

รายงานการเข้าร่วมโครงการ

19-IN-54-GE-TRC-A

Training of Trainers on Big Data Analytics for Public Sector Productivity

ระหว่างวันที่ 25 - 29 พฤศจิกายน 2562

ณ กรุงเทพมหานคร ประเทศไทย

จัดทำโดย นายพิชญุตม์ ทองจันทร์

ตำแหน่ง รักษาการผู้ช่วยผู้อำนวยการฝ่าย ฝ่ายบริหารข้อมูลลูกค้า ธนาคารออมสิน

วันที่ 7 มกราคม 2563

ที่มาของโครงการ

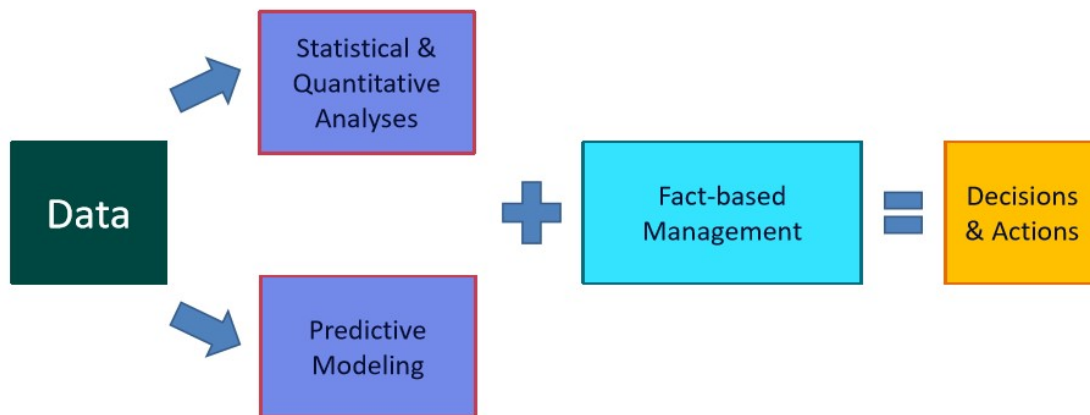
โครงการ Training of Trainers on Big Data Analytics for Public Sector Productivity จัดขึ้นที่กรุงเทพมหานคร ระหว่างวันที่ 25 - 29 พฤศจิกายน 2562 โดย องค์การเพิ่มผลผลิตแห่งเอเชีย (Asian Productivity Organization) หรือ APO และ สถาบันเพิ่มผลผลิตแห่งชาติ เป็นหนึ่งในโครงการชุดฝึกอบรม มีวัตถุประสงค์เพื่อเพิ่มศักยภาพและขีดความสามารถให้กับหน่วยงานและองค์กรสาธารณะภายใต้ประเทศสมาชิกขององค์การเพิ่มผลผลิตแห่งเอเชีย

โครงการนี้เป็นโครงการฝึกอบรมเชิงปฏิบัติการ 5 วัน โดยวิทยากรผู้เชี่ยวชาญด้านการวิเคราะห์ข้อมูล 1 ท่าน จาก ประเทศสิงคโปร์ ได้แก่ Prof. Michelle CHEONG Lee Fong โดยเนื้อหาประกอบด้วย

- วันที่ 1 และวันที่ 2 พูดถึงภาพรวมในการฝึกอบรม การวิเคราะห์ข้อมูลและความสำคัญในการตั้งคำถามเพื่อวิเคราะห์ข้อมูลที่ถูกตั้งโดยใช้ Microsoft Excel ในการคำนวณข้อมูล พร้อมทำแบบฝึกหัด
- วันที่ 3 พูดถึงการทำความเข้าใจเกี่ยวกับการวิเคราะห์แบบ Descriptive และ Visualization โดยใช้ QlikSense Program พร้อมทำแบบฝึกหัด
- วันที่ 4 พูดถึงการทำความเข้าใจเกี่ยวกับการวิเคราะห์แบบ Predictive โดยใช้ WEKA Program และการทำ Forecasting พร้อมทำแบบฝึกหัด
- วันที่ 5 พูดถึงการพัฒนาคิดวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลที่นำเสนอโดย DELTA Framework พร้อมทั้งการประเมินความสามารถในการวิเคราะห์ขององค์กร



