บทที่2
แนวคิดทฤษฎีเอกสารที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลปริมาณการส่งออกข้าวหอมมะลิจากกรมการค้าต่างประเทศเพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับระบบ เพื่อใช้เป็นแนวทางการศึกษา ประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยวกับการพยากรณ์ข้อมูล (Forecasting data)

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.1.4 แนวคิดเกี่ยวกับแผงหน้าปัด (Dashboard)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล

2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3 ทฤษฎีเกี่ยวกับอนุกรมเวลา (Time Series)

2.2.4 ทฤษฎีเกี่ยวกับกระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization)

2.3 เครื่องมือในการออกแบบ และวิเคราะห์ข้อมูล

2.3.1 แบบจำลองอนุกรมเวลา (Time series analysis)

2.3.2 การแสดงผลแบบ (visualization)

2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)

2.3.4 โปรแกรม Tableau Public

2.3.5 โปรแกรม Minitab

2.3.6 โปรแกรม IBM SPSS

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

เป็นศาสตร์ในการวิเคราะห์และประมวลผลข้อมูลด้วยวิธีทางคณิตศาสตร์ สถิติ และการวิเคราะห์ข้อมูลที่มีอยู่ในปัจจุบัน อดีต และทำนายอนาคต เพื่อนำผลนั้นมาปรับใช้ให้เกิดประโยชน์ ความรู้เรื่องดังกล่าวสำคัญยิ่งขึ้น เมื่อเทคโนโลยีในปัจจุบันมีการเก็บข้อมูลมหาศาลในระดับ Big Data มากเกินกว่าคนทั่วไปจะสามารถใช้งานได้เต็มที่ Data Analytics จึงเหมาะที่จะเป็นเครื่องมือสำหรับธุรกิจ (Business Intelligence) ทั้งขนาดใหญ่และขนาดเล็ก

2.1.1.1 ความหมายการวิเคราะห์ข้อมูล (Data analytic)

Data Analytics นั้นเป็น Business Intelligence อย่างหนึ่ง ซึ่งเป็นศาสตร์ของการใช้ข้อมูลต่าง ๆ จากที่ต่าง ๆ มาร่วมวิเคราะห์รวมกัน เพื่อปรับปรุงธุรกิจ หรือการตลาด ซึ่ง Data Analytics นั้นเป็น Business Analytics อย่างหนึ่ง โดย Business Analytics เป็นวิธีการที่ใช้ข้อมูลเทคโนโลยีด้านสารสนเทศ เข้ามาทำการวิเคราะห์ข้อมูล เพื่อช่วยในการดำเนินธุรกิจ ซึ่ง Business Analytics แบ่งเป็น 3 แบบคือ

1) Descriptive Analytics เป็นรูปแบบการใช้ข้อมูลแบบพื้นฐานที่สุด โดยเน้นการอธิบายว่ากำลังเกิดขึ้น หรืออาจจะเกิดอะไรขึ้น สามารถอธิบายถึงสาเหตุการเกิดต่าง ๆ ได้ว่าทำไม ซึ่ง Descriptive Analytics ตัวอย่างคือรายงานธุรกิจ รายงานด้านการทำ Campaign หรือโฆษณา หรือรายงานผลดำเนินงานที่ผ่านมา เป็นข้อมูลพื้นฐานที่แสดงผลในการทำกิจกรรมต่าง ๆ ให้เราได้รับทราบ

2) Predictive Analytics เป็นรูปแบบการใช้ข้อมูลที่มีความซับซ้อนขึ้นมา โดยจะเป็นการ “พยากรณ์” หรือ “ทำนาย” สิ่งที่กำลังเกิดขึ้น โดยใช้ข้อมูลในอดีต ร่วมกับโมเดลทางคณิตศาสตร์ต่าง ๆ หรือร่วมกับการทำ Data Mining นอกจากนี้ Predictive Analytics ยังทำให้เราสามารถวิเคราะห์หาโอกาสและความเสี่ยงต่าง ๆ ที่จะเกิดขึ้นในอนาคตได้ด้วย เช่นการรู้เทรนด์ทาง

การตลาด การพยากรณ์ยอดการขายหรือการทำ Campaign ว่าจะมีคนร่วมเท่าไหร่ หรือมี ROI เท่าไหร่

3) Prescriptive Analytics เป็นรูปแบบการวิเคราะห์ข้อมูล ที่มีความซับซ้อนและยากที่สุด เพราะไม่เพียงพยากรณ์หรือทำนายว่าอะไรจะเกิดขึ้น แต่ยังให้คำแนะนำในทางเลือกต่าง ๆ และผลแต่ละทางเลือกว่าจะมี Pros & Cons อย่างไร โมเดลของ Prescriptive Analytics นั้นจะสามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามามากขึ้น และ Prescriptive Analytics นี้ยังเป็นการใช้ข้อมูลที่มากที่สุด และเกี่ยวพันกับเรื่อง Big Data เป็นอย่างมาก

