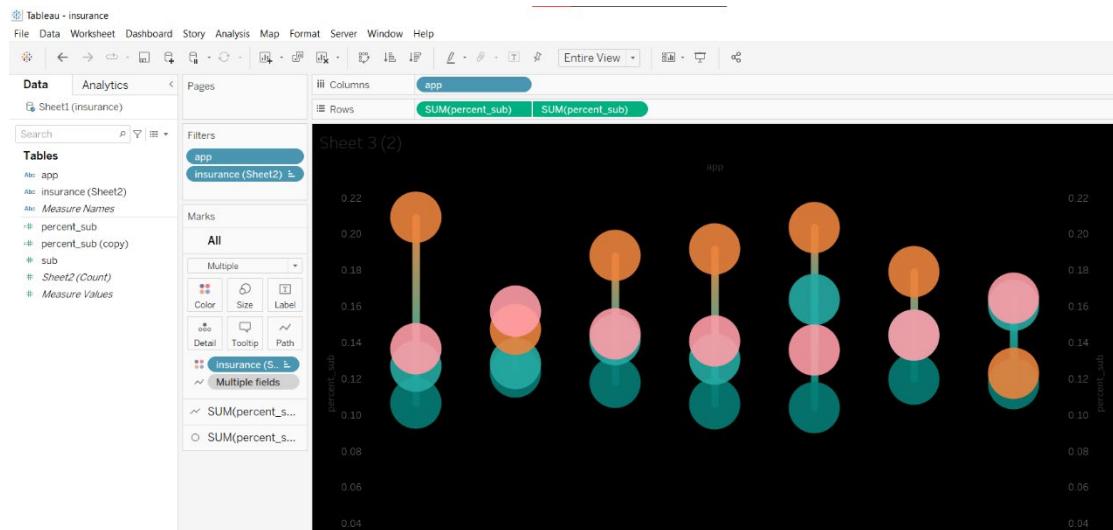
	action	date
1	Close	2016-07-28
2	Close	2016-07-28
3	Open	2016-07-28
4	Close	2016-07-28
5	Open	2016-07-28
6	Open	2016-07-28
7	Close	2016-07-28

Truncated results, showing first 1000 rows.

Command took 1.85 seconds -- by a user at 1/28/2019, 4:37:51 PM on unknown cluster

ภาพที่ 3.3 ตัวอย่างการใช้ Pyspark Query ข้อมูลบน Databrick

นอกจากนั้นจะเป็นการสร้าง Visualization จากข้อมูลที่พิเลี้ยงเตรียมมาให้ ส่วนมากเป็นการใช้โปรแกรม Tableau เป็นหลัก ในการทำ Visualization ซึ่งจากการสังเกตพบว่า Tableau สามารถสร้างกราฟได้หลายรูปแบบ สามารถใส่สูตรในการสร้างกราฟได้หลากหลาย อีกทั้งยังสามารถตอบแต่งเพิ่มเติมได้ตามใจชอบ ขึ้นอยู่กับความต้องการและความสามารถของผู้ใช้งาน



ภาพที่ 3.4 Visualization ที่สร้างด้วยโปรแกรม Tableau

3.2.2 โปรเจคอื่น ๆ

นอกเหนือจาก Insurance Project ที่ได้ช่วยเป็นโปรเจคหลักแล้ว ยังมีโปรเจคอื่น ๆ ที่ได้รับมอบหมายให้ช่วยเหลือ เช่น Banking Project และ Retail Project

Banking Project ได้รับมอบหมายให้ช่วยงานในด้านการจัด Category หรือการจัดหมวดหมู่ของข้อมูลซึ่งเน้นหลักไปที่การจัดข้อมูลในส่วนของ Personal Finance หรือสินเชื่อส่วนบุคคล เพื่อให้พร้อมต่อการเรียกใช้งานในส่วนถัดไป

Retail Project เป็นโปรเจคด่วนที่ลูกค้าต้องการข้อมูลและการวิเคราะห์ข้อมูลจากมุมมองของบุคคลภายนอก เพื่อที่จะได้รับมุมมองใหม่ ๆ นอกเหนือจากการวิเคราะห์กันเองภายในองค์กรของลูกค้า ที่เลี้ยงจึงมอบหมายให้ช่วยทำกราฟในรูปแบบที่พิเศษต้องการคือ Sankey Diagram หรือกราฟแสดงการไหลของข้อมูล โดยพิเศษมีการเตรียมข้อมูลไว้ให้แล้ว จึงได้สร้างกราฟโดยประยุกต์ใช้ Code จาก Google Developer [9]

```

<html>
  <head>
    <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script>
    <script type="text/javascript">
      google.charts.load('current', {'packages':['sankey']});
      google.charts.setOnLoadCallback(drawChart);

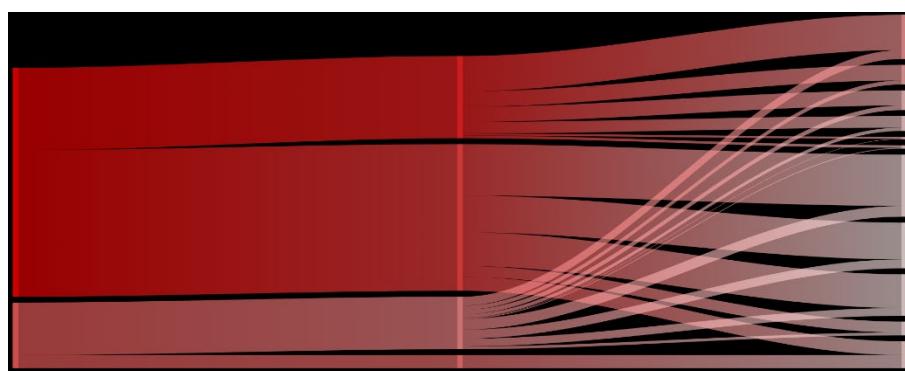
      function drawChart() {
        var data = new google.visualization.DataTable();
        data.addColumn('string', 'From');
        data.addColumn('string', 'To');
        data.addColumn('number', 'Weight');
        data.addRows([
          [ 'A', 'X', 5 ],
          [ 'A', 'Y', 7 ],
          [ 'A', 'Z', 6 ],
          [ 'B', 'X', 2 ],
          [ 'B', 'Y', 9 ],
          [ 'B', 'Z', 4 ]
        ]);

        // Sets chart options.
        var options = {
          width: 600,
        };

        // Instantiates and draws our chart, passing in some options.
        var chart = new google.visualization.Sankey(document.getElementById('sankey_basic'));
        chart.draw(data, options);
      }
    </script>
  </head>
  <body>
    <div id="sankey_basic" style="width: 900px; height: 300px;"></div>
  </body>
</html>

```

ภาพที่ 3.5 ตัวอย่าง Code สร้าง Sankey Diagram



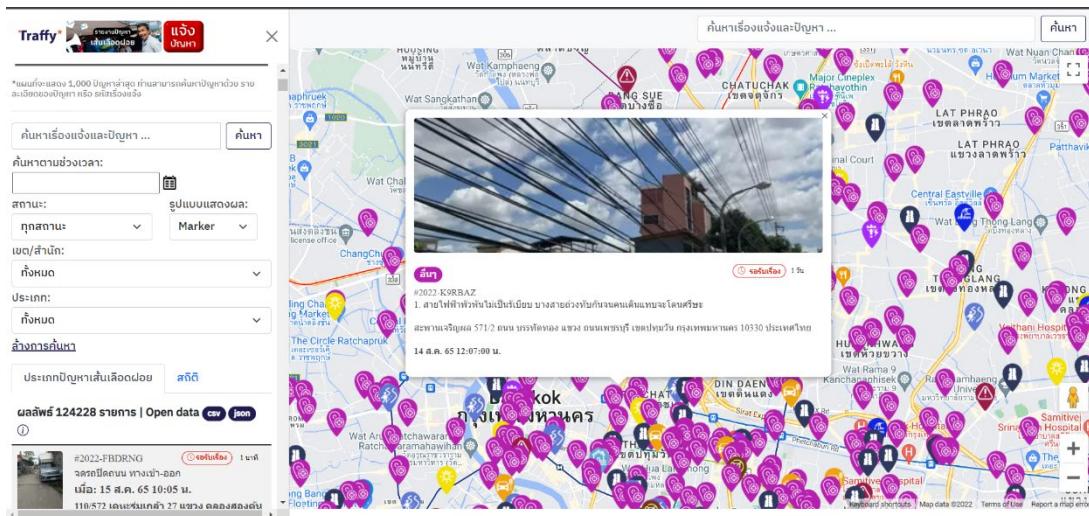
ภาพที่ 3.6 Sankey Diagram ที่สร้าง

3.3 ขั้นตอนการดำเนินโครงการสหกิจศึกษา

3.3.1 กำหนดหัวข้อโครงการและหาข้อมูลที่จะนำวิเคราะห์

เนื่องจากโครงการนี้มีความจำเป็นต้องเผยแพร่ ข้อมูลที่จะนำวิเคราะห์จึงควรจะเป็นข้อมูลที่เปิดเผยได้ มีปริมาณข้อมูลมากพอสมควร และมีความหลากหลายของข้อมูลเพื่อให้สามารถนำมาร่วมกันวิเคราะห์ได้ ดังนั้นจึงได้ตัดสินใจเลือก ข้อมูลรายงานปัญหาเส้นเลือดฟ้อยภายในเขตพื้นที่จังหวัดกรุงเทพฯ มาเป็นข้อมูลที่จะนำวิเคราะห์

เนื่องจากข้อมูลเป็นข้อมูลที่เป็นสาธารณะ สามารถขอข้อมูลในระบบ เพื่อนำไปจัดทำสถิติ หรือวิเคราะห์ข้อมูล ผ่าน API ได้ นอกจากนั้นข้อมูลยังมีปริมาณมากพอสมควรที่จะสามารถวิเคราะห์ได้ และมีความหลากหลายของข้อมูล คือปัญหาต่าง ๆ ที่ประชาชนรายงานเข้ามา มีความน่าสนใจในการวิเคราะห์ รวมไปถึงข้อมูลดังกล่าว yang สามารถวิเคราะห์ในเชิงลึกและหลากหลายมุมมองได้มากขึ้น และยังเป็นประโยชน์ต่อส่วนรวมได้ จึงตัดสินใจกำหนดหัวข้อโครงการเป็นเนื้อหาดังกล่าว



ภาพที่ 3.7 ข้อมูลรายงานปัญหาเส้นเลือดฟ้อยภายในเขตพื้นที่จังหวัดกรุงเทพฯ

3.3.2 รายละเอียดข้อมูลที่เลือก



ภาพที่ 3.8 ภาพจาก Website Traffy โครงการ Traffy x TeamChadChart

Traffy x TeamChadChart เป็น แพทบอทให้บริการรวมข้อมูลปัญหาเส้นเลือดฟอยย์ ในกรุงเทพมหานคร การแก้ปัญหาเส้นเลือดฟอยย์ เป็นหนึ่งในนโยบายของ รองศาสตราจารย์ ชัชชาติ สิทธิพันธุ์ ผู้ว่าราชการกรุงเทพมหานคร โดยการเริ่มแก้ปัญหาจากจุดเล็กๆ เพื่อให้คุณภาพชีวิตของทุกคนดีขึ้น

ระบบต่าง ๆ ในกรุงเทพฯ เปรียบได้กับร่างกายคน มีระบบเส้นเลือดใหญ่ที่เป็นหลัก อยู่ส่วนกลาง และมีเส้นเลือดฟอยย์ที่กระจายไปยังส่วนต่าง ๆ ของร่างกาย ตัวอย่างเช่น เปรียบเส้นเลือดใหญ่เส้นหนึ่งเป็นการระบายน้ำผ่านอุโมงค์กษัตรีขนาดใหญ่ มูลค่าเป็นหมื่นล้านบาท ในขณะที่เส้นเลือดฟอยย์เปรียบเป็นท่อระบายน้ำตามชุมชนต่าง ๆ หลายที่อุดตัน ระบายน้ำไม่ได้ ฉะนั้นแล้วระบบเส้นเลือดฟอยย์ที่มีปัญหา สุดท้ายแล้วจะส่งผลกระทบต่อการทำงานและประสิทธิภาพของเส้นเลือดใหญ่ จึงสมควรได้รับการแก้ไขก่อนที่จะสายเกินแก้



ภาพที่ 3.9 QR code สำหรับการแจ้งปัญหา

3.3.3 ปัญหาที่พบจากข้อมูลที่เลือก

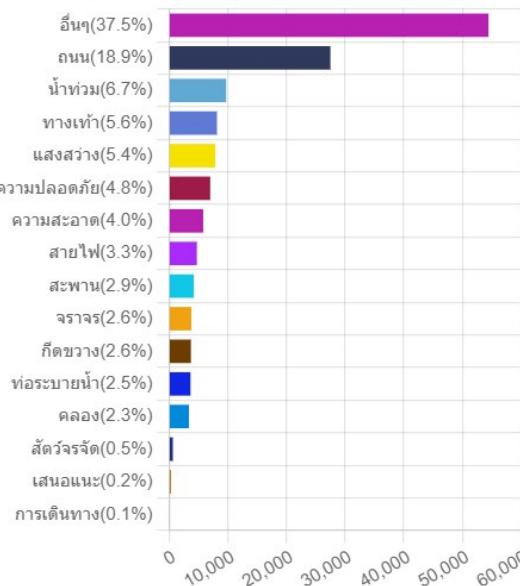
จากข้อมูลที่ได้เลือกมาคือ ข้อมูลรายงานปัญหาเส้นเลือดฟ้อยภายในเขตพื้นที่จังหวัดกรุงเทพมหานคร ซึ่งมีหน้าข้อมูลหลักอยู่ที่ Website <https://share.traffy.in.th/teamchadchart> ที่จะแสดงข้อมูล สติ๊ต และจุดที่มีการแจ้งปัญหาเข้ามายาบบันแพนที่เป็นบางส่วน ดังภาพที่ 3.7

โดยข้อมูลจะถูกจัดแบ่งตามประเภทของปัญหาที่แจ้งเข้ามา เดิมจะมีอยู่ 7 ประเภท ประกอบไปด้วย 1. ประเภทขยะ 2. ประเภทความปลอดภัย 3. ประเภทจราจร 4. ประเภททางเท้า 5. ประเภทน้ำท่วม 6. ประเภทเสนอแนะ และ 7. ประเภทอื่น ๆ

จากการสังเกตสถิติที่แสดงบน Website พบว่า ข้อมูลนั้นมีความผิดปกติ นั่นคือ ประเภทของปัญหาที่แจ้งเข้ามา ประเภท “อื่น ๆ” กลับมีจำนวนมากที่สุด

จะสังเกตเห็นได้ว่า ทาง Website ได้พยายามแบ่งประเภทของปัญหาออกมาใหม่ ดังภาพที่ 3.10 แต่ปัญหาประเภท “อื่น ๆ” กลับยังคงมากเป็นอันดับ 1 ดังนั้น ปัญหาจากข้อมูลที่เลือกนี้คือ ประเภทของปัญหาที่แจ้งเข้ามา ประเภท “อื่น ๆ” มีจำนวนมากเกินไป

ประเภทปัญหาเส้นเลือดฟอย



ภาพที่ 3.10 ภาพปัญหาประเภท “อื่น ๆ” ที่มีจำนวนมาก

3.3.4 ดึงข้อมูลจาก Website (Web Scraping : API)

ความหมายของ API (Application Programming Interface) แปลเป็นภาษาไทยตรงตัวได้ว่า “ส่วนต่อประสานโปรแกรมประยุกต์” หรือก็คือ “ตัวกลาง” ที่ให้ Website, Software, Application สามารถสื่อสารและทำงานร่วมกันได้ โดยเชื่อมถึงข้อมูลกับฐานข้อมูล (Database) ของอีกฝ่าย ในรูปแบบของ Code ที่เป็นคำสั่งเรียกใช้ API

การดึงข้อมูลจาก Website โดยผ่าน API จึงสามารถทำได้ง่ายและสะดวก เพราะ API เป็นตัวกลางเชื่อมต่อให้แล้ว เวลาจะใช้ข้อมูลก็จะหยิบข้อมูลจากที่ตัว API ส่งมา โดยทาง Website traffy.in.th เปิดให้ผู้นำข้อมูลไปพัฒนาต่อยอดสามารถขอข้อมูลในระบบ เพื่อนำไปจัดทำสถิติหรือวิเคราะห์ข้อมูล ผ่าน API โดยข้อมูลจะอัปเดตทุกๆ 10 นาที และยังอำนวยความสะดวกสะดวกแก่ผู้นำข้อมูลไปพัฒนาต่อยอดด้วยการระบุ Resource URL สำหรับการเข้าถึงข้อมูลที่ต้องการผ่าน URL ดังกล่าว ไว้แล้ว ยิ่งไปกว่านั้นยังระบุ Parameters ที่เป็นการตั้งค่าสำหรับดึงข้อมูลผ่าน API