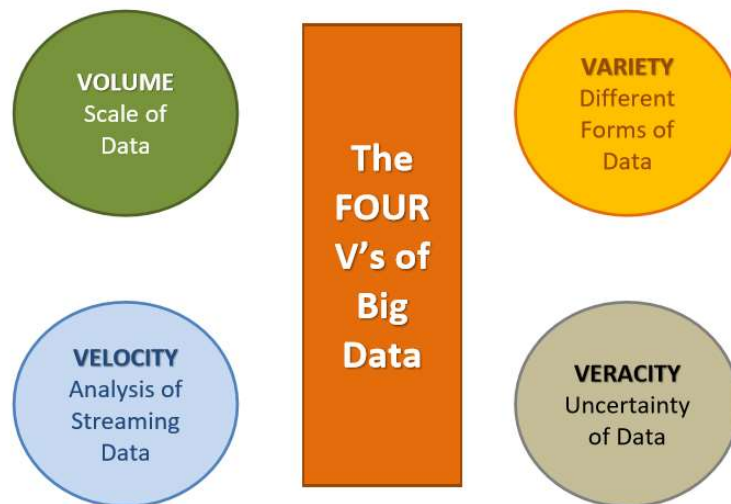
เนื้อหา/องค์ความรู้ที่ได้รับ



ภาพที่ 1 การวิเคราะห์ข้อมูล

เริ่มจากการเรียนรู้พื้นฐานเกี่ยวกับการวิเคราะห์ข้อมูล กำหนดว่าการวิเคราะห์ข้อมูลคืออะไร ซึ่งเป็นการใช้ข้อมูล การวิเคราะห์เชิงสถิติและเชิงปริมาณ มาประกอบการวิเคราะห์โดยมีการฝึกตั้งโจทย์เพื่อวิเคราะห์ข้อมูลที่ถูกต้อง ฝึกตั้งคำถาม เพื่อตอบโจทย์และนำผลการวิเคราะห์ข้อมูลไปสนับสนุนการตัดสินใจในการดำเนินธุรกิจ

โดย Big Data มีคุณลักษณะสำคัญอยู่ 4 อย่าง คือ



ภาพที่ 2 คุณลักษณะของ Big Data

1. Volume คือ ข้อมูลที่มีปริมาณมาก ในระดับ Terabyte หรือ Petabyte ขึ้นไป
2. Velocity คือ ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาและรวดเร็ว หรืออาจจะหมายถึงความเร็วของการสร้างข้อมูลและประมวลผลข้อมูล

3. Variety คือ ชนิดและลักษณะของข้อมูลมีความหลากหลาย ทั้งแบบมีโครงสร้าง แบบกึ่งโครงสร้าง หรือแบบไม่มีโครงสร้าง เช่น รูปภาพ ข้อความ วิดีโอ
4. Veracity คือ คุณภาพและความน่าเชื่อถือของข้อมูล เพราะถ้าข้อมูลไม่มีคุณภาพ ไม่มีความน่าเชื่อถือ ซึ่งมีความเกี่ยวเนื่องโดยตรงกับผลลัพธ์การวิเคราะห์ข้อมูลที่มีประสิทธิภาพ

ซึ่งในการจัดการหรือการวิเคราะห์ข้อมูลเชิงลึกจะถูกแบ่งเป็น 4 ระดับ ได้แก่



ภาพที่ 3 ระดับการวิเคราะห์ข้อมูลเชิงลึก ที่มา : ดร.ธนชาตย์ ฤทธิบูรณ์ ผู้เชี่ยวชาญด้าน Data Analytics

1. Descriptive Analytics

ลำดับขั้นแรกของการวิเคราะห์ข้อมูล จะเริ่มต้นจากการวิเคราะห์แบบพื้นฐาน โดยวิเคราะห์เพียงแค่ตัวแปรเดียวเท่านั้น (Univariate Analysis) ทำการวิเคราะห์ด้วยเทคนิคประเภทสถิติเชิงพรรณนา หาค่ากลาง ค่าการกระจาย เช่น ค่าเฉลี่ย ผลรวม เป็นต้น

ตัวอย่างคำถามการวิเคราะห์ในระดับนี้ จะเป็นเพียงคำถามเพื่อให้ทราบว่าเกิดอะไรขึ้น เช่น ยอดขายเป็นเท่าไร จำนวนลูกค้าทั้งหมดเป็นเท่าไร ความถี่ในการใช้บริการบ่อยแค่ไหน เป็นต้น