2.1.2 แนวคิดเกี่ยวกับการพยากรณ์ข้อมูล (Forecasting data)

การพยากรณ์เป็นทั้งศาสตร์และศิลป์ในการทำนายเหตุการณ์ในอนาคต ซึ่งอาจนำหลาย ๆ วิธีมาใช้แล้วแต่สถาณการณ์ เช่น นำข้อมูลในอดีตมาพยากรณ์อนาคตโดยอาศัยหลักการทางคณิตศาสตร์เข้าช่วย หรือใช้ดุลยพินิจของผู้พยากรณ์เพียงอย่างเดียว หรืออาจใช้หลาย ๆ วิธีร่วมกันเพื่อให้การพยากรณ์มีความแม่นยำมากขึ้น

2.1.2.1 ความหมายของการพยากรณ์ข้อมูล (Forecasting data)

การพยากรณ์ หมายถึงการคาดการณ์ (Predict) เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่สนใจที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศ (Information) ประกอบการตัดสินใจซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดก่อนการวางแผน หรือการเตรียมการที่จะเริ่มกระทำการใด ๆ เพื่อความถูกต้องและแม่นยำในการตัดสินใจ ดังนั้นในการดำเนินธุรกิจภายใต้ความไม่แน่นอนและการเปลี่ยนแปลงที่รวดเร็วเช่นปัจจุบันจึงจำเป็นที่จะต้องทราบถึงความเป็นไปในอนาคต โดยอาศัยเทคนิคการพยากรณ์ต่าง ๆ เพื่อเพิ่มประสิทธิภาพในการดำเนินงานทางธุรกิจ สามารถตอบสนองต่อความต้องการของลูกค้า เพิ่มผลตอบแทนทางธุรกิจ ลดต้นทุนและความสูญเสียต่าง ๆ ที่อาจเกิดขึ้นเป็นต้น

2.1.2.2 การเลือกเทคนิคการพยากรณ์

1) ระยะของการพยากรณ์ ผู้วิเคราะห์จะต้องกำหนดว่าต้องการพยากรณ์ระยะสั้น ระยะปานกลาง หรือระยะยาว

2) ลักษณะของข้อมูล การเลือกวิธีการพยากรณ์ควรเลือกจากลักษณะหรือรูปแบบของข้อมูล ข้อมูลบางชุดอาจจะไม่ค่อยเปลี่ยนแปลงมากนัก บางชุดมีแต่ความผันแปรไม่แน่นอน หรืออาจมีปัจจัย หรือตัวแปรประเภทอื่น ๆ ที่มีอิทธิพลต่อข้อมูลที่ต้องการศึกษา ซึ่งอาจต้องใช้วิธีการวิเคราะห์ความถดถอย และสหสัมพันธ์

3) ค่าใช้จ่าย การพิจารณาเลือกเทคนิคการพยากรณ์จะต้องคำนึงถึงค่าใช้จ่ายด้วย เนื่องจากแต่ละวิธีจะต้องมีจำนวนข้อมูลที่ใช้วิเคราะห์แตกต่างกัน ถ้าใช้ข้อมูลมากจะยิ่งต้องเสียค่าใช้จ่ายมาก นอกจากนั้น ถ้าเลือกใช้การวิเคราะห์ความถดถอย และสหสัมพันธ์จะต้องหาข้อมูลของตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรที่จะพยากรณ์ จะต้องเสียค่าใช้จ่ายในการหาข้อมูลตัวแปรอิสระเหล่านั้น โดยเฉพาะอย่างยิ่งถ้าตัวแปรอิสระเป็นข้อมูลภายนอกองค์กร ก็จะยิ่งเสียค่าใช้จ่ายมากขึ้น

4) ความถูกต้อง สิ่งที่สำคัญที่สุดในการพิจารณาเลือกเทคนิคพยากรณ์คือ ความถูกต้องของค่าพยากรณ์ นั้นต้องการค่าคลาดเคลื่อนในการพยากรณ์ต่ำสุด

2.1.2.3 ช่วงเวลาของการพยากรณ์

สามารถแบ่งตามระยะเวลาของการพยากรณ์ได้ 3 ประเภทคือ

1) การพยากรณ์ในระยะสั้น เป็นการพยากรณ์เหตุการณ์ที่ไม่เกิน 1 ปี โดยทั่วไปมักจะอยู่ในช่วงไม่เกิน 3 เดือน เช่น การพยากรณ์การวางแผนจัดซื้อ การจัดตารางการทำงาน การมอบหมายงาน การพยากรณ์ยอดขาย และการพยากรณ์ระดับการผลิต