Parameters				
Name	Required	Description	Default Value	Example
limit	optional	Limit จำนวน ผลลัพธ์การ ค้นหา (ไม่เกินครึ่ง ละ 1000)	1000	100
offset	optional	ข้อมูลเริ่ม ต้นที่ ต้องการ	0	100
start	optional	วันที่เริ่บต้น (ถ้าระบุ start จำเป็นต้อง ระบุ end มา ด้วย)		2021-10-01
end	optional	วันที่เริ่มสิ้น สุด (ถ้าระบุ end จำเป็นต้อง ระบุ start มาด้วย)		2021-10-05

ภาพที่ 3.11 Parameters ที่ traffy.in.th ระบุ

แต่ทว่า Parameters ที่ชื่อ Limit มีการระบุรายละเอียดไว้ว่า มีการจำกัดจำนวนผลลัพธ์การค้นหา ไม่เกินครั้งละ 1,000 ซึ่งหมายความว่า ข้อมูลที่แสดงผลต่อ 1 การค้นหา สามารถแสดงได้สูงสุดแค่ 1,000 ผลลัพธ์เท่านั้น ซึ่งถ้าหากต้องการข้อมูลมากกว่านั้น ต้องทำการเขียน Code Python สำหรับวน Loop เพื่อดึงข้อมูลครั้งละ 1,000 แต่วนเก็บข้อมูลหลายครั้ง จนครบจำนวนที่ต้องการดังภาพที่ 3.12

```
Cmd 6
1 n = 1000
2 while n <= 90000:
3     url = f"https://publicapi.traffy.in.th/share/teamchadchart/search?&offset={n}"
4
5     res = requests.get(url)
6     print(res)
7     print(n)
8     if res.status_code != 201:
9         continue
10    else:
11        soup = BeautifulSoup(res.text, 'html.parser')
12        text = soup.get_text()
13        df = pd.json_normalize(json.loads(text)['results'])
14        df_sum = df_sum.append(df, ignore_index=True)
15        n += 1000
```

ภาพที่ 3.12 ตัวอย่าง Code สำหรับวน loop เพื่อดึงข้อมูลผ่าน API

โดยข้อมูลที่ได้จะประกอบไปด้วย 15 คอลัมน์ ดังตารางที่ 3.2 ด้านล่าง

ตารางที่ 3.2 โครงสร้างข้อมูลรายงานปัญหาเส้นเลือดฟอย

Column	Datatype	Description
Id	String	ลำดับปัญหาที่แจ้ง
Type	String	ประเภทปัญหาที่แจ้ง
Org	String	หน่วยงานที่รับผิดชอบ
Comment	String	รายละเอียดปัญหาที่แจ้ง
Ticket_id	String	รหัสปัญหาที่แจ้ง
Coords	Array	ตำแหน่งของปัญหาที่แจ้ง
Photo	String	รูปภาพปัญหาที่แจ้ง
After_photo	String	รูปภาพหลังแก้ไขปัญหาที่แจ้งแล้ว
Address	String	ที่อยู่ของปัญหาที่แจ้ง
District	String	แขวงของปัญหาที่แจ้ง
Subdistrict	String	เขตของปัญหาที่แจ้ง
Province	String	จังหวัดของปัญหาที่แจ้ง
Timestamp	String	เวลาที่แจ้งปัญหา
Problem_Type_Abdul	String	ประเภทของปัญหาที่ทาง Website พยายามแบ่งใหม่
State	String	สถานะปัญหาที่แจ้ง

3.3.5 ทำความสะอาดข้อมูล (Data Cleaning)

ข้อมูลที่ดึงมาจาก Website ผ่าน API มา ยังไม่สามารถนำมารวบรวมได้ เพราะอาจจะปนด้วยสิ่งไม่พึงประสงค์ต่างๆ

เมื่อได้ข้อมูลมาครั้งแรก ต้องทำความสะอาดข้อมูล นำส่วนที่ไม่ต้องการจะใช้ออกก่อน เช่น ข้อมูลที่ว่างเปล่า (Null) หรือข้อมูลผิดประเภทที่ไม่ต้องการ ซึ่งหากไม่นำออก จะทำให้เกิดปัญหาในขั้นตอนถัดไปได้ รวมไปถึงการทำหนดวันที่ของข้อมูลที่จะนำมาวิเคราะห์ หรือการทำหนดขอบเขต เพราะข้อมูลที่เลือกมา มีการยัดเตหอย่างต่อเนื่อง จึงจำเป็นต้องทำหนดขอบเขต ซึ่งในที่นี้จะใช้ข้อมูลปัญหาเส้นเลือดฟอยที่แจ้งเข้ามาในช่วงเดือนพฤษภาคม - เดือนมิถุนายน ปีพ.ศ. 2565 และเมื่อทำความสะอาดเรียบร้อยแล้วจะมีข้อมูลรวมทั้งหมด 15 คอลัมน์ 61,523 แถว (รวมส่วนหัวของตาราง)

```

Cmd 9
1 df = df[df['comment'].notna()]

Command took 0.16 seconds -- by sanmooks@ais.co.th at 8/9/2022, 10:03:26 AM on TRAINEE_DA

Cmd 10
1 df = df[df['province'] == "กรุงเทพมหานคร"]

Command took 0.07 seconds -- by sanmooks@ais.co.th at 8/9/2022, 10:03:27 AM on TRAINEE_DA

Cmd 11
1 df = df[~(df['comment'].str.contains(' goo'))]

Command took 0.09 seconds -- by sanmooks@ais.co.th at 8/9/2022, 10:03:27 AM on TRAINEE_DA

Cmd 12
1 df = df[(df['state'] == 'รอรับเรื่อง') | (df['state'] == 'ส่งเรื่องแล้ว') | (df['state'] == 'เสร็จสิ้น')]

Command took 0.10 seconds -- by sanmooks@ais.co.th at 8/9/2022, 10:03:28 AM on TRAINEE_DA

```

ภาพที่ 3.13 ตัวอย่าง Code เพื่อคัดแยกส่วนที่ไม่ต้องการ

	id	type	org	comment	ticket_id	coords	photo	after_photo	address	district	subdistrict	province	time
0	104979	รีบูน ชั้น ชาติ	เพื่อน	ไฟเบอร์ไฟฟ้าติดใน ห้อง	2022- 4F7NNZ	[100.37604276909 13.716807044081]	https://storage.googleapis.com/traffy_public_b...	29 ซอย เกษตรธิ 27 แขวงบางนา เขต บาง นา กรุงเทพฯ	บางนา	บางนา	กรุงเทพมหานคร	2022-0 09 03:15:079391	
1	104978	รีบูน ชั้น ชาติ	เพื่อน	1. เนื้อเยื่อกระดูก หัวใจ เป็นสี เขียว	2022- ADYKUV	[100.546365759324 13.7218863042152]	https://storage.googleapis.com/traffy_public_b... https://storage.googleapis.com/traffy_public_b...	54 ซอย รามคำแหง แขวง คลองสามวา เขต คลอง สามวา กรุงเทพฯ	คลอง สามวา	คลอง สามวา	กรุงเทพมหานคร	2022-0 09 02:55:696801	
2	104977	รีบูน ชั้น ชาติ	เพื่อน	บางส่วนของกระดูก สมองได้ร้าบ หัก ขาด	2022- 3V2K2WR	[100.57535, 13.797149]	https://storage.googleapis.com/traffy_public_b...	121 ซอย โชคชัย 22 แขวง สามเสนนนท์ เขต ห้วยขวาง กรุงเทพฯ	สามเสนนนท์	สามเสนนนท์	กรุงเทพมหานคร	2022-0 09 02:19:137591	
3	104976	รีบูน ชั้น ชาติ	เพื่อน	กระดูกกระดูก หัวใจ หายไป หายไป	2022- A8MYUK	[100.473346 13.787717]	https://storage.googleapis.com/traffy_public_b... https://storage.googleapis.com/traffy_public_b...	62 ถนน รุ่งเรือง แขวง บางปะอิน เขต บางปะ อิน กรุงเทพฯ	บางปะ อิน	บางปะ อิน	กรุงเทพมหานคร	2022-0 09 02:14:972591	
4	104975	รีบูน ชั้น ชาติ	เพื่อน	มีรอยบาก ไม่สามารถ รักษา รักษาไม่ ได...	2022- HVAPP2	[100.59833737451 13.8148029620789]	https://storage.googleapis.com/traffy_public_b... https://storage.googleapis.com/traffy_public_b...	239/61 ถนนรัชดาภิเษก แขวงหลักสอง เขต หลัก สอง กรุงเทพฯ	หลักสอง	หลักสอง	กรุงเทพมหานคร	2022-0 09 01:36:8799	
...
62542	109602	รีบูน ชั้น ชาติ	เพื่อน	ไม่แน่ใจไม่ได้ขอรับ ช่วย	2022- FEV1N3	[100.59917398811 13.7491003214901]	https://storage.googleapis.com/traffy_public_b...	436 ถนน พระราม 9 แขวง บางกะปิ เขต ห้วยขวาง กรุงเทพฯ	บางกะปิ	บางกะปิ	กรุงเทพมหานคร	2022-0 02 15:01:469491	
62543	109601	รีบูน ชั้น ชาติ	เพื่อน	ประกายกระดูกที่ 133 มีความดันสูง ไม่สามารถรักษา ^{...}	2022- 88W7CC	[100.371115365895 13.6378178132873]	https://storage.googleapis.com/traffy_public_b... https://storage.googleapis.com/traffy_public_b...	4/11 ถนนรัชดาภิเษก แขวง บางเขน เขต บาง เขน กรุงเทพฯ	บางเขน	บางเขน	กรุงเทพมหานคร	2022-0 02 14:44:49792	
62544	109600	รีบูน ชั้น ชาติ	เพื่อน	แมลงสาบตื้อที่หัว ล้วนล้ำเข้ามาในร่อง เส้นเลือด...	2022- HN4C73	[100.527912 13.778553]	https://storage.googleapis.com/traffy_public_b...	214 ถนนรัชดาภิเษก แขวง บางเขน เขต บาง เขน กรุงเทพฯ	บางเขน	บางเขน	กรุงเทพมหานคร	2022-0 02 14:43:3939	
62545	109599	รีบูน ชั้น ชาติ	เพื่อน	สีสันดี แต่กระดูก แข็งเป็นปูน มีไฟ สว่าง...	2022- BTR43K	[100.53584219761 13.7101369785007]	https://storage.googleapis.com/traffy_public_b...	2nd floor Makro แขวง ห้วยขวาง เขต ห้วย ขวาง กรุงเทพฯ	ห้วยขวาง	ห้วยขวาง	กรุงเทพมหานคร	2022-0 02 14:41:030631	
62546	109598	รีบูน ชั้น ชาติ	เพื่อน	หัวใจล้มเหลว บุบบัน เสียชีวิต ^{...}	2022- F4HVYU	[100.52160322662 13.8063322591836]	https://storage.googleapis.com/traffy_public_b...	แยกบ้านไว้ แขวง หนองบัว เขต หนองบัว ^{...}	หนองบัว	หนองบัว	กรุงเทพมหานคร	2022-0 02 14:40:89940	

61523 rows × 15 columns

ภาพที่ 3.14 ตัวอย่างข้อมูล

3.3.6 แก้ไขปัญหา

จากปัญหาที่พบคือ ประเภทของปัญหาที่แจ้งเข้ามานั้น ประเภท “อื่น ๆ” มีจำนวนมากเกินไป หากไม่ทำการแก้ไข จะทำให้ไม่สามารถวิเคราะห์ข้อมูลต่อได้ หรือหากสามารถวิเคราะห์ได้ผลการวิเคราะห์อาจจะออกมาคลาดเคลื่อน หรือผลลัพธ์ออกมาไม่ตรงตามความคาดหมาย ดังนั้นควรแก้ไขปัญหานี้ด้วยการแจกแจงปัญหาประเภท “อื่น ๆ” ให้ได้มากที่สุดเพื่อที่จะได้สามารถนำไปวิเคราะห์ต่อได้อย่างมีประสิทธิภาพ

```
Out[13]: type
          อื่นๆ      57261
          ความปลอดภัย     1069
          ทางเท้า        1028
          จราจร         900
          น้ำท่วม        562
          ขยาย          495
          เสนอแนะ       201
Name: ticket_id, dtype: int64
```

ภาพที่ 3.15 จำนวนของข้อมูลแต่ละประเภทที่มีอยู่เดิม

จึงมีการแก้ปัญหาโดยการใช้ Natural Language Processing (NLP) เข้ามาจัดการกับรายละเอียดปัญหาที่แจ้งเข้ามา Natural Language Processing (NLP) แปลเป็นไทยได้ว่า การประมวลผลภาษาธรรมชาติ ภาษาธรรมชาติ (Natural Language) คือภาษาที่มนุษย์ใช้สื่อสารกันทั่วไป เมื่อมาร่วมกับคำว่า Processing จึงกลายเป็นคำว่า Natural Language Processing (NLP) หรือ การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ ช่วยให้อุปกรณ์คอมพิวเตอร์ต่าง ๆ สามารถสื่อสารกับมนุษย์ได้ด้วยการใช้งานภาษาของเครื่องเอง และดำเนินการทำางานต่าง ๆ ที่เกี่ยวข้องกับภาษาได้ จึงจำเป็นต้องใช้ NLP มาจัดการกับรายละเอียดปัญหาที่แจ้งเข้ามา ที่เป็นภาษาที่มนุษย์ใช้สื่อสารกัน

เนื่องจากภาษาที่มนุษย์สื่อสารมีอยู่เป็นจำนวนมาก จึงต้องเลือกใช้ NLP ให้เหมาะสมกับภาษาของข้อมูลที่จะทำการประมวลผล ในกรณีรายละเอียดปัญหาที่แจ้งเข้ามาเป็นภาษาไทย จึงจะนำ PyThaiNLP (Thai Natural Language Processing in Python) ซึ่งเป็น Python Library ที่ถูกพัฒนาขึ้นเพื่องานวิจัยและพัฒนาการประมวลภาษาธรรมชาติภาษาไทยโดยเฉพาะ มาใช้งาน

PyThaiNLP ประกอบไปด้วยการทำงานที่หลากหลาย และสามารถเรียกใช้ได้หลายรูปแบบ จึงจะยกมากล่าวแค่การทำงานที่จะใช้งานดังต่อไปนี้

- **Word Tokenization** การตัดคำ
- **Contextual Extraction** กระบวนการสกัดข้อมูลที่มีโครงสร้างอุอกมาจากข้อมูลรูปแบบข้อความทั่วไป

ความสามารถในการตัดคำ (Word Tokenization) ของ PyThaiNLP เป็นการตัดประโยคภาษาไทยอุอกมาเป็นคำอย่างถูกต้องตามหลักภาษา ซึ่งภาษาอังกฤษจะมีช่องว่างระหว่างคำอยู่แล้ว ทำให้ง่ายต่อการแบ่งคำ สำหรับภาษาไทยนั้นมีความยุ่งยากมากกว่าเนื่องจากเป็นภาษาที่ไม่มีช่องว่างระหว่างคำ [10] ซึ่งเมื่อหลังจากการตัดคำเรียบร้อยแล้ว จะสามารถนำข้อมูลไปประมวลผลในแบบต่าง ๆ ได้ ดังตัวอย่างการใช้งานการตัดคำ (Word Tokenization) ภาพที่ 3.16

Word

Default word tokenizer ("newmm") use maximum matching algorithm.

```
[25]: from pythainlp import word_tokenize
text = "ก็จะรู้ความชั่วร้ายที่ทำไว้ และคงจะไม่ยอมให้ทำงานหนังคน"
print("default (newmm):")
print(word_tokenize(text))
print("\nnewmm and keep_whitespace=False:")
print(word_tokenize(text, keep_whitespace=False))

default (newmm):
['ก็', 'จะ', 'รู้', 'ความ', 'ชั่วร้าย', 'ที่', 'ทำ', 'ไว้', 'และ', 'คงจะ', 'ไม่', 'ยอมให้', 'ทำงาน', 'หนัง', 'คน']
newmm and keep_whitespace=False:
['ก็', 'จะ', 'รู้', 'ความ', 'ชั่วร้าย', 'ที่', 'ทำ', 'ไว้', 'และ', 'คงจะ', 'ไม่', 'ยอมให้', 'ทำงาน', 'หนัง', 'คน']
```