2. Diagnostics Analytics

ในลำดับขั้นถัดไปของการวิเคราะห์นั้น จะเป็นการหาความสัมพันธ์ของตัวแปรตั้งแต่สองตัวขึ้นไป (Multivariate Analysis) เช่น การวิเคราะห์สหสัมพันธ์ (Correlation Analysis) หรือ การวิเคราะห์ข้อมูลเชิงหลายมิติ (Multidimensional Analysis) การวิเคราะห์ข้อมูลในขั้นตอนนี้จะเป็นการวิเคราะห์เพื่อประเมินว่าตัวแปรแต่ละตัวนั้นมีผลต่อกันหรือไม่

ตัวอย่างคำถามการวิเคราะห์ในระดับนี้ เช่น อุณหภูมิมีผลต่อยอดขายหรือไม่ ยอดขายในแต่ละพื้นที่แตกต่างกันหรือไม่ เป็นต้น

3. Predictive Analytics

ระดับการวิเคราะห์ในขั้นนี้ จะมุ่งเน้นการพยากรณ์หรือทำนายอนาคต เป็นการสร้างโมเดลเพื่อสกัดรูปแบบความสัมพันธ์ที่สามารถนำไปใช้กับเหตุการณ์ที่ยังไม่เกิดขึ้นได้ เทคนิคการวิเคราะห์จะเป็นกลุ่มเทคนิคการทำเหมืองข้อมูล (Data Mining) หรือ การเรียนรู้ของเครื่องจักร (Machine Learning)

ตัวอย่างคำถามการวิเคราะห์จะเป็นคำถามเชิงพยากรณ์ เช่น ลูกค้าคนไหนจะหยุดใช้บริการ เมื่อลูกค้าซื้อสินค้าแล้ว จะซื้อสินค้าใดต่อ เป็นต้น

4. Prescriptive Analytics

ลำดับขั้นที่ถูกพิจารณาว่าเป็น Maturity Level ที่สูงสุดของการทำ Data Analytics ก็คือ การทำ Prescriptive Analytics หรือ การวิเคราะห์ที่สามารถบอกได้ว่าต้องทำอะไรเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดภายใต้เงื่อนไขที่จำกัด โดยเทคนิคที่ใช้ ก็คือ Optimization

ตัวอย่างคำถามการวิเคราะห์ เช่น ควรตั้งราคาสินค้าเท่าไรดี ถึงจะได้กำไรสูงสุด หรือ จะต้องวางแผนการส่งสินค้าในเวลาใด หรือ เส้นทางใดจึงจะใช้ต้นทุนต่ำที่สุด

ฟังก์ชัน Microsoft Excel พื้นฐาน ที่นำมาใช้ในการวิเคราะห์ข้อมูล มีดังนี้

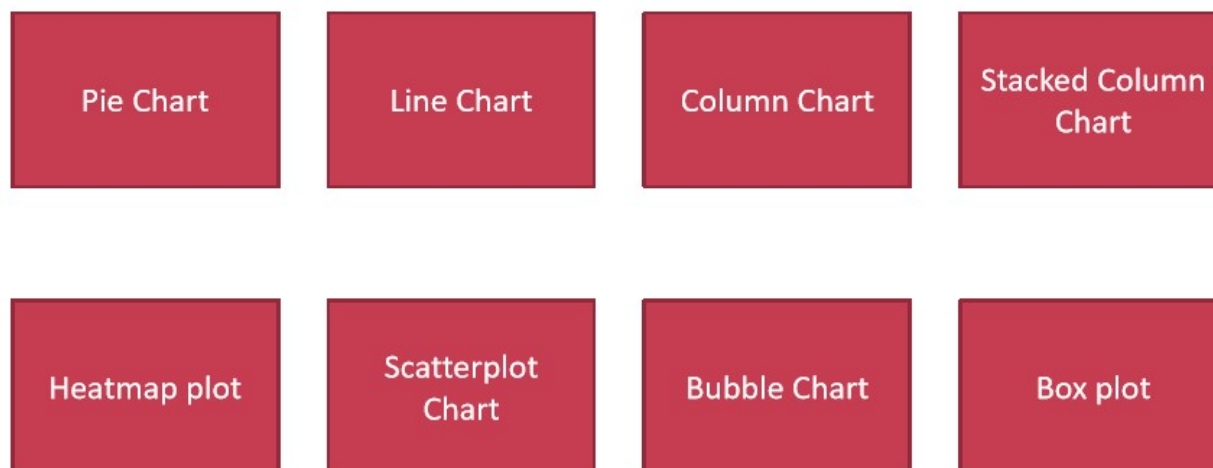
1. Excel Basic statistical information – COUNT(), AVERAGE(), MAX(), MIN(), COUNTIF(), COUNTIFS(), SUMIF(), AVERAGEIF()
2. Data filtering – AUTOFILTERS, Advanced Filters o Data Lookup – LOOKUP(), VLOOKUP(), MATCH(), INDEX()
3. Univariate Data Analysis – Central Tendency (Mean, Median, Mode), Dispersion (SD, Variance)
4. Bivariate Data Analysis – Causality (Linear relationship), RSQ(), TREND(), COVARIANCE(), CORREL()
5. Linear Regression – LINEST()
6. Graphs and Charts – Column, Bar, Pie, Area, XY Scatter
7. Curve Fitting – XY Scatter with Trendline
8. Pivot Table and Pivot Chart