2) การพยากรณ์ระยะกลาง เป็นการพยากรณ์เหตุการณ์ที่อยู่ในช่วง 3 เดือน ถึง 3 ปี จะใช้มากในการพยากรณ์การวางแผนการขาย การวางแผนการผลิต การวางแผนด้านงบประมาณเงินสด และการวิเคราะห์การวางแผนการดำเนินงานต่าง ๆ

3) การพยากรณ์ระยะยาว เป็นการพยากรณ์เหตุการณ์ที่มากกว่า 3 ปีขึ้นไป มักใช้สำหรับการวางแผนผลิตภัณฑ์ใหม่ ค่าใช้จ่ายในการลงทุน การขยายทำเลที่ตั้ง และการวิจัยพัฒนา

2.1.2.4 เทคนิคการพยากรณ์ (Forecasting Techniques)

การพยากรณ์เป็นเรื่องการคาดคะเนในอนาคต ดังนั้นการพยากรณ์กับความจริงนั้นอาจจะเหมือนหรือไม่เหมือนกัน แต่การพยากรณ์ที่ดีควรจะต้องให้ใกล้เคียงกับความเป็นจริงมากที่สุด ดังนั้นเทคนิคและวิธีการพยากรณ์จึงเป็นเรื่องที่จะต้องทำความเข้าใจร่วมกัน เทคนิคของการพยากรณ์นั้นจะแบ่งออกเป็น 2 ประเภท ตามลักษณะวิธีการพยากรณ์ที่อาศัยข้อมูลในอดีตประกอบการสร้างรูปแบบในการพยากรณ์ได้ดังนี้ (1) เทคนิคการพยากรณ์แบบไม่เป็นทางการ (Informal Forecasting Technique) วิธีพยากรณ์แบบนี้จะใช้ประสบการณ์และดุลพินิจของผู้พยากรณ์ โดยผู้พยากรณ์จะเป็นผู้ที่มีความรู้ความเชี่ยวชาญในเรื่องนั้น ๆ อย่างดีจึงสามารถคาดคะเนสิ่งต่าง ๆ ที่คาดว่าจะเกิดขึ้นได้ (2) เทคนิคการพยากรณ์แบบเป็นทางการ (Formal Forecasting Technique) การพยากรณ์วิธีนี้จะต้องอาศัยข้อมูลมาสนับสนุนและใช้ความรู้ทางสถิติและคณิตศาสตร์เข้ามาช่วยในการวิเคราะห์ข้อมูล ดังนั้นผู้พยากรณ์จะต้องเข้าใจวิธีการและขั้นตอนในการนำข้อมูลมาวิเคราะห์ เทคนิควิธีการพยากรณ์แบบเป็นทางการนี้ยังแบ่งออกได้เป็น 2 จำพวก คือ การพยากรณ์เชิงคุณภาพ (Qualitative Methods) และการพยากรณ์เชิงปริมาณ (Quantitative Methods)

1) เทคนิคเชิงคุณภาพ (Qualitative Techniques) เทคนิคการพยากรณ์เชิงคุณภาพ เป็นเทคนิคที่อาศัยประสบการณ์ผู้พยากรณ์เป็นส่วนใหญ่ ซึ่งอาจจะไม่มีการใช้ข้อมูลในอดีต เนื่องจากไม่ได้มีการเก็บรวบรวมข้อมูลในอดีตไว้หรือมีแต่มีไม่พอเพียงต่อการนำมาสร้างรูปแบบในการพยากรณ์ ดังนั้นความถูกต้องของการพยากรณ์เชิงคุณภาพจะขึ้นอยู่กับประสบการณ์และความสามารถของผู้พยากรณ์เป็นหลัก ซึ่งได้แก่ ทัศนคติของผู้จัดการการพยากรณ์โดยกลุ่มผู้บริหาร (A Jury of Executive Opinion) พนักงานขาย ทำการพยากรณ์ (Sale forecast Estimate) สำรวจตลาด (Market Research) เทคนิคเดลฟี (Delphi Technique)

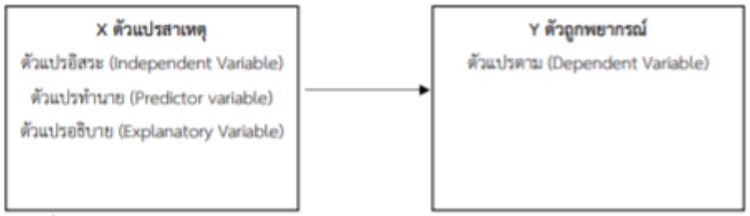
2) เทคนิคเชิงปริมาณ (Quantitative Techniques) เทคนิคการพยากรณ์เชิงปริมาณ จะเป็นเทคนิคที่ต้องใช้ข้อมูลในอดีตมาสร้างรูปแบบการพยากรณ์