ภาพที่ 3.16 ตัวอย่างการใช้งานการตัดคำ (Word Tokenization)

โดยเริ่มการตัดคำ (Word Tokenize) ด้วยการนำรายละเอียดปัญหาที่แจ้งเข้ามา ทำการรวบรวม แบ่งเป็นกลุ่มย่อยเพื่อลดเวลาการตัดคำ เพราะขั้นตอนนี้ใช้เวลามาก หากข้อมูลที่ต้องการตัดคำมีปริมาณมาก เวลาที่ใช้ในการตัดจะมากตามไปด้วย

```

Cmd 91
1 text_other1 = ""
2 text_other2 = ""
3 text_other3 = ""
4 for index, row in df.iterrows():
5     if index <= 20000:
6         text_other1 = text_other1 + " "+row['comment']
7     elif index <= 40000:
8         text_other2 = text_other2 + " "+row['comment']
9     elif index <= 60000:
10        text_other3 = text_other3 + " "+row['comment']

Command took 1.19 minutes -- by sanmooks@ais.co.th at 7/22/2022, 9:32:14 AM on TRAINEE_DA

Cmd 92
1 sum_text = word_tokenize(text_other1)

Command took 39.02 minutes -- by sanmooks@ais.co.th at 7/22/2022, 9:33:35 AM on TRAINEE_DA

Cmd 93
1 sum_text += word_tokenize(text_other2)
2 sum_text += word_tokenize(text_other3)

Command took 1.20 hours -- by sanmooks@ais.co.th at 7/22/2022, 10:47:30 AM on TRAINEE_DA

```

ภาพที่ 3.17 การตัดคำ (Word Tokenization)

เมื่อทำการตัดคำเรียบร้อยแล้ว จะเข้าสู่การทำ Contextual Extraction หรือ กระบวนการสกัดข้อมูลที่มีโครงสร้างของมาจากการหักลูปแบบข้อความทั่วไป เพื่อแสดงกลุ่มคำ จำนวน และความถี่ โดยจะใช้ Wordcloud ที่เป็น Python Library

Wordcloud หรือกลุ่มคำที่จะแสดงขนาดตามจำนวนความถี่ของคำ โดยเมื่อมีความถี่มากขนาดของคำจะมากตามไปด้วย ทำให้ในหลายวงการหรือในทาง Data Analytics มักจะใช้แสดงคำที่นิยมใช้ในบทความ เป็นต้น

Wordcloud ที่เป็น Python Library ประกอบด้วย Parameter สำคัญดังนี้

- **Font Path** ที่อยู่ของไฟล์ Font ที่จะให้ Wordcloud แสดงผล
- **Stopword** เป็นการลบคำที่ไม่ต้องการจะให้แสดงใน Wordcloud จากข้อความทั้งหมด ในที่นี่คือคำสันธานที่ใช้เชื่อมประโยค เช่นคำว่า และ แล้ว จึง และอื่น ๆ เพื่อให้ Wordcloud นับแต่คำสำคัญ หรือ Keyword ที่ต้องการ

```

1 import matplotlib.pyplot as plt # ใช้ในการแสดง Word Cloud
2 from pythainlp.tokenize import word_tokenize # ใช้ในการตัดคำ
3 from wordcloud import WordCloud # ใช้ทำ Word Cloud
4 from pythainlp.corpus import thai_stopwords # ใช้ลบคำที่ไม่ใช้ออก
5 wordcloud = WordCloud( prefer_horizontal=2,
6                         scale=2,
7                         max_words=100,
8                         font_path= endfontfile,
9                         # path ตั้ง Font
10                        stopwords=list(thai_stopwords()) + ['ที่นี่', 'แจ้ง', 'เรื่อง', 'ปัญหา', 'บริเวณ', 'เขต', 'นัด', 'ขอ'],
11                        background_color="white", # ตั้งค่าพื้นหลัง
12                        regexp=r"[\u0E00-\u0E7Fa-zA-Z]+"\# ป้องกัน bug วรรณยุกต์
13 ).generate(' '.join(sum_text))
14 plt.figure(figsize=[10,10])
15 plt.imshow(wordcloud, cmap=plt.cm.gray, interpolation='bilinear')
16 plt.axis("off")
17 plt.show()

```

ภาพที่ 3.18 Code สร้าง Wordcloud ด้วยภาษา Python

และนำไปแสดงผลด้วย Matplotlib ที่เป็น Python Library เพื่อแสดงผล Wordcloud ดังภาพที่ 3.19



ภาพที่ 3.19 Wordcloud จากข้อมูลปัญหาประเภท “อื่น ๆ”

จากภาพที่ 3.19 เป็น Wordcloud ที่สร้างจากข้อมูลปัญหาที่แจ้งเข้ามาประเภท “อื่น ๆ” ทั้งหมด จะเห็นได้ว่า มีคำที่ตรงกับประเภทเดิมอยู่เป็นจำนวนมาก ดังนั้นแล้วควรจัดคำเหล่านั้นกลับสู่ประเภทเดิม

```

def generate_new_type(comment):
    word = ['ຂໍພະ']
    for i in word:
        if i in comment:
            return 'ຂໍພະ'

    word = ['ຮອດີຕີ', 'ຮອຈອດ', 'ຈອດຮອ', 'ຈຣາຈຣ']
    for i in word:
        if i in comment:
            return 'ຈຣາຈຣ'

    word = ['ທ່າງເຫົວ', 'ໜົດນາທ', 'ທ່າງເຕັ້ນ']
    for i in word:
        if i in comment:
            return 'ທ່າງເຫົວ'

    word = ['ນ້າກ່ວມ', 'ໜ່ວມ', 'ຄລອງ', 'ນ້າ']
    for i in word:
        if i in comment:
            return 'ນ້າກ່ວມ'

    word = ['ອັນດຽຍ', 'ອັນດີເໜີ', 'ປລອດສັຍ', 'ຄນ', 'ກລ້ອງ']
    for i in word:
        if i in comment:
            return 'ຄວາມປລອດສັຍ'
    #6
    word = ['ເສນອແນະ', 'ເສນອ', 'ແນະ', 'ແນຍຝາ', 'ທ່ານບ', 'ອໝາກ', 'ດ້ວຈ']
    for i in word:
        if i in comment:
            return 'ເສນອແນະ'

```

ກາພທີ 3.20 ຕ້ວອຍ່າງ code ສໍາຫຼັບແບ່ງປະເທດເດີມ

ຊື່ນອກຈາກການແສດງ Wordcloud ແລ້ວ Wordcloud ທີ່ເປັນ Python Library ຍັງສາມາຮັດຮະບຸລຳດັບຂອງຄຳທີ່ນໍາຂຶ້ນໄປແສດງເປັນ Wordcloud ໄດ້ ຊື່ຈະໃຫ້ລຳດັບນີ້ປະກອບການຕັດສິນໃຈໃນການແບ່ງປະເທດໃນຂັ້ນຕອນຕ່ອໄປ ຊື່ຄ້າຫາກພິຈາລະນາດູແລ້ວ ຄຳທີ່ອູ່ໃນລຳດັບຕົ້ນ ຖ້າ ຄຳໃຫຍ່ທີ່ໄມ່ມີຄວາມໝາຍຫຼືໄມ່ເປັນຄຳສຳຄັນເທົ່າທີ່ຄວຮ ກີ່ຈະທຳການເປົ້າມາໃນລຳດັບຄັດໄປທັນທີ

```

1 wordcloud.layout_

```

```

Out[32]: [((‘ຄນນ’, 1.0), 113, (23, 175), None, ‘rgb(112, 207, 87)’),
((‘ຮ’, 0.6110657172872651), 91, (93, 36), None, ‘rgb(68, 59, 132)’),
((‘ທ່າງເຫົວ’, 0.5100968833897513), 83, (91, 132), None, ‘rgb(38, 129, 142)’),
((‘ຕີດ’, 0.41122913505311076), 75, (2, 85), None, ‘rgb(48, 105, 142)’),
((‘ຄນ’, 0.35555036768997317), 70, (8, 2), None, ‘rgb(39, 128, 142)’),
((‘ອັນດຽຍ’, 0.3390918641298004), 44, (150, 1), None, ‘rgb(45, 112, 142)’),
((‘ໄຟ’, 0.33792459437375977), 44, (72, 123), None, ‘rgb(32, 146, 140)’),
((‘ທ່າ’, 0.3153962880821758), 40, (49, 19), None, ‘rgb(221, 227, 24)’),
((‘ນ້ານ’, 0.29613633710750553), 29, (116, 1), None, ‘rgb(43, 116, 142)’),
((‘ໜີ’, 0.293334889693008), 29, (72, 376), None, ‘rgb(64, 69, 136)’),
((‘ທ່າມ’, 0.29170071203455117), 21, (90, 237), None, ‘rgb(72, 193, 110)’),
((‘ຈຸດ’, 0.27244076105988096), 20, (60, 63), None, ‘rgb(46, 109, 142)’),
((‘ໜົດນາທ’, 0.2659040504260535), 20, (91, 173), None, ‘rgb(45, 112, 142)’),
((‘ຮັບກວນ’, 0.26205206023111943), 20, (175, 310), None, ‘rgb(142, 214, 69)’),
((‘ແຍກ’, 0.2558655305241041), 20, (172, 178), None, ‘rgb(176, 221, 47)’),

```

ກາພທີ 3.21 ລຳດັບຂອງຄຳເຮືອງຕາມຈຳນວນທີ່ໜ້າ

หลังจากแบ่งข้อมูลเข้าประเภทเดิมแล้ว จะนำข้อมูลที่เหลือมาสร้างเป็น Wordcloud เพื่อพิจารณาคำที่มีความถี่มากที่สุดไปสร้างเป็นประเภทใหม่ โดยวิธีการแบ่งประเภท จะพิจารณาจากคำที่มีความถี่มากและประเภทเดิมที่มีอยู่ หากคำนั้นสามารถจัดเป็นประเภทเดิมได้ ก็จะทำการใช้คำนั้นเป็นหนึ่งในการแบ่งประเภทเดิม หากคำนั้นไม่สามารถจัดเป็นประเภทเดิมได้ จึงจะแบ่งออกเป็นประเภทใหม่



ภาพที่ 3.22 Wordcloud จากข้อมูลที่เหลือจากการแบ่งประเภทเดิม

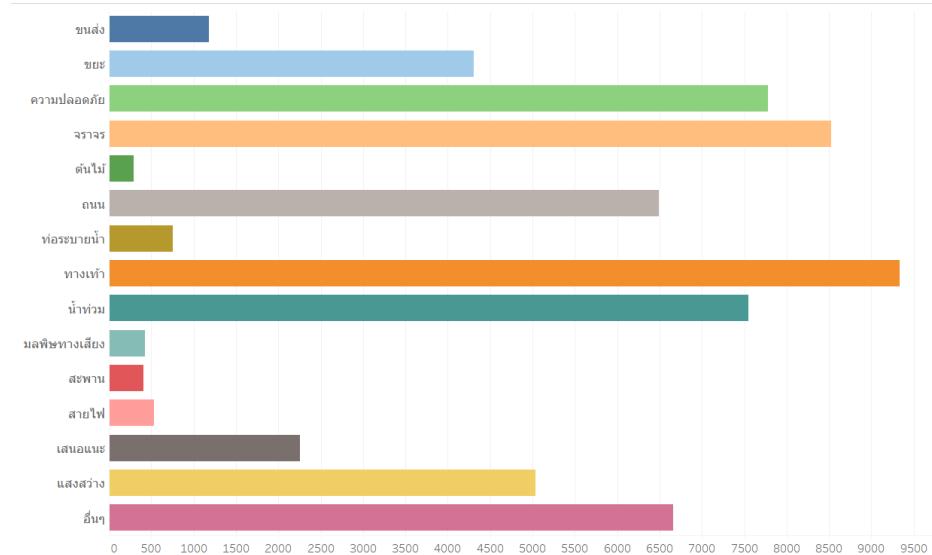
จากภาพที่ 3.22 จะเห็นได้ชัดว่ามีคำว่า “ถนน” ปรากฏเด่นชัด เมื่อพิจารณาจากประเภทเดิมที่มีอยู่พบว่าไม่สามารถจัดคำว่าถนนเป็นหนึ่งในประเภทเดิมได้ จึงทำการแบ่งประเภทใหม่เป็นประเภท “ถนน”

นอกจากนั้นใน Wordcloud ยังปรากฏคำว่า “ไฟ” จากการสังเกตพบว่ามีคำที่มาคู่กับคำว่า “ไฟ” ได้แก่คำว่า “สายไฟ” และ “ไฟ ส่องสว่าง” ซึ่งทั้งสองคำสามารถจัดเป็นประเภทใหม่ ได้แก่ ประเภท “แสงสว่าง” และ ประเภท “สายไฟ”

ท้ายสุดแล้ว ปัญหาจะถูกแบ่งประเภทใหม่ออกรเป็น 15 ประเภท ได้แก่ 1.ทางเท้า 2.จราจร 3.ความปลอดภัย 4.น้ำท่วม 5.ถนน 6.แสงสว่าง 7.ขยาย 8.เสนอแนะ 9.ขนส่ง 10.ท่อระบายน้ำ 11.สายไฟ 12.มลพิษทางเสียง 13.สะพาน 14.ต้นไม้ และ 15.อื่น ๆ

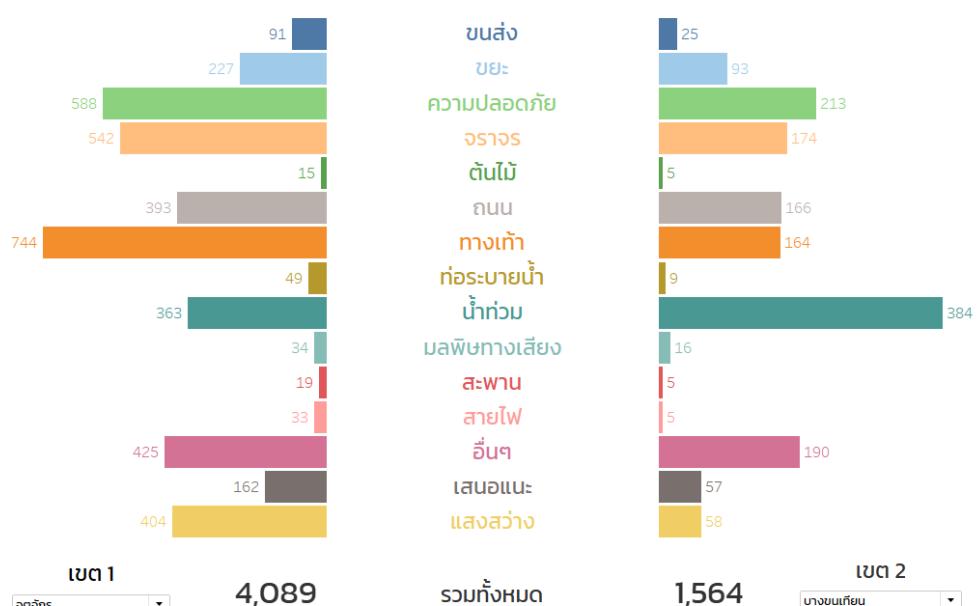
3.3.7 วิเคราะห์ข้อมูลเชิงลึก

ในขั้นตอนการวิเคราะห์ข้อมูลเชิงลึก จะมีการนำ Tableau มาใช้งานร่วมด้วย เพื่อเป็นตัวช่วยในการวิเคราะห์ข้อมูลได้มากขึ้น เช่น การนำข้อมูลมาแสดงเป็นกราฟ หรือแผนภูมิในรูปแบบต่าง ๆ เพื่อการมองเห็นภาพรวมได้ดียิ่งขึ้น