การวิเคราะห์แบบ Descriptive และ Visualization

การเตรียมข้อมูลเพื่อการวิเคราะห์หลักๆจะมีข้อมูลอยู่ 2 ประเภท คือ

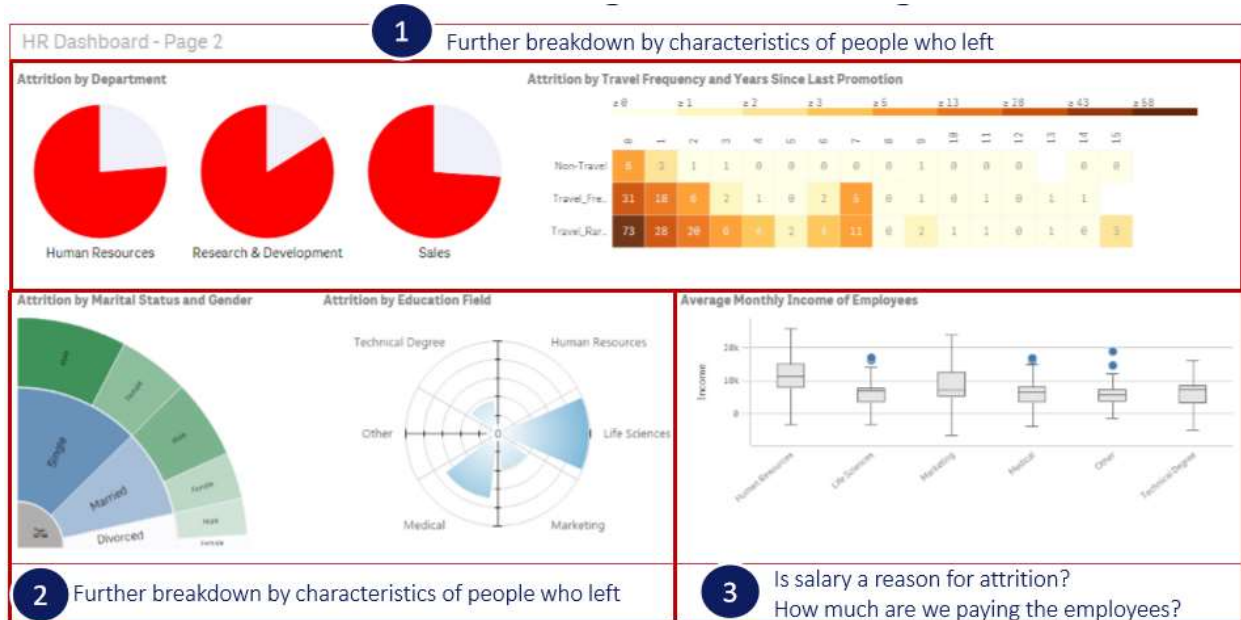
1. Dimensions จะเป็นข้อมูลประเภทตัวอักษร เช่น เพศ ชื่อผลิตภัณฑ์ ที่อยู่ ปี เป็นต้น และ
2. Measures จะเป็นข้อมูลประเภทตัวเลข เช่น ยอดขาย กำไร ต้นทุน เป็นต้น

ซึ่งข้อมูลทั้ง 2 ประเภทสามารถนำมาวิเคราะห์ได้ตามที่ต้องการ โดยการวิเคราะห์แบบ Descriptive ที่ต้องการให้มีการแสดงผล Visualization สามารถใช้ QlikSense Program เป็นเครื่องมือช่วยในการวิเคราะห์ข้อมูลให้ดูง่ายขึ้นในรูปแบบของกราฟต่างๆ



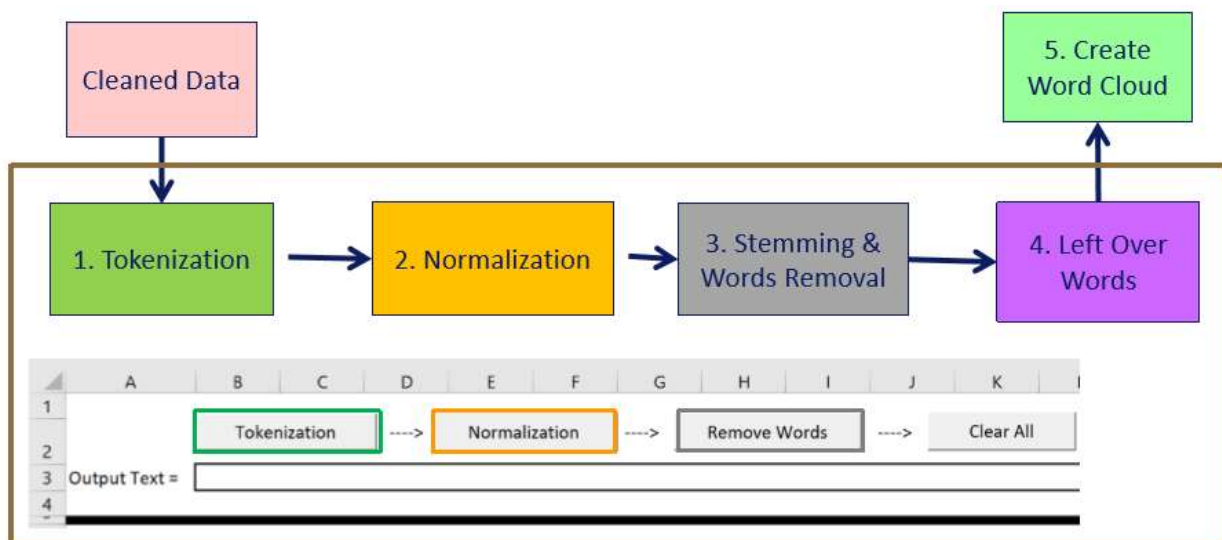
ภาพที่ 4 กราฟแบบต่างๆ

จะเห็นว่า QlikSense Program มีกราฟแบบต่างๆหลากหลายรูปแบบให้เลือกใช้ตามความเหมาะสมของข้อมูลแต่ละประเภท โดยสามารถนำเข้าไฟล์ข้อมูลที่เตรียมไว้เข้าสู่ QlikSense Program และ คลิกเลือกกราฟที่ต้องการพร้อมเลือกข้อมูลที่ต้องการจะวิเคราะห์ ก็จะได้ผลการวิเคราะห์ดังภาพ



ภาพที่ 5 ตัวอย่างหน้าจอผลการวิเคราะห์ข้อมูลโดยใช้ QlikSense Program

ถ้าต้องการทราบจำนวนของคำพูด หรือประโยคที่ถูกกล่าวถึงมากที่สุด จะมีวิธีในการดูได้อย่างง่ายผ่านการทำ word cloud ซึ่งก่อนที่จะทำ word cloud ได้นั้นจะต้องทำการ cleaned data ตามขั้นตอนดังนี้