ภาพที่ 3.23 การนำข้อมูลมาแสดงเป็นกราฟเพื่อวิเคราะห์ข้อมูล

จำนวนปัญหาในแต่ละประเภทเบื้องต้น



ภาพที่ 3.24 การนำกราฟมาเทียบกันเพื่อวิเคราะห์ข้อมูล

รวมไปถึงการสร้างคอัมน์ใหม่ เพื่อการหาข้อมูลเชิงลึกอีกอีกหนึ่ง เช่น เพศของผู้รายงานปัญหา เป็นต้นจากภาพที่ 3.27 เป็นการจำแนกเพศโดยคร่าวจากการกำหนดคำเฉพาะเจาะจงที่ใช้จำแนกเพศ นั้น ๆ จากการรายงานปัญหาสืบเนื่องเลือดฝอยเข้ามา ทำให้ได้มาซึ่งเพศ ซึ่งสามารถนำไปสร้างกราฟเพื่อวิเคราะห์ข้อมูลเชิงลึกดูได้ว่า เพศไหนมีการแจ้งปัญหารือไม่ได้เข้ามามากกว่ากัน

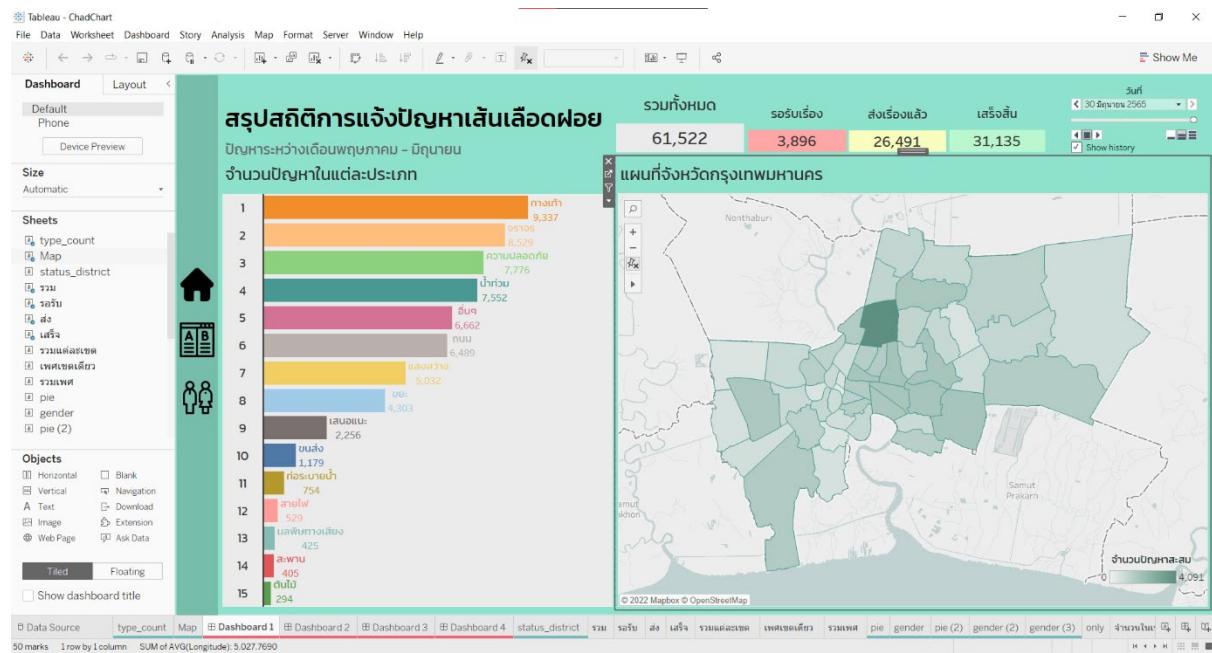
```
def generate_gender(comment):
    m_word = ['ครับ', 'ผม', 'กระผม', 'นจะคับ', 'จะ', 'คับ']
    f_word = ['ค่ะ', 'คะ', 'หນ', 'ดีลัน', 'นะคะ', 'ค่า']
    for i in m_word:
        if i in comment:
            return 'Male'
    for j in f_word:
        if j in comment:
            return 'Female'
    return 'Null'
```

ภาพที่ 3.25 ตัวอย่าง code จำแนกเพศโดยคร่าว

3.3.8 สร้าง Visualization และ Dashboard

เมื่อวิเคราะห์หาข้อมูลเชิงลึกเรียบร้อยแล้ว ถัดมาจะเป็นการสร้างแผนภูมิ หรือกราฟ (Visualization) เพื่อให้ผู้ฟังเข้าใจยิ่งขึ้น ในบางครั้งเมื่อทำหรือวิเคราะห์ข้อมูลเรียบร้อยแล้ว ก็ต้องทำกราฟออกมาให้สวยงาม และเข้าใจง่าย เพื่อให้ผู้อื่นสามารถอ่านกราฟแล้วเข้าใจได้มากที่สุด ขั้นตอนนี้จึงเป็นสิ่งจำเป็น

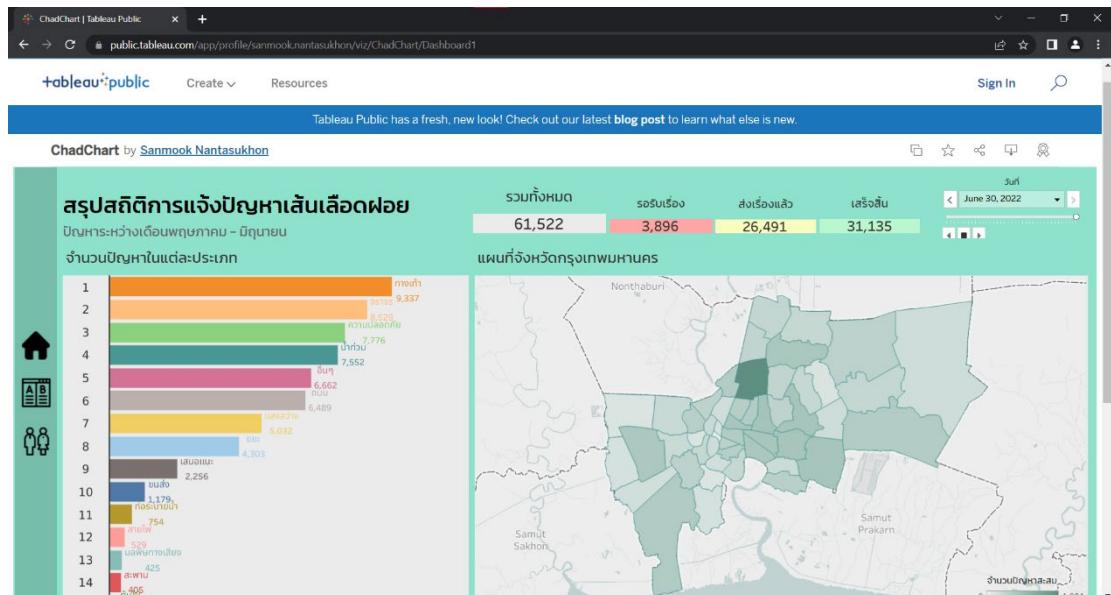
หลังจากที่ได้ทำ Visualization เพื่อนำเสนอข้อมูลเชิงลึกในรูปแบบของกราฟต่างๆ ก็ต้องนำกราฟต่างๆเหล่านั้นมารวมเข้าด้วยกันเป็น Dashboard เพื่อการนำเสนอโครงงาน และเรียบเรียงกราฟให้มีประสิทธิภาพมากที่สุด เพื่อเตรียมการนำเสนอต่อไป



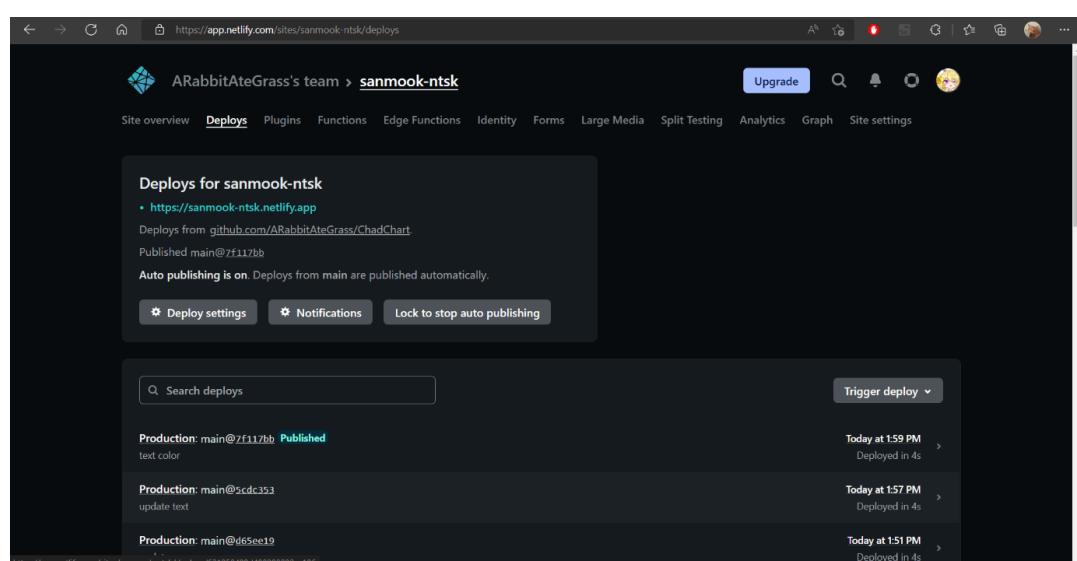
ภาพที่ 3.26 การสร้าง Visualization และ Dashboard ใน Tableau

3.3.9 นำ Dashboard เพย์พร์บัน Website

เมื่อสร้าง Visualization และ Dashboard ใน Tableau และสามารถนำขึ้นไปแสดงผล และเพย์พร์บัน Website ได้ ผ่าน Tableau Public ซึ่งเป็นตัวช่วยในการแสดงผล Dashboard บน Website ที่มีการให้ Code HTML เพื่อนำ Dashboard ไปแสดงบน Website ต่าง ๆ ได้ ในที่นี้จะแสดงผล Dashboard บน Website ส่วนตัวที่จัดทำขึ้นและเพย์พร์บัน Netlify



ภาพที่ 3.27 Tableau Public ตัวช่วยในการแสดงผล Dashboard บน Website



ภาพที่ 3.28 เพย์พร์ Website ผ่าน Netlify

```

File Edit Selection View Go Run Terminal Help
Get Started index.html
index.html - Visual Studio Code
104     <div class="row">
105         <div class="col-lg-4 ms-auto"><p class="lead">
106             Traffy x TeamChadchart is the Chatbots provide information on capillary (micro issues) problems in Bangkok.
107             They encourage users to report everyday city problems □ uneven pavements, blocked access, no garbage collection, etc □ so they can be fixed.
108         </p>
109         <div class="col-lg-4 me-auto "><p class="lead">
110             This is my cooperative education project which was made in my 1st term of 4th year in faculty of Information Technology.
111             There are the full report and Github of this project, link in the icon below.
112         </p>
113     </div>
114 </div>
115 </section>
116 <!-- Footer-->
117 <footer class="footer text-center">
118     <div class="container">
119         <div class="row">
120             <!-- Footer Location-->
121             <div class="col-lg-4 mb-5 mb-lg-0">
122                 <h4 class="text-uppercase mb-4">Data from</h4>
123                 <p class="lead mb-0">
124                     <a href="https://www.traffy.in.th/">traffy.in.th</a>
125                 </p>
126             </div>
127             <!-- Footer Social Icons-->
128             <div class="col-lg-4 mb-5 mb-lg-0">
129                 <h4 class="text-uppercase mb-4">Around the web</h4>
130                 <a class="btn btn-outline-light btn-social mx-1" href="#"><i class="fa-solid fa-book"></i></a>
131                 <a class="btn btn-outline-light btn-social mx-1" href="https://github.com/ARabbitAteGrass/ChadChart"><i class="fab fa-fw fa-github"></i></a>
132             </div>
133             <!-- Footer About Text-->
134             <div class="col-lg-4">
135                 <h4 class="text-uppercase mb-4">About</h4>
136                 <p class="lead mb-0">Sammook Nantatsukon<br>
137                     Study in faculty of Information Technology,
138                     Thai-Nichi Institute of technology
139                     <a href="mailto:na.sammook_st@tni.ac.th">na.sammook_st@tni.ac.th</a>
140             </p>

```

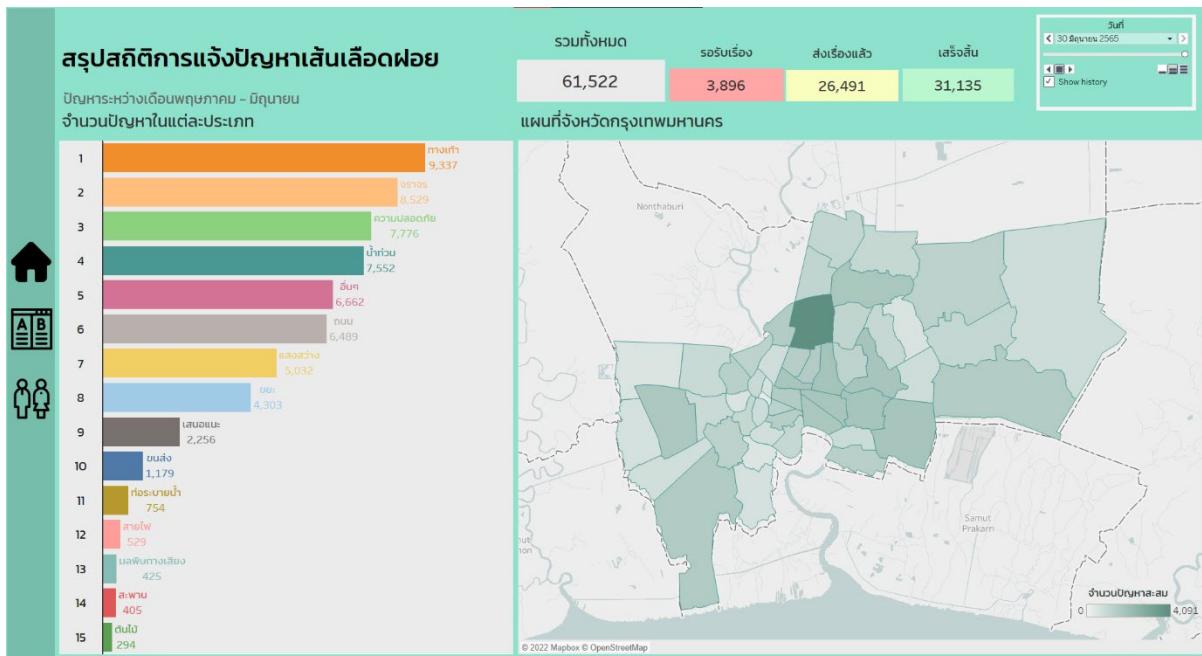
ภาพที่ 3.29 ตัวอย่าง Code ของ Website

บทที่ 4

สรุปผลการดำเนินงาน

4.1 ผลการดำเนินงานและผลการวิเคราะห์ข้อมูล

4.1.1 Dashboard 1 แสดงสรุปสถิติการแจ้งปัญหาเส้นเลือดฟอย



ภาพที่ 4.1 Dashboard แสดงสรุปสถิติการแจ้งปัญหาเส้นเลือดฟอย

ประกอบไปด้วย

- กราฟแสดงจำนวนปัญหาในแต่ละประเภท
- แผนที่แสดงจังหวัดกรุงเทพมหานครที่มีระดับความเข้มสีตามจำนวนปัญหา
- จำนวนรวมของข้อมูลทั้งหมด
- จำนวนรวมของข้อมูลปัญหาที่รอรับเรื่อง
- จำนวนรวมของข้อมูลปัญหาที่ส่งเรื่องแล้ว
- จำนวนรวมของข้อมูลปัญหาที่เสร็จสิ้น
- ตัวเลือกวันที่ของข้อมูลปัญหาที่แจ้งเข้ามา

จากจำนวนปัญหาที่แจ้งเข้ามาระหว่างเดือนพฤษภาคม – เดือนมิถุนายน รวมทั้งสิ้น 61,522 เรื่อง แบ่งเป็นปัญหาจัดการเสร็จสิ้นไปแล้ว 31,135 เรื่อง ปัญหาที่ส่งเรื่องให้หน่วยงานที่เกี่ยวข้องแล้ว 26,491 เรื่อง และปัญหาที่รอรับเรื่อง 3,896 เรื่อง