ภาพที่ 6 ขั้นตอนการเตรียมข้อมูลเพื่อทำ word cloud

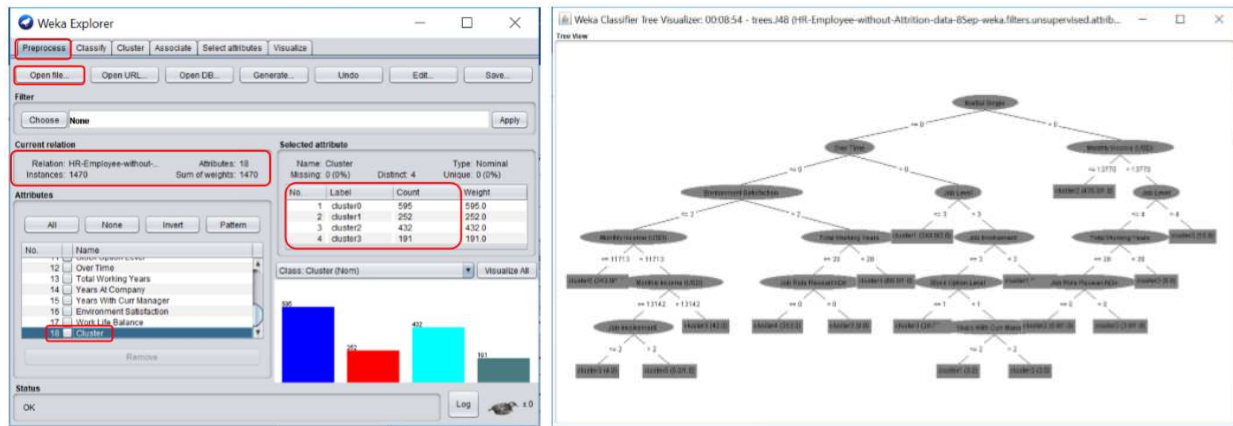


การ Forecasting จะสามารถช่วยให้เราทราบเทรนด์ที่จะเกิดขึ้นในอนาคตเพื่อที่จะได้เตรียมการรับมือกับสิ่งที่จะเกิดขึ้น เช่น การ Forecasting ยอดขายในอีก 3 เดือนข้างหน้าสามารถทำได้โดยดูจากยอดขายในเดือนที่ผ่านมา เป็นต้น

ภาพที่ 8 สรุปหลักการในการทำ Forecasting

1. เลือกวิธีการพยากรณ์ที่จะใช้โดยทำการพล็อตข้อมูลที่ผ่านมา ก่อน เพื่อดูว่าเป็นไปตามแนวโน้มหรือฤดูกาลต่างๆที่ผ่านมาหรือไม่
2. การเลือกวิธีที่เหมาะสม สามารถกำหนดค่าพารามิเตอร์ปรับให้เหมาะสม
3. เปรียบเทียบวิธีการที่จะเลือกใช้ โดยใช้การดูข้อมูลเชิงปริมาณ

การวิเคราะห์ข้อมูลแบบ Predictive สามารถวิเคราะห์ผ่าน WEKA Program โดยการนำเข้าข้อมูลเข้าสู่ระบบแล้ว และสามารถเลือก Attributes ที่ต้องการวิเคราะห์ พร้อมทั้งเลือกโมเดล Linear Regression Clustering & Decision Tree ที่ต้องการวิเคราะห์ข้อมูลก็จะแสดงผลได้ดังภาพ



ภาพที่ 9 การวิเคราะห์แบบ Predictive โดยใช้ WEKA Program

DELTA Framework

เป็นกรอบการทำงานที่ผู้บริหารสามารถใช้ในการประเมินองค์ประกอบที่สามารถทำให้ การวิเคราะห์ในการทำงานได้ตรงจุดและยังประสบความสำเร็จในการสร้างมูลค่าจากการวิเคราะห์



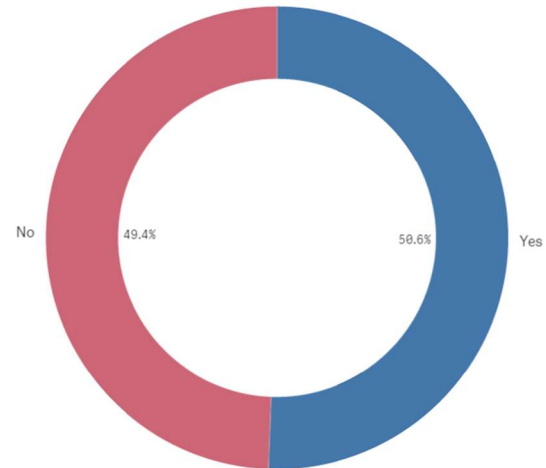
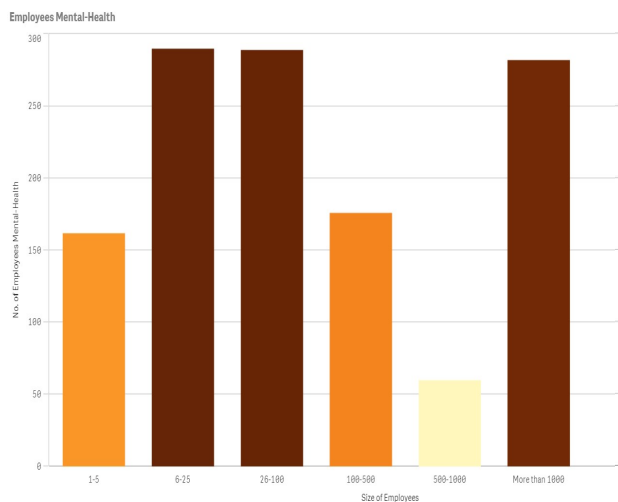
ภาพที่ 10 DELTA Framework