สำหรับ Dashboard 1 นี้จะแสดงปัญหาทั้งสิ้นที่สามารถจำแนกเป็นประเภทต่างๆได้ 15 ประเภท และเรียงตามจำนวนจากมากไปน้อย ได้ดังนี้ ปัญหาประเภทที่ 1 ทางเท้า มีจำนวน 9,337 เรื่อง, ปัญหาประเภทที่ 2 จราจร จำนวน 8,529 เรื่อง, ปัญหาประเภทที่ 3 ความปลอดภัย จำนวน 7,776 เรื่อง, ปัญหาประเภทที่ 4 น้ำท่วม จำนวน 7,552 เรื่อง, ปัญหาประเภทที่ 5 อื่น ๆ จำนวน 6,662 เรื่อง, ปัญหาประเภทที่ 6 ถนน จำนวน 6,489 เรื่อง, ปัญหาประเภทที่ 7 แสงสว่าง จำนวน 5,032 เรื่อง, ปัญหาประเภทที่ 8 ขยาย จำนวน 4,303 เรื่อง, ปัญหาประเภทที่ 9 เสนอแนะ จำนวน 2,256 เรื่อง, ปัญหาประเภทที่ 10 ขยะ จำนวน 1,179 เรื่อง, ปัญหาประเภทที่ 11 ท่อระบายน้ำ จำนวน 754 เรื่อง, ปัญหาประเภทที่ 12 สายไฟ จำนวน 529 เรื่อง, ปัญหาประเภทที่ 13 mLพิษทางเสียง จำนวน 425 เรื่อง, ปัญหาประเภทที่ 14 สะพาน จำนวน 405 เรื่อง, ปัญหาประเภทที่ 15 ต้นไม้ จำนวน 294 เรื่อง สามารถดูได้ที่กราฟด้านซ้ายของภาพที่ 4.1

โดยหากคุณตามจำนวนของปัญหาที่แจ้งเข้ามาในแต่ละเขตแล้ว จะพบว่า เขตที่มีการแจ้งปัญหาเข้ามามากที่สุด 3 ลำดับแรกได้แก่ ลำดับที่ 1 เขตตุจักร จำนวน 4,091 เรื่อง, ลำดับที่ 2 เขตหัวหมาก จำนวน 1,903 เรื่อง, ลำดับที่ 3 เขตสวนหลวง จำนวน 1,833 เรื่อง โดยสามารถดูจำนวนปัญหาที่แจ้งเข้ามาในแต่ละเขตได้ในกราฟด้านขวาของภาพที่ 4.1 โดยจะเรียงตามความเข้มสี ยิ่งสีเข้มมาก จำนวนปัญหาที่แจ้งเข้ามาก็มากตาม

นอกจากนั้น เมื่อกดไปที่แท่งประเภทของปัญหานั้น ๆ ในกราฟด้านซ้ายของภาพที่ 4.1 กราฟแผนที่ทางด้านขวาจะเปลี่ยนแปลง เป็นการเลือกประเภทของปัญหาเพื่อดูว่ามีเขตไหนแจ้งปัญหาเข้ามา โดยกราฟด้านขวาเมื่อที่เปลี่ยนแปลงไปจะแสดงจำนวนปัญหาที่แจ้งเข้ามาแต่ละเขตในวันที่ที่ได้ทำการเลือกไว้เท่านั้น

4.1.2 Dashboard 2 เทียบจำนวนสถานะในแต่ละเขต



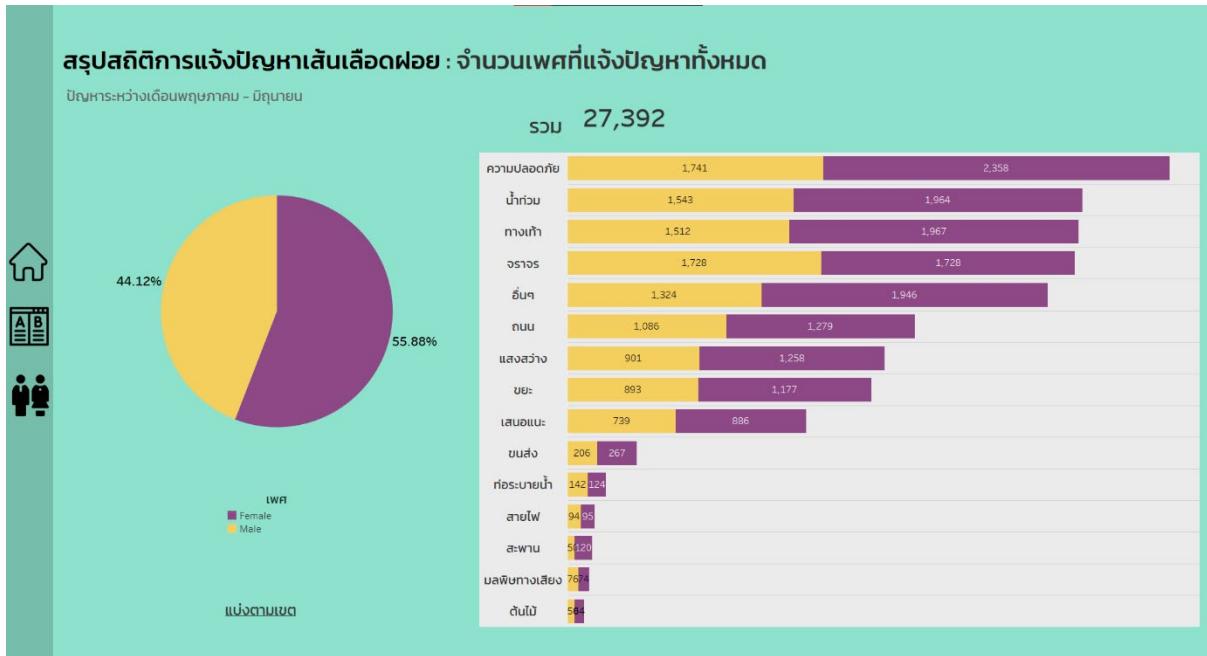
ภาพที่ 4.2 Dashboard แสดงการเทียบจำนวนสถานะในแต่ละเขต

ประกอบไปด้วย

- กราฟแสดงการเทียบจำนวนสถานะของปัญหาที่แจ้งเข้ามาในแต่ละเขต
- จำนวนรวมทั้งหมดของปัญหาในเขตนั้น
- ตัวเลือกเขตที่จะเปรียบเทียบ

Dashboard 2 นี้เป็นการแสดงการเปรียบเทียบจำนวนสถานะของปัญหาที่แจ้งเข้ามาระหว่างเขต 2 เขตที่สามารถเลือกเปรียบเทียบได้ โดยจะมีสถานะทั้ง 3 สถานะดังนี้ สถานะที่ 1 รอรับเรื่อง, สถานะที่ 2 ส่งเรื่องแล้ว, สถานะที่ 3 เสร็จสิ้น และมีการแสดงจำนวนปัญหาในสถานะนั้น ๆ เป็นเปอร์เซ็นต์เทียบกับในเขตที่เลือก แต่หากจะดูจำนวนปัญหาในสถานะนั้น ๆ ให้อ่านมาส์เพรเวนแท่งกราฟนั้น ๆ จะสามารถดูจำนวนก่อนการคิดเป็นเปอร์เซ็นต์ได้ รวมไปถึงมีการแสดงจำนวนปัญหาทั้งหมดในเขตที่เลือกข้างใต้กราฟ พร้อมกับตัวเลือกเขตด้านข้าง

4.1.3 Dashboard 3 จำนวนเพศที่แจ้งปัญหาทั้งหมด



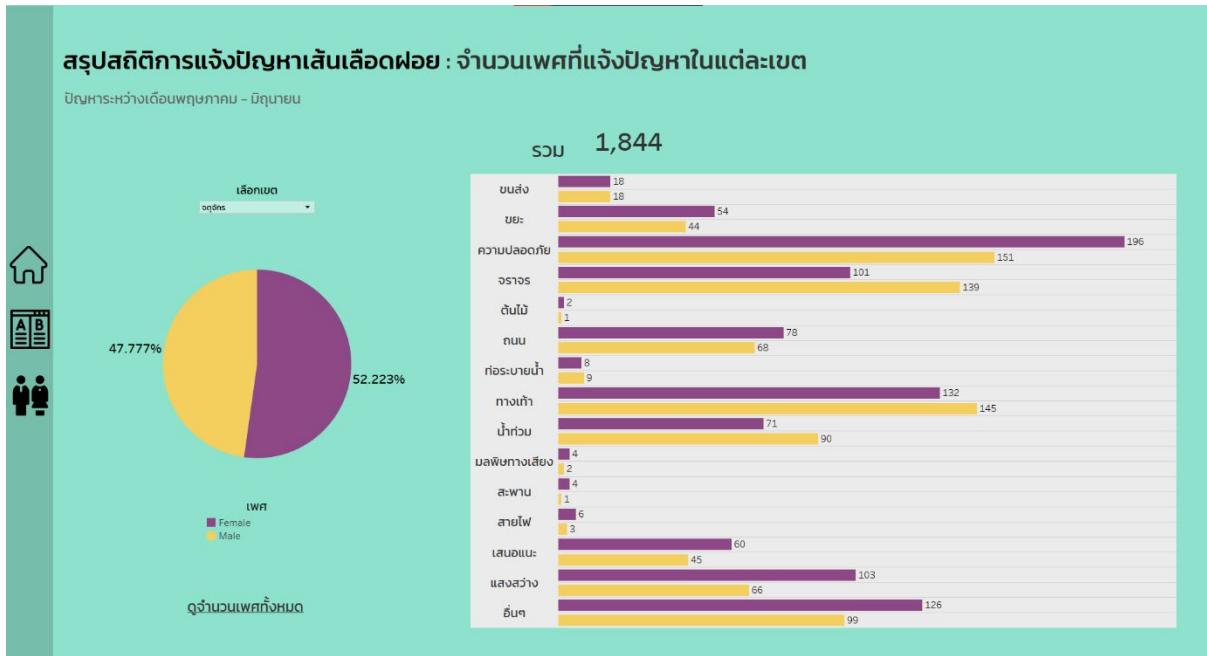
ภาพที่ 4.3 Dashboard แสดงจำนวนเพศที่แจ้งปัญหาทั้งหมด

ประกอบไปด้วย

- กราฟแสดงจำนวนเพศที่แจ้งปัญหาเข้ามา โดยแยกตามประเภทของปัญหา
- กราฟแสดงจำนวนเพศจากทั้งหมดโดยคิดเป็นเปอร์เซ็น
- จำนวนรวมของจำนวนเพศที่แจ้งปัญหาเข้ามา

Dashboard 3 นี้จะแสดงกราฟแสดงจำนวนเพศที่แจ้งปัญหาเข้ามาทั้งหมด โดยที่เพศจะจำแนกจากข้อมูลปัญหาที่แจ้งเข้ามาด้วยการตรวจสอบว่ามีการใช้คำลงท้าย และคำสรรพนามบุรุษที่ 1 ตรงกับเพศใด เป็นการจำแนกโดยคร่าว เพื่อให้ได้มาซึ่งเพศมาสร้างกราฟเพื่อวิเคราะห์เพิ่มได้ ซึ่งตัวกราฟจะแบ่งเป็น 2 ประเภท 1 ด้านซ้ายของภาพที่ 4.3 เป็นกราฟรูปวงกลม และ 2 ด้านขวาของภาพที่ 4.3 เป็นกราฟแบบแท่ง ซึ่งกราฟรูปวงกลมจะแสดงสัดส่วนระหว่างเพศชาย และหญิง เป็นแบบเปอร์เซ็นต์ ซึ่งถ้าเลื่อนมาส์ไปทางบนสัดส่วนนั้น ๆ ก็จะทราบจำนวนปัญหาที่แจ้งของเพศนั้นๆ และบนกราฟแบบแท่งก็มีจำนวนรวมของเพศทั้งหมดแสดงอยู่ เช่นกัน โดยที่กราฟแบบแท่ง จะแสดงจำนวนปัญหาที่แจ้งเข้ามาแต่ละเพศ โดยจำแนกตามประเภทของปัญหาทั้ง 15 ประเภท

4.1.4 Dashboard 4 จำนวนเพศที่แจ้งปัญหาในแต่ละเขต



ภาพที่ 4.4 Dashboard แสดงจำนวนเพศที่แจ้งปัญหาในแต่ละเขต

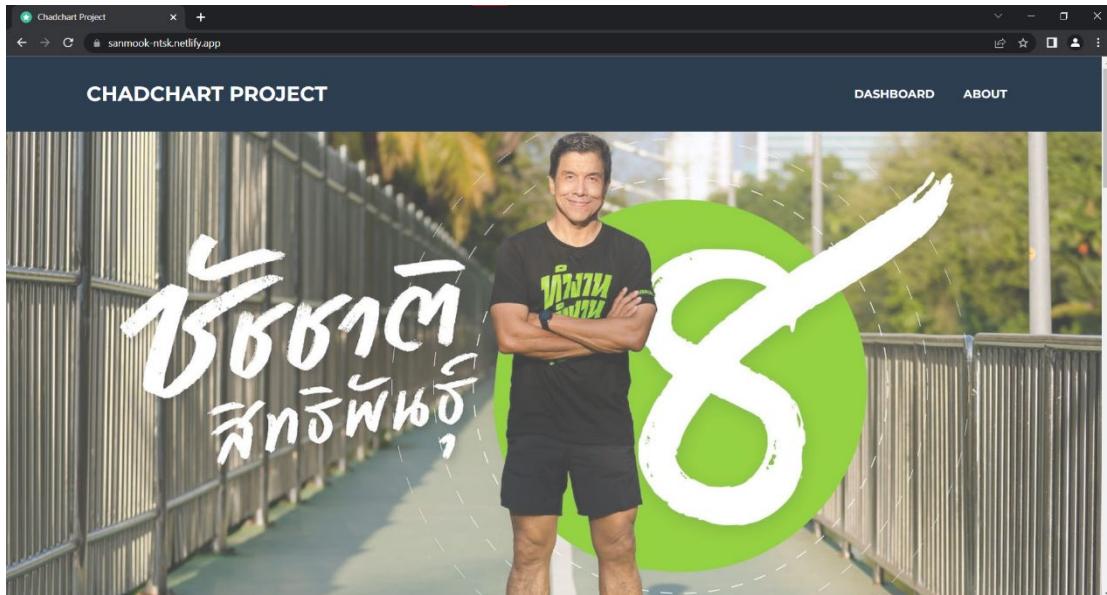
ประกอบไปด้วย

- กราฟแสดงจำนวนเพศที่แจ้งปัญหาเข้ามาแยกตามประเภทของปัญหา
- จำนวนรวมทั้งหมดของจำนวนเพศที่แจ้งปัญหาเข้ามาตามเขตที่เลือก
- ตัวเลือกเขตที่ต้องการดู
- กราฟแสดงจำนวนเพศจากทั้งหมดเขตที่เลือกโดยคิดเป็นเปอร์เซ็น

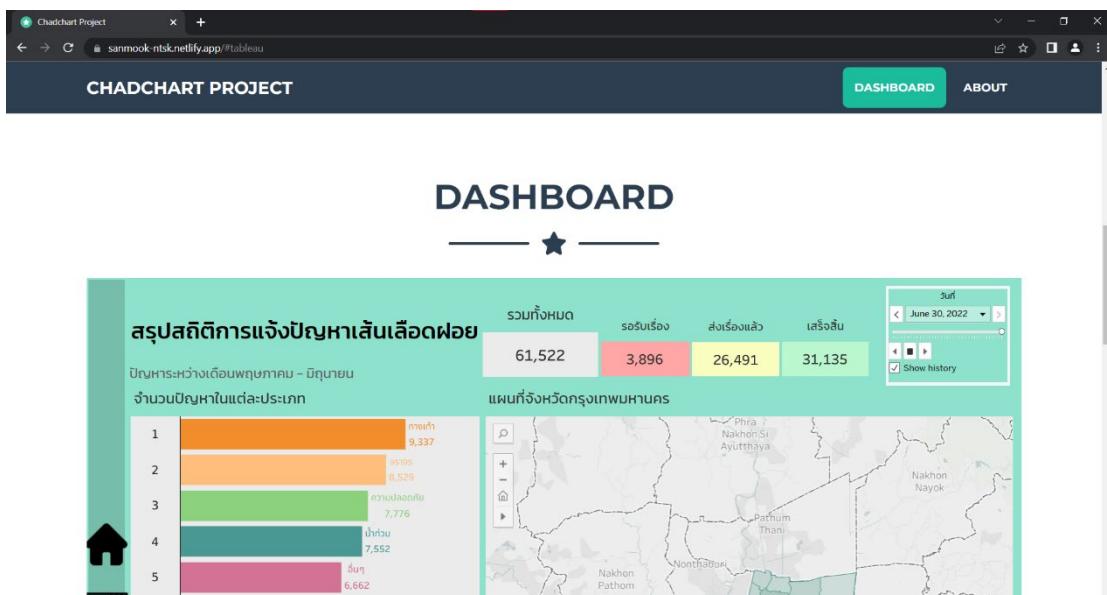
สืบเนื่องจาก Dashboard 3 มีการแสดงจำนวนปัญหาที่แจ้งเข้ามาโดยแบ่งตามเพศทั้งหมด Dashboard 4 นี้จะแสดงจำนวนเพศที่แจ้งปัญหาเข้ามาทั้งหมดเช่นกัน เพียงแต่แบ่งตามเขต ซึ่งตัวกราฟจะแบ่งเป็น 2 ประเภท 1 ด้านซ้ายของภาพที่ 4.3 เป็นกราฟรูปวงกลม และ 2 ด้านขวาของภาพที่ 4.3 เป็นกราฟแบบแท่ง เช่นเดียวกับ Dashboard 3 ด้วยการแบ่งตามเขตที่สามารถเลือกเองได้ ทำให้สามารถวิเคราะห์ข้อมูลเชิงลึกได้มากขึ้น ข้อมูลเชิงลึกในหน่วยอยู่ได้มากขึ้น