DELTA Framework ประกอบไปด้วยดังนี้

1. **Data** คือ ข้อเท็จจริงต่าง ๆ ที่มีอยู่ในธรรมชาติ เป็นกลุ่มสัญลักษณ์ แทนปริมาณหรือการกระทำต่าง ๆ ที่ยังไม่ผ่านการประมวลผล ข้อมูลอาจอยู่ในรูปของตัวเลข ตัวหนังสือ และท้ายสุดของข้อมูลก็คือ วัตถุประสงค์ของสารสนเทศ
2. **Enterprise** คือ การบริหารจัดการองค์กรให้มีประสิทธิภาพ
3. **Leadership** คือ ความสามารถด้านอิทธิพลต่อบุคคลในกลุ่มเพื่อนำไปสู่ความสำเร็จตามเป้าหมาย ใช้กระบวนการสั่งการ การมีอิทธิพลต่อผู้อื่น การมีปฏิสัมพันธ์ โดยถ่ายทอดแนวคิดไปสู่การปฏิบัติ
4. **Targets** คือ ระดับหรือมาตรฐานผลการปฏิบัติงานที่คาดหวัง องค์กรกำหนด เป้าหมายเป็นหลักสำหรับเปรียบเทียบ เพื่อวัดความก้าวหน้าของความสำเร็จขององค์กร
5. **Analysts** คือ นักวิเคราะห์ ที่ทำหน้าที่เหมือนเป็นนักแก้ปัญหาให้องค์กร โดยผ่านการวิเคราะห์ของข้อมูล

ภาพการเข้าร่วมกิจกรรมกลุ่ม (Group Discussion)

ทางกลุ่มได้เลือก Data Set ในหัวข้อผู้ป่วยที่เป็นโรคซึมเศร้ามาทำการวิเคราะห์เพื่อดูว่าปัจจัยอะไรบ้างที่เกี่ยวข้องและส่งผลทำให้เป็นโรคซึมเศร้า



ภาพที่ 11 ภาพการเข้าร่วมกิจกรรมกลุ่ม

ซึ่งจากการวิเคราะห์ข้อมูล ดังภาพพบว่า คนที่ป่วยเป็นโรคซึมเศร้าจะอยู่ในองค์กรที่มีขนาดเล็ก และขนาดใหญ่ โดยได้เข้ารับการรักษาย่างจริงจังที่ 50.6% ซึ่งจะเป็นได้ว่าจากผลการวิเคราะห์ข้อมูลกลุ่มตัวอย่าง จะยังมีผู้ป่วยโรคซึมเศร้าที่ยังไม่ได้เข้ารับการรักษาที่ค่อนข้างสูง ส่วนองค์กรที่มีขนาดกลางจะส่งผลให้เกิดโรคซึมเศร้าในจำนวนที่น้อย

ประโยชน์ที่ได้รับจากการเข้าร่วมโครงการ

สามารถนำความรู้ที่ได้จากการอบรมมาประยุกต์ใช้ในการทำงาน โดยสามารถวิเคราะห์หาความสัมพันธ์ของตัวแปรแต่ละตัวจากฐานข้อมูล และนำข้อมูลมาวิเคราะห์ เพื่อประกอบการตัดสินใจในการสร้างโอกาสทางธุรกิจให้กับธนาคาร อีกทั้งการพัฒนาผลิตภัณฑ์ หรือนำเสนอผลิตภัณฑ์/บริการที่ตอบสนองความต้องการของลูกค้าในแต่ละ Segment ได้อย่างมีประสิทธิภาพ