4.1.5 Website <https://sanmook-ntsk.netlify.app>

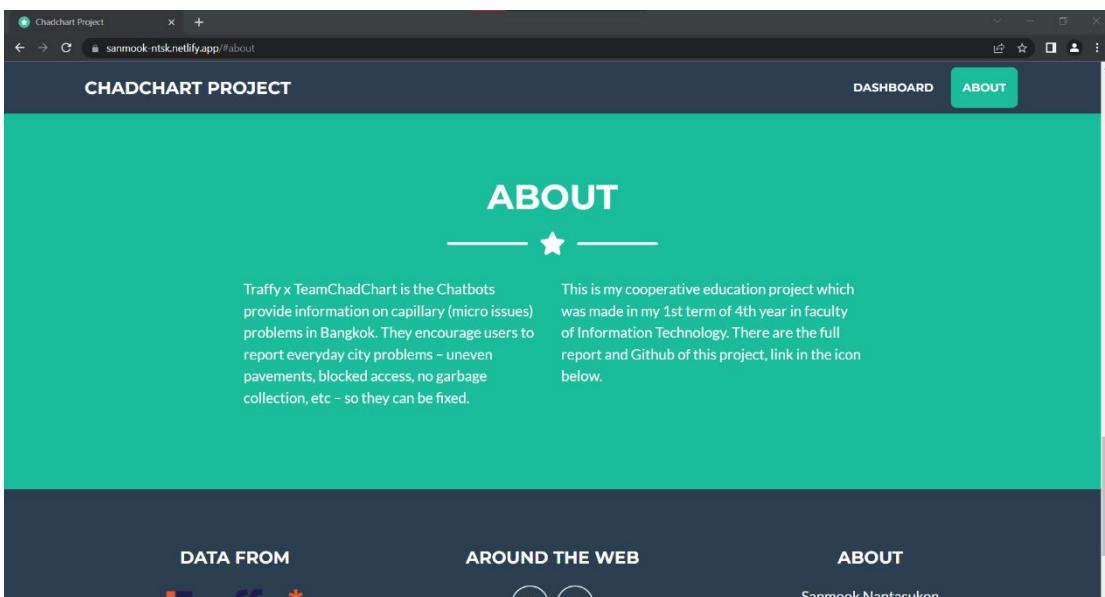
โดย Dashboard ทั้ง 4 จะถูกจัดแสดงไว้บน Website <https://sanmook-ntsk.netlify.app>



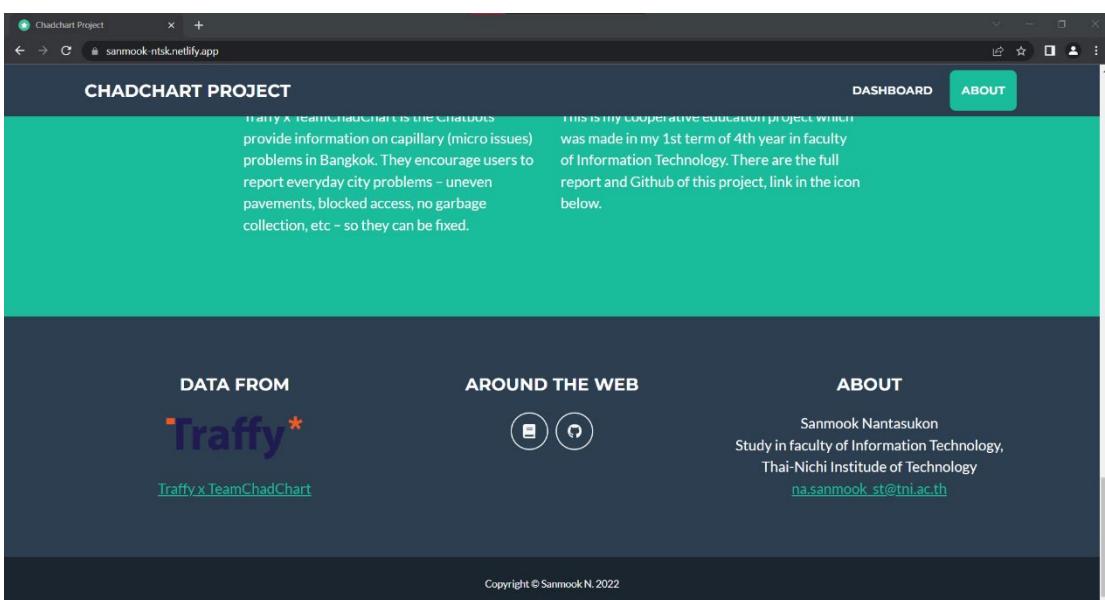
ภาพที่ 4.5 Website <https://sanmook-ntsk.netlify.app> หน้าแรก



ภาพที่ 4.6 Website <https://sanmook-ntsk.netlify.app> หน้า Dashboard



ภาพที่ 4.7 Website <https://sanmook-ntsk.netlify.app/#about> หน้า About



ภาพที่ 4.8 Website <https://sanmook-ntsk.netlify.app> ส่วน Footer

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 สรุปผลดำเนินโครงการ

จากการที่ข้าพเจ้าได้ออกปฏิบัติสหกิจศึกษา ณ บริษัท แอดวานซ์ อินฟอร์ เซอร์วิส จำกัด (มหาชน) ในตำแหน่ง Business Data Analyst ซึ่งมีรายละเอียดงานตัวอย่างเช่น ทำความสะอาดข้อมูล เตรียมพร้อมข้อมูลก่อนการทำการวิเคราะห์ข้อมูล สร้างแผนภูมิหรือกราฟ รวมไปถึงสร้าง Dashboard สำหรับการนำเสนอ ทำให้ข้าพเจ้าได้รับความรู้ และประสบการณ์เพิ่มเติมที่สามารถนำไปประยุกต์ใช้กับการทำโครงการที่ข้อมูลเชิงลึกนี้ได้

จากการลงมือวิเคราะห์ข้อมูลเชิงลึก โดยใช้ข้อมูลรายงานปัญหาเส้นเลือดฟ้อยในกรุงเทพมหานคร ซึ่งเป็นข้อมูลที่เปิดให้ประชาชนแจ้งปัญหาขนาดอยู่ต่าง ๆ ในกรุงเทพฯ เพื่อส่งเรื่องให้หน่วยงานที่เกี่ยวข้องแก้ไข โดยใช้ข้อมูลในช่วงเดือนพฤษภาคม - เดือนมิถุนายน ปีพ.ศ. 2565 รวมทั้งสิ้น 61,522 เรื่อง เดิมพบว่า ข้อมูลประเภท “อื่น ๆ” มีจำนวนมากผิดปกติ จึงได้ทำการแก้ปัญหาโดยการใช้ Natural Language Processing (NLP) และ WordCloud เข้ามาจัดการ ทำให้สามารถแก้ไขปัญหาและวิเคราะห์ข้อมูลต่อได้ จากการวิเคราะห์พบว่า ปัญหาที่ถูกแจ้งเข้ามามากที่สุดคือ ปัญหาประเภท “ทางเท้า”, เขตที่มีการแจ้งปัญหามากที่สุดคือเขต “จตุจักร”, เพศหญิงมีการแจ้งปัญหาประเภท “ความปลอดภัย” มากกว่าเพศชาย และเขต “บางรัก” มีปัญหาสถานะ “รอรับเรื่อง” มากที่สุด

ท้ายสุดแล้ว จากการดำเนินงานดังกล่าว นับเป็นประสบการณ์ที่ข้าพเจ้าได้ทำการวิเคราะห์ข้อมูลเชิงลึก รวมไปถึงข้าพเจ้ายังได้รับความรู้และทักษะต่าง ๆ อาทิ ความรู้ด้านการทำ Data Cleaning ความรู้ด้านการทำ Visualization ทักษะการสื่อสาร ทักษะการนำเสนอ ทักษะการทำงานด้านเอกสาร และอื่น ๆ อันเป็นความรู้และทักษะที่จะกล่าวเป็นพื้นฐานอันสำคัญ ซึ่งจำเป็นต่อการเข้าปฏิบัติงานในภายภาคหน้า

5.2 ปัญหาและแนวทางในการแก้ไข

5.2.1 Response (HTTP Response) 500

ปัญหานี้เกิดขึ้นระหว่างการดึงข้อมูลผ่าน API ตั้งเหตุมาจากการที่ Website ต้นทางมีระบบรักษาความปลอดภัย คือ การตัดการติดต่อสื่อสารชั่วคราวกับช่องทางสื่อสารนั้น ๆ หากช่องทางนั้น ๆ มีการติดต่อสื่อสารเข้ามาที่ทาง Website ต้นทางซึ่งมากเกินไป เพื่อเป็นการป้องกัน Website ล่มเนื่องจากมีการติดต่อเข้ามาใช้งานเป็นจำนวนมาก

ซึ่งการดึงข้อมูลผ่าน API ในที่นี้ ต้องติดต่อไปที่ตัว Website ต้นทางซึ่งอยู่แล้วเพื่อให้ได้ข้อมูลครบ ทำให้ตัว Website ส่ง Response 500 (มีบางอย่างผิดพลาดเกิดขึ้นบน Website) เพื่อหยุดการติดต่อสื่อสาร ทำให้ข้อมูลบางส่วนขาดหาย

ทางแก้คือการเขียน Code ตรวจสอบการตอบกลับจาก Website ต้นทางที่จะดึงข้อมูลมา ว่ามีการส่ง Response 201 (การเชื่อมต่อปกติ) กลับมาหรือไม่ ถ้าการเชื่อมต่อปกติ แสดงว่าสามารถดึงข้อมูลมาได้ แต่ถ้าหากไม่ปกติให้ส่งคำขอร้องเชื่อมต่อไปใหม่ จนกว่าต้นทางจะยอมรับ เพื่อข้อมูลที่ต้องการจะได้ไม่ขาดหาย

```
<Response [201]>
25000
<Response [201]>
26000
<Response [201]>
27000
<Response [500]>
28000
<Response [201]>
28000
<Response [201]>
29000
<Response [500]>
30000
<Response [201]>
30000
<Response [201]>
31000
<Response [201]>
32000
```

ภาพที่ 5.1 ตัวอย่าง HTTP Response 500 ที่ได้รับ

5.2.2 Font path บน Databrick

เนื่องจากการใช้ Wordcloud ต้องใส่ Font Path ด้วย โดยเฉพาะภาษาไทย ที่หากไม่ใส่ก็จะไม่สามารถแสดงตัวอักษรภาษาไทยได้ ดังนั้นแล้วจึงจำเป็นที่จะต้องใส่ Font Path และเพื่อความสวยงามด้วยส่วนนึง ซึ่งถ้าหากใช้งาน Wordcloud บนเครื่องคอมพิวเตอร์ ส่วนตัว จะไม่เกิดปัญหาดังกล่าว เพราะสามารถใส่ Font Path เป็นที่อยู่ของ Font File ที่โหลดมาเองได้ แต่ทว่า Databrick เป็น Cloud Service ไม่สามารถอัปโหลด Font File ส่วนตัวขึ้นไปได้

ทางแก้จึงจำเป็นต้องใช้ Library fontTools เพื่อดึง Font Path สาธารณะจาก Google มาแทน เป็นอันใช้ได้

```
1 pip install fontTools
Python interpreter will be restarted.
Collecting fontTools
  Downloading fonttools-4.36.0-py3-none-any.whl (950 kB)
Installing collected packages: fontTools
Successfully installed fontTools-4.36.0
Python interpreter will be restarted.

Command took 9.92 seconds -- by sanmooks@ais.co.th at 8/23/2022, 9:39:42 AM on TRAINEE_DA
```

```
Cmd 23
1 pip install Brotli
Python interpreter will be restarted.
Collecting Brotli
  Downloading Brotli-1.0.9-cp38-cp38-manylinux1_x86_64.whl (357 kB)
Installing collected packages: Brotli
Successfully installed Brotli-1.0.9
Python interpreter will be restarted.

Command took 5.74 seconds -- by sanmooks@ais.co.th at 8/23/2022, 9:39:43 AM on TRAINEE_DA
```

```
Cmd 24
1 startfontfile = "./sarabun.woff2"
2 urllib.request.urlretrieve(url="https://fonts.gstatic.com/s/sarabun/v13/DtVhJx26TKEr37c9YHZ5iXwJ1gk.woff2",
3                             filename=startfontfile)
4 endfontfile = "./sarabun.ttf"
5
6 fontTools.ttLib.woff2.decompress(input_file=startfontfile,
7                                   output_file=endfontfile)

Command took 0.12 seconds -- by sanmooks@ais.co.th at 8/23/2022, 9:41:24 AM on TRAINEE_DA
```

ภาพที่ 5.2 การติดตั้งและใช้งาน fontTools

5.3 ข้อเสนอแนะจากการดำเนินงาน

จากการดำเนินโครงการสหกิจศึกษา มีข้อเสนอแนะดังนี้

1. Traffy Founde ควรมีการจัดเก็บข้อมูลช่อง “Status” หรือช่องสถานะปัญหาที่แจ้งใหม่ จากเดิมที่เมื่อตัวปัญหาที่ถูกแจ้งเข้ามานั้น เมื่อถูกเปลี่ยนไปอยู่สถานะใด ข้อมูลก็จะถูก Replace หรือเขียนทับทันที ทำให้ไม่สามารถทราบได้ว่า ปัญหานั้น ๆ ถูกทิ้งไว้กี่วันก่อนจะมีการเปลี่ยนสถานะ หรือแก้ไข ถ้าหากมีการปรับปรุงส่วนนี้ ก็จะสามารถวิเคราะห์เวลาการทำงานได้มากขึ้น
2. ควรใช้ Pyspark ในการจัดการกับข้อมูลขนาดใหญ่ (Big Data) แทน Pandas หรือ SQL เนื่องจาก Pyspark ถูกออกแบบมาเพื่อจัดการกับข้อมูลขนาดใหญ่โดยเฉพาะ ทำให้สามารถเรียกแสดงผลข้อมูลที่ต้องการได้รวดเร็วกว่า Pandas หรือ SQL

เอกสารอ้างอิง

- [1] AIS, 2558, **About AIS** [Online], Available : <https://www.ais.th/aboutais/index.html> [8 กันยายน 2565].
- [2] Riccoprint, 2559, **รับทำความสะอาดข้อมูล Data cleaning - Riccoprint** [Online], Available : <https://riccoprint.com/data-cleaning-service> [10 กันยายน 2565].
- [3] Ball Weera Kasetsin, 2562, **Data Pipeline Explanation** [Online], Available : <https://medium.com/linedevth/data-pipeline-explanation-d3a2d7f21700> [10 กันยายน 2565].
- [4] Sirasit Boonklang, 2564, **Web Scraping ด้วย Node.JS** [Online], Available : <https://www.borntodev.com/2021/11/07/web-scraping-%E0%B8%94%E0%B9%89%E0%B8%A7%E0%B8%A2-node-js> [14 กันยายน 2565].
- [5] Saksit Srimarong, 2564, **Visual Analytics การวิเคราะห์ด้วยภาพ** [Online], Available : <https://bigdata.go.th/big-data-101/visual-analytics> [14 กันยายน 2565].
- [6] Meow, 2564, **Dashboard คืออะไร พร้อมประโยชน์และตัวอย่าง** [Online], Available : <https://blog.datath.com/dashboard-data-analysis> [14 กันยายน 2565].
- [7] Talil, 2564, **API หรือ Application Programming Interface คืออะไร ?** [Online], Available : <https://tips.thaiware.com/1728.html> [15 กันยายน 2565].
- [8] Echo, 2564, **Argument Python คืออะไร ใช้ทำอะไร** [Online], Available : <https://www.devkit.com/post/1023> [15 กันยายน 2565].
- [9] GoogleDeveloper, 2564, **Sankey Diagram | Charts - Google Developers** [Online], Available : <https://developers.google.com/chart/interactive/docs/gallery/sankey> [26 กันยายน 2565].
- [10] Kraisak, 2564, **การประมวลผลภาษาธรรมชาติ** [Online], Available : <https://csit.nu.ac.th/kraisak/ds/ds/chapter10/Chapter10.pdf> [29 กันยายน 2565].

ภาคผนวก

รายงานประจำสัปดาห์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัชคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 1

ชื่อ-สกุลนักศึกษา แสงนฤทธิ์ พันธุ์สุคันธ์ รหัสนักศึกษา..... 19.1.31.1.0084
 คณะวิชา..... ภาคโนร์ดอีสต์สันเตอร์ สาขาวิชา..... เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน/...../.....				
อังคาก/...../.....				
พุธ 1./.6./65	8	ภาษา Data Visualization tools ต่างๆ	รู้จัก Data Visualization tools ex. Power BI, Tableau	ไม่เคยใช้งาน Data Visualization tools ก่อน
พฤหัสบดี 2./.6./65	8	- ฝึกใช้ Tableau เสร็จข้อมูลของมีเป็นไฟล์ชั้งต้น - ซึ่งมีไฟล์ clean data public transport	รู้จักเครื่องมือ tools เช่น Tableau	Tableau นี้ tools ทางประมวลผลและประยุกต์ 90% ไม่ได้แลกเปลี่ยน
ศุกร์/...../.....				
เสาร์/...../.....				
อาทิตย์/...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	16	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	—	ลงชื่อ..... แสงนฤทธิ์ พันธุ์สุคันธ์ (..... แสงนฤทธิ์ พันธุ์สุคันธ์) วัน/เดือน/ปี 6/6/65 นักศึกษา	ลงชื่อ..... อรุณรัตน์ โภคสันติ์ (..... อรุณรัตน์ โภคสันติ์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 6/6/65 ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	16			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมเม้นท์ ทุกสัปดาห์หรือครึ่งครึ่ง อย่างน้อย 1 ครั้ง/เดือน ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร : 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 2

ชื่อ-สกุลนักศึกษา แสงฤทธิ์ พันธุ์สุดารงค์ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอัตโนมัติ	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน 6/./65	8	ฝึกการใช้งาน calculated fields เป็นต้นใน Tableau	รู้จักการใช้งาน calculated fields เป็นต้น	ผู้รับสอนใน calculated fields ไม่มากพอ
อัจฉริยะ 7/./65	8	เข้าอบรมเกี่ยวกับ Data Introduction ของบริษัท	รู้จัก Data Architecture และ Data Domain เป็นต้น ฝึกใช้งาน	—
พื้นฐาน 8/./65	8	- ฝึกใช้ function ก้าแต่ละอย่าง เช่นใน Tableau - ฝึกใช้ export ภาพเป็น powerpoint	รู้จักการใช้งานของ เป็นต้นใน Tableau และ export	ยังรู้จักการนำไปสู่การพัฒนาต่อไปไม่รู้จะต่อไปอย่างไร
พฤหัสบดี 9/./65	8	เข้าฟังบรรยายเรื่องภาพรวมธุรกิจ ของบริษัท	เรียนรู้ภาพรวมธุรกิจ ของบริษัท	—
ศุกร์ 10/./65	8	ฝึกสร้างกราฟ แบบเส้นเพื่อ highlight ค่า Min, Max, Med ต้องการใช้ calculated fields	เรียนรู้การใช้งาน calculated fields เพิ่มขึ้น	ผู้รับสอนคำสั่งใน calculated fields ไม่มากพอ
เสาร์ /.....				
อาทิตย์ /.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	16	ลงชื่อ แสงฤทธิ์ พันธุ์สุดารงค์ (..... แสงฤทธิ์ พันธุ์สุดารงค์) วัน/เดือน/ปี 13/6/65 นักศึกษา	ลงชื่อ วิชิตา โภคสุพันธุ์ (..... วิชิตา โภคสุพันธุ์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 13/6/65 ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	56			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้ร่วมงานทุกคนทุกวิชา ทุกสัปดาห์หรือถ่วงครวด อย่างน้อยถ้าไม่เกิน ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ ๓

ชื่อ-สกุณักศึกษา แสงนุก นันดร์สุดาร์ รหัสนักศึกษา..... 1913110084
 คณะวิชา..... เทคโนโลยีสารสนเทศ สาขาวิชา..... เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน ๑๓/๖/๖๕	๘	ภาษา python กับ SQL	ทักษะการใช้ python, SQL	-
อังค์ ๑๔/๖/๖๕	๘	เข้าร่วมฝึกอบรมเรื่องการทำ Visualization	ได้รู้จักการทำกราฟทางมาเขียน	-
พุธ ๑๕/๖/๖๕	๘	ศึกษา pyspark เพื่อใช้ query ข้อมูลจาก database	ทักษะการใช้ pyspark	เขียนรู้จัก function ใน pyspark ไม่มากพอ
พฤหัสบดี ๑๖/๖/๖๕	๘	ซ้อมใช้ query ข้อมูลจาก database	ทักษะการใช้ pyspark และ SQL	รู้จักคำสั่งjoin pyspark ไม่มากพอ
ศุกร์ ๑๗/๖/๖๕	๘	ฝึกทำขั้นตอนที่ query บน anaconda นำรับเข้ามาใน tableau	ทักษะการใช้ tableau	-
เสาร์ /.....				
อาทิตย์ /.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	56	ลงชื่อ..... แสงนุก นันดร์สุดาร์ (..... แสงนุก นันดร์สุดาร์) วัน/เดือน/ปี..... 24/6/65	ลงชื่อ..... อรุณรัตน์ โภคสุขพันธุ์ (..... อรุณรัตน์ โภคสุขพันธุ์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี..... 24/6/65	
จำนวนชั่วโมง รวมทั้งหมด	96	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หนัทวิชา ทุกสัปดาห์หรือถ่วงครั้ด อย่างน้อยถ้ามีเวลาเที่ยงวันไว้เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 4

ชื่อ-สกุณักศึกษา แสงนุก พันธุ์สุคุณ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นที่ 20/6/65	8	ศึกษาวิธีการดึงข้อมูลจาก Website	ฝึกวิธี 1. ตั้งตาก API 2. Web Scraping	-
อังค์ 21/6/65	8	เรียนรู้วิธีการทำ Web Scraping เป็งตันตัวง Beautiful Soup จากผู้	Web Scraping เป็งตัน	-
พุธ 22/6/65	8	ฝึกทำ Web Scraping เป็งตัน	Web Scraping เป็งตันเพิ่มเติม	สรุปชุด function 9 用Beautiful Soup ไม่สามารถ
พฤหัสบดี 23/6/65	8	เรียนรู้วิธี clean data ภาคผ่า	Clean data เป็งตัน	-
ศุกร์ 24/6/65	8	เรียนรู้วิธี clean data กด code ของผ่า	Clean data เป็งตัน เพิ่มเติม	สรุปชุด function 9 用Pandas ไม่สามารถ
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	96	ลงชื่อ แสงนุก พันธุ์สุคุณ (..... แสงนุก พันธุ์สุคุณ) วัน/เดือน/ปี 24/6/65	ลงชื่อ อริติภา ใจสุกันต์ (..... อริติภา ใจสุกันต์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 29/6/65	
จำนวนชั่วโมง รวมทั้งหมด	136	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หน่วยวิชา ทุกสัปดาห์หรือถ่วงครั้งครวด อย่างน้อยถ้ามีเวลาเที่ยงวัน ให้เพื่อทำการงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัจฉริยา สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 5

ชื่อ-สกุลนักศึกษา แสงอุไร นันทน์ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นที่ 27/6/65	8	หนทางการใช้ SQL	ทักษะ SQL	-
อังคาร 28/6/65	8	เรียนรู้การใช้ SQL ในการ Query data ทดลองใช้งานฝึกหัด	ทักษะ SQL	-
พุธ 29/6/65	8	เรียนรู้การใช้ SQL ในการ Query data ปрактиcin ทักษะฝึกหัด	ทักษะ SQL	ไม่รู้จักข้อมูลที่มีอยู่มากนัก
พฤหัสบดี 30/6/65	8	ติดตั้ง PySpark เพื่อฝึกการใช้งาน	Pyspark เพื่องาน	-
ศุกร์ 1/7/65	8	ติดตั้ง PySpark เพื่อฝึกการใช้งาน	Pyspark เพื่องาน	รู้จัก function ใน PySpark ไม่很多
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	136	ลงชื่อ แสงอุไร นันทน์ (..... แสงอุไร นันทน์) วัน/เดือน/ปี 8/7/65 นักศึกษา	ลงชื่อ อรุณรัตน์ ใจดี (..... อรุณรัตน์ ใจดี) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 8/7/65 ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	176			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมpetence ทุกสับดาห์หรือครึ่งครึ่ง อย่างน้อยครั้งเดือนหนึ่ง ไว้เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่.....๖.....

ชื่อ-สกุณักศึกษาแสหนุกนนทกรสุคุณ รหัสนักศึกษา.....1913110084
 คณะวิชา.....ภาคโนริชาระสัน美德 สาขาวิชา.....เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันท์ ๔./๗./๖๕	๘	ตึกษา Pandas สำหรับการ clean data เป็นต้น	Pandas เป็นต้น	-
อังคาร ๕./๗./๖๕	๘	clean data ด้วยภาษา Pandas	Pandas, clean data เป็นต้น	-
พุธ ๖./๗./๖๕	๘	clean data ด้วยภาษา Pandas	Pandas, clean data เป็นต้น	ต้อง data ที่ต้อง clean ต้องใช้รีซัปเปน และ = ล็อก explode ต้อง
พฤหัสบดี ๗./๗./๖๕	๘	clean data ด้วยภาษา Pandas	Pandas, clean data เป็นต้น	-
ศุกร์ ๘./๗./๖๕	๘	ตึกษาวิธีการลงข้อมูลจาก Website ตัว API	ต้อง data ผ่าน API เป็นต้น	บาง website มีการบันลิง API ผ่อนให้งาน ท่อน website บางแห่งไม่เข้า
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	176	ลงชื่อ..... <u>แสหนุก นนทกรสุคุณ</u> (..... <u>แสหนุก นนทกรสุคุณ</u>) วัน/เดือน/ปี..... <u>๙/๗/๖๕</u>	ลงชื่อ..... <u>ธีรดา ໂຄສະນາ</u> (..... <u>ธีรดา ໂຄສະນາ</u>) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี..... <u>๙/๗/๖๕</u>	
จำนวนชั่วโมง รวมทั้งหมด	216	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมเม้นท์ ทุกสัปดาห์อ่อนคร่าวด อย่าลืมเขียนลงในแบบฟอร์มที่แนบมาให้ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 7

ชื่อ-สกุลนักศึกษา แสงฤทธิ์ พันธุ์สุคันธ์ รหัสนักศึกษา.....1913110084.....
 คณะวิชา.....เทคโนโลยีสารสนเทศ สาขาวิชา.....เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอัตโนมัติ	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
เดือนที่ 11/7/65	8	ติดตั้ง API ดังข้างต้นจาก Website ผ่าน API	ตั้ง Up ฝ่าย API เบื้องต้น	-
เดือนที่ 12/7/65	9	ติดตั้งข้อมูลจาก Website ผ่าน API	ตั้ง Data ผ่าน API เบื้องต้น	Parameters มีการจำกัด ผลลัพธ์การดังนี้หากต้อง ^{ส่ง} ก็เป็น
พุธ...../...../.....				
พฤหัสบดี 14/7/65	8	ติดตั้งข้อมูลจาก Website ผ่าน API	ตั้ง Up ฝ่าย API เบื้องต้น	# Requests limit per minute ทำให้ต้องรอรับ ^{ส่ง} response 500
ศุกร์ 15/7/65	8	ติดตั้งข้อมูลจาก Website ผ่าน API	ตั้ง Data ผ่าน API เบื้องต้น	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	32	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	216	ลงชื่อ..... แสงฤทธิ์ พันธุ์สุคันธ์ (..... แสงฤทธิ์ พันธุ์สุคันธ์) วัน/เดือน/ปี..... 26/7/65	ลงชื่อ..... ธิตา ໂຄສ່ວນ (..... ธิตา ໂຄສ່ວນ) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี..... 26/7/65	
จำนวนชั่วโมง รวมทั้งหมด	248	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้ดูแลนักศึกษา ทุกสัปดาห์หรือถ่วงครึ่ง อย่างน้อยครั้งละหนึ่งใบ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ ๘

ชื่อ-สกุลนักศึกษา แสงสุก นันดรศุภวงศ์ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอิสระ	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันพ. ๑๔./๗./๖๕	๘	clean data จากข้อมูลที่ได้รับ	clean data เป็นต้น เพิ่มเติม	-
อังค์. ๑๙./๗./๖๕	๘	clean data จากข้อมูลที่ได้รับ	clean data เป็นต้น เพิ่มเติม	-
พุ. ๒๐./๗./๖๕	๘	ศึกษาวิธีกรอก text preprocessing	กรอก text cleaning	-
พฤหัสบดี ๒๑./๗./๖๕	๘	ศึกษาวิธีกรอก text preprocessing	กรอก word tokenize	-
ศุกร์ ๒๒./๗./๖๕	๘	ศึกษาวิธีกรอก text preprocessing	การ count words และการคำนวณ TF IDF	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	248	ลงชื่อ..... แสงสุก นันดรศุภวงศ์ (..... แสงสุก นันดรศุภวงศ์)	ลงชื่อ..... ธิติพงษ์ วงศ์สุวรรณ (..... ธิติพงษ์ วงศ์สุวรรณ)	ตำแหน่ง Business Data Intelligence Analyst Manager
จำนวนชั่วโมง รวมทั้งหมด	288	วัน/เดือน/ปี ๒๖/๗/๖๕ นักศึกษา	วัน/เดือน/ปี ๒๔/๗/๖๕ ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้ดูแลนักศึกษา ทุกสัปดาห์หรือถ่วงครึ่ง อย่างน้อยครั้งละหนึ่งใบ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัจฉริภาพ สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร : 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ ๗

ชื่อ-สกุลนักศึกษา แสงนุก พันธุ์ศุภนธ์ รหัสนักศึกษา 191310084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน ๒๕/๗/๖๕	-	ลา กิจ	-	-
อังคาก ๒๖/๗/๖๕	8	text preprocessing เมืองทัน	ทำ word tokenize เมืองทัน	ตัด text ฝึกภาษา word tokenize ตัดออก ทำงานตาม
พุทธศักราช ๒๗/๗/๖๕	8	text preprocessing เมืองทัน	ทำ word tokenize เมืองทัน	-
พฤหัสบดี/...../.....				
ศุกร์/...../.....				
เสาร์/...../.....				
อาทิตย์/...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	16	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	238	ลงชื่อ แสงนุก พันธุ์ศุภนธ์ (..... แสงนุก พันธุ์ศุภนธ์) วัน/เดือน/ปี 26/7/65	ลงชื่อ ติติมา ໂຄສ່ານິຕີ (..... ติติมา ໂຄສ່ານິຕີ) ตำแหน่ง Business, Data, Intelligence, Analyst, Manager วัน/เดือน/ปี 26/7/65	ผู้ควบคุมการปฏิบัติงาน
จำนวนชั่วโมง รวมทั้งหมด	304	นักศึกษา		

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้ร่วมงานทุกคนจะเป็น ทุกสัปดาห์หรือถ่องคร่ำครั้ด อย่างน้อย 1 ครั้ง ไม่เกิน 1 เดือน ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 10

ชื่อ-สกุลนักศึกษา แสงนฤทธิ์ พันธุ์สุคันธ์ รหัสนักศึกษา..... 1913110084
 คณะวิชา..... เทคโนโลยีสารสนเทศ สาขาวิชา..... เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอัลกอริทึม	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นที่ 1./๘./๖๕	8	ตัดช่วงตัดคำ text mining	ทำ text mining เพื่อแบ่งกลุ่มหัวสังเคราะห์	-
ขั้นที่ 2./๘./๖๕	8	ตัดช่วงตัดคำ text mining	text mining เป็นต้นหัวสังเคราะห์ word cloud	-
ขั้นที่ 3./๘./๖๕	8	ตัดช่วงตัดคำ text mining	Py Thai NLP	-
พฤหัสบดี 4./๘./๖๕	8	ตัดช่วงตัดคำ word cloud	library wordcloud	-
ศุกร์ 5./๘./๖๕	8	ตัดช่วงตัดคำ word cloud	library PyThaiNLP	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	304	ลงชื่อ..... แสงนฤทธิ์ พันธุ์สุคันธ์ (..... แสงนฤทธิ์ พันธุ์สุคันธ์.....) วัน/เดือน/ปี..... 22/๘/๖๕..... นักศึกษา	ลงชื่อ..... Thitiporn Phanthong (..... ทิติพน พันธุ์สุคันธ์.....) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี..... 22/๙/๖๕..... ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	344			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้รับผิดชอบวิชา ทุกสัปดาห์หรือถ่วงครั้ด อย่างน้อยครั้งละหนึ่งใบ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร : 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือนที่

เดือนที่ 11

ชื่อ-สกุลนักศึกษา แสงนุก พันธุ์สุกนร รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน ๑๕/๑๕/๖๕	8	word cloud python จาก data ที่ tokenize แล้ว	Deep learning	import font file ขึ้น databrick ไม่ได้
อังค์ ๗/๗/๖๕	8	word cloud python จาก data ที่ tokenize แล้ว	Deep learning	-
พุธ. ๑๘/๙/๖๕	8	word cloud python จาก data ที่ tokenize แล้ว	Deep learning	word cloud ไม่ลงผลลัพธ์ ลงมาหนาเกินไป
พฤหัสบดี. ๒๓/๙/๖๕	8	word cloud python จาก data ที่ tokenize แล้ว	Deep learning	-
ศุกร์ /...../.....				
เสาร์ /...../.....				
อาทิตย์ /...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	32	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	344	ลงชื่อ แสงนุก พันธุ์สุกนร (..... แสงนุก พันธุ์สุกนร) วัน/เดือน/ปี 22/๙/๖๕ นักศึกษา	ลงชื่อ T.WATKIN.C.J (..... อรุณรัตน์ วงศ์สุกนร) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 22/๙/๖๕ ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	378			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้ร่วมงานทุกคนจะวิชา ทุกสัปดาห์หรือถ่วงครั้ด อย่างน้อย 1 ครั้ง ให้เข้าสู่ระบบเท่านั้น ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 12

ชื่อ-สกุลนักศึกษา แสงนุ๊ก พัฒนาศุภวงศ์ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอิสระ	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขั้นพื้นฐาน 15/๘/๖๕	๘	Word cloud python จาก data ที่ tokenize แล้ว	Deep learning	-
อังคาร 16/๘/๖๕	๙	Word cloud python จาก data ที่ tokenize แล้ว	Deep learning	-
พุธ 17/๘/๖๕	๙	ศึกษาเรื่องการจัดเก็บข้อมูลด้วย pandas	pandas	-
พฤหัสบดี 18/๘/๖๕	๙	ศึกษาเรื่องการจัดเก็บข้อมูลด้วย pandas	pandas	-
ศุกร์ 19/๘/๖๕	๙	ศึกษาเรื่องการจัดเก็บข้อมูลด้วย pandas	pandas	-
เสาร์ /...../.....				
อาทิตย์ /...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	378	ลงชื่อ แสงนุ๊ก พัฒนาศุภวงศ์ (..... แสงนุ๊ก พัฒนาศุภวงศ์) วัน/เดือน/ปี 22/๙/๖๕	ลงชื่อ Thitima J. (..... อริสา โภคสุวรรณ) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 22/๙/๖๕	
จำนวนชั่วโมง รวมทั้งหมด	418	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หน่วยงาน ทุกสัปดาห์หรือครึ่งครึ่ง อย่างน้อย 1 ครั้ง ไว้เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัจฉริภาพ สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่.....13.....

ชื่อ-สกุลนักศึกษา ไชยา พันธุ์สุวนันท์ รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันทร์ 22/๘/๖๕	๘	ฝึกวิธีการ จัดการข้อมูลด้วย pandas	pandas	-
อังคาร 23/๘/๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาจัดการแบบ群 ร่วมกัน	กรอก customer segments	ข้อมูลไม่สม่ำเสมอ ทำให้เปลี่ยนกลุ่มไม่ได้
พุธ 24/๘/๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาจัดการ แบบ群ร่วมกัน	กรอก customer segments	-
พฤหัสบดี 25/๘/๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาจัดการ แบบ群ร่วมกัน	กรอก customer segments	-
ศุกร์ 26/๘/๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาจัดการ แบบ群ร่วมกัน	กรอก customer segments	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	418	ลงชื่อ..... ไชยา พันธุ์สุวนันท์ (..... ไชยา พันธุ์สุวนันท์) วัน/เดือน/ปี..... 22/๘/๖๕	ลงชื่อ..... T.M.I.M.A.D (..... ไชยา พันธุ์สุวนันท์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี..... 24/๘/๖๕	
จำนวนชั่วโมง รวมทั้งหมด	458	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หน้า ทุกสัปดาห์หรือถ่องคร่ำครัด อย่างน้อยถ้าไม่เกิน 1 วัน เพื่อทำการงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพัฒนาการ แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 14

ชื่อ-สกุณักศึกษา แสงฤทธิ์ พนกนันต์ รหัสนักศึกษา 1913110084
 คณะวิชา ภาควิชาระดับบัณฑิต สาขาวิชา หลักสูตรสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันท์ ๒๑./๘./๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาตัด แบ่งกลุ่มข้อมูล	กราฟ customer segments	-
อังคาร ๓๐./๘./๖๕	๘	นำข้อมูลที่ได้จาก word cloud มาตัด แบ่งกลุ่มข้อมูล	กราฟ customer segments	-
พุธ ๓๑./๘./๖๕	๘	ศึกษาการวิเคราะห์ข้อมูล เพื่อนำไปใช้งาน	ໄใจเด็งสำนึกรากฐาน insight	-
พฤหัสบดี ๑./๙./๖๕	๘	ศึกษาการวิเคราะห์ข้อมูล เพื่อนำไปใช้งาน	ໄใจเด็งสำนึกรากฐาน insight	-
ศุกร์ ๒./๙./๖๕	๘	ศึกษา กรณีศenario นำข้อมูล เพื่อนำไปใช้งาน	ໄใจเด็งสำนึกรากฐาน insight	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	458	ลงชื่อ แสงฤทธิ์ พนกนันต์ (..... แสงฤทธิ์ พนกนันต์) วัน/เดือน/ปี 1/9/65	ลงชื่อ สิรินา ใจสุกนันต์ (..... สิรินา ใจสุกนันต์) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 1/9/65	
จำนวนชั่วโมง รวมทั้งหมด	498	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หน่วยงาน ทุกสัปดาห์หรือถ่วงครั้ด อย่างน้อยถ้าไม่น้อยกว่า ๔ ไฟล์ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ 15

ชื่อ-สกุลนักศึกษา ไสว พานิช รหัสนักศึกษา 1913110084
 คณะวิชา สถาปัตยกรรมศาสตร์ สาขาวิชา สถาปัตยกรรมศาสตร์

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอิสระ	ความรู้ทักษะที่ได้รับ	บัญหา/อุปสรรค
ขันพร. ๕./๙./๖๕	๘	ศึกษาคร่าวๆเเผนน้ำที่บ้านมาเพื่อนำไปใช้	โน๊ตสีเข้มน้ำเงิน Insight	—
อังคาร ๖./๙./๖๕	๘	ลงฝังวิเคราะห์ข้อมูล	กราฟเคราะห์ข้อมูล	—
พุธ ๗./๙./๖๕	๘	ลงฝังวิเคราะห์ข้อมูล	กราฟเคราะห์ข้อมูล	—
พฤหัสบดี ๘./๙./๖๕	๘	ลงฝังวิเคราะห์ข้อมูล	กราฟเเพร์เซ็นต์	—
ศุกร์ ๙./๙./๖๕	๘	ลงฝังวิเคราะห์ข้อมูล	กราฟ insight	—
เสาร์/...../.....				
อาทิตย์/...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	498	ลงชื่อ ไสว พานิช (..... ไสว พานิช) วัน/เดือน/ปี 26/๙/๖๕ นักศึกษา	ลงชื่อ ดิษณุ ใจมาน (..... ดิษณุ ใจมาน) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 26/๙/๖๕ ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	538			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมpetence ทุกสับ派ห่อถ่องครั้งครัด อย่างถูกต้องตามที่กำหนดไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตสวนหลวง กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 ต่อ 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่ ๑๖

ชื่อ-สกุลนักศึกษา [เสียงุก นันดร์สุกานดา] รหัสนักศึกษา [๑๓๑๐๐๔๔]
 คณะวิชา [เทคโนโลยีสารสนเทศ] สาขาวิชา [เทคโนโลยีสารสนเทศ]

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันพ. ๑๒./๙./๖๕	๘	ลงมือวิเคราะห์ข้อมูล	ทราบ insight	-
อังค. ๑๓./๙./๖๕	๙	ลงมือวิเคราะห์ข้อมูล	ทราบ insight	-
พุ. ๑๔./๙./๖๕	๘	ลงมือวิเคราะห์ข้อมูล	ทราบ insight	-
พฤหัสบดี ๑๕./๙./๖๕	๘	ลงมือวิเคราะห์ข้อมูล	ทราบ insight	-
ศุกร. ๑๖./๙./๖๕	๘	ลงมือวิเคราะห์ข้อมูล	ทราบ insight	-
เสาร์...../...../.....				
อาทิตย์...../...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	๔๐	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	๕๓๘	ลงชื่อ [เสียงุก นันดร์สุกานดา] (..... [เสียงุก นันดร์สุกานดา]) วัน/เดือน/ปี ๒๔/๙/๖๕	ลงชื่อ [เสียงุก นันดร์สุกานดา] (..... [เสียงุก นันดร์สุกานดา]) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี ๒๔/๙/๖๕	
จำนวนชั่วโมง รวมทั้งหมด	๕๗๘	นักศึกษา	ผู้ควบคุมการปฏิบัติงาน	

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมเม้นท์ ทุกสับดาห์ อีเมลถ้าไม่ได้รับตอบกลับ ให้ทำการงานฉบับสมบูรณ์

Thanh P.

(..... ลงชื่อ โน้ตผู้สอน)

ตำแหน่ง Business Data Intelligence Analyst Manager

วัน/เดือน/ปี ๒๔/๙/๖๕



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือนที่

เดือนที่.....17.....

ชื่อ-สกุลนักศึกษา แสง พงษ์สุขุม รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอัล	ความรู้/ทักษะที่ได้รับ	บัญหา/อุปสรรค
ขั้นพื้นฐาน 19/9/65	๘	- ฝึกอบรม Visualizatior สรุป - ทำใบสรุป	การทำ Visualization	-
อังคาร 20/9/65	๙	- ฝึกอบรม Visualizatior สรุป - ทำใบสรุป	การทำ Visualization	-
พุธ 21/9/65	๘	- ทำ Visualization - ทำใบสรุป	การทำ Visualization	-
พฤหัสบดี 22/9/65	๘	- ทำ Visualization - ทำใบสรุป	การทำ Visualization	-
ศุกร์ 23/9/65	๘	- ทำ Visualization - ทำใบสรุป	การทำ Visualization	-
เสาร์/...../.....				
อาทิตย์/...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	578	ลงชื่อ แสง พงษ์สุขุม ^(..... แสง พงษ์สุขุม) วัน/เดือน/ปี 26/9/65 นักศึกษา	ลงชื่อ รังษิมา พ. ^(..... รังษิมา พ.) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 26/9/65 ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	618			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้สอนทุก周หน่วยงาน ทุกสัปดาห์หรือครึ่งครึ่ง อย่างน้อยครั้งละหน้าเล็กน้อย ไว้ เพื่อทำรายงานฉบับสมบูรณ์



CCC-Co 04

ศูนย์สหกิจศึกษาและอัคคหางาน สถาบันเทคโนโลยีไทย - ญี่ปุ่น

Co-operative Education and Career Center

1771/1 ถนนพหลโยธิน แขวงสวนหลวง เขตคลองเตย กรุงเทพฯ 10250 โทรศัพท์: 0-2763-2700 โทร 2750, 2762 โทรสาร: 0-2763-2754

แบบฟอร์มรายงานการปฏิบัติงานประจำเดือน

เดือนที่.....18.....

ชื่อ-สกุลนักศึกษา แสงนุก นันทน์สุดา รหัสนักศึกษา 1913110084
 คณะวิชา เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ

วัน/เดือน/ปี	จำนวนชั่วโมง	งานที่ปฏิบัติโดยอื่น	ความรู้/ทักษะที่ได้รับ	ปัญหา/อุปสรรค
ขันท์ 26/9/65	9	- ฝึกหัดใช้igrab ทำ Dashboard - ทำเดสก์	กราฟ Dashboard	—
อังคาร 27/9/65	8	- ฝึกหัดใช้igrab ทำ Dashboard - ทำเดสก์	กราฟ Dashboard	—
พุธ 28/9/65	8	- อบรม Visualization ทำ Dashboard - ทำเดสก์	กราฟ Dashboard	—
พฤหัสบดี 29/9/65	8	- อบรม Visualization ทำ Dashboard - ทำเดสก์	กราฟ Dashboard	—
ศุกร์ 30/9/65	8	- อบรม Visualization ทำ Dashboard - ทำเดสก์	กราฟ Dashboard	—
เสาร์/...../.....				
อาทิตย์/...../.....				
จำนวนชั่วโมงรวม ในรายงานฉบับนี้	40	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	ขอรับรองว่ารายงานฉบับนี้ เป็นความจริงทุกประการ	
จำนวนชั่วโมง ในรายงานฉบับก่อน	618	ลงชื่อ แสงนุก นันทน์สุดา (..... แสงนุก นันทน์สุดา) วัน/เดือน/ปี 26/9/65 นักศึกษา	ลงชื่อ อธิษฐาน โภคสุธรรม (..... อธิษฐาน โภคสุธรรม) ตำแหน่ง Business Data Intelligence Analyst Manager วัน/เดือน/ปี 26/9/65 ผู้ควบคุมการปฏิบัติงาน	
จำนวนชั่วโมง รวมทั้งหมด	658			

หมายเหตุ นักศึกษาต้องส่งรายงานฉบับนี้ถึงอาจารย์ที่ปรึกษาสาขาวิชา/ผู้งานทุกคอมเม้นท์ ทุกสับดาห์อ่อนครั้งครั้ด อย่างน้อย 2 ครั้ง ไว้เพื่อทำรายงานฉบับสมบูรณ์

ประวัติผู้จัดทำโครงการ

ชื่อ – สกุล	นางสาวแสตนมุก นันตะสุคนธ์	
วัน เดือน ปีเกิด	10 พฤษภาคม 2544	
ประวัติการศึกษา		
ระดับประถมศึกษา	ประถมศึกษาตอนปลายพ.ศ. 2553	
	โรงเรียนพระธรรมหนักสวนกุหลาบ มหาสารคาม	
ระดับมัธยมศึกษา	มัธยมศึกษาตอนต้นพ.ศ. 2556 โรงเรียนอัมพรไพรีศาล มัธยมศึกษาตอนปลายพ.ศ. 2559 โรงเรียนโพธิสารพิทยากร	
ระดับอุดมศึกษา	คณะเทคโนโลยีสารสนเทศ สาขาวิชาเทคโนโลยีสารสนเทศพ.ศ. 2562 สถาบันเทคโนโลยีไทย - ญี่ปุ่น	
ทุนการศึกษา	ทุนสถาบันเทคโนโลยีไทย-ญี่ปุ่น ประเภทที่ 4	
ประวัติการฝึกอบรม	1. หลักการ MONOZUKURI ณ สถาบันเทคโนโลยีไทย – ญี่ปุ่น 2. อบรมหลักการ KAIZEN ณ สถาบันเทคโนโลยีไทย – ญี่ปุ่น 3. Project Management ผ่านช่องทางออนไลน์	
ผลงานที่ได้รับการตีพิมพ์	- ไม่มี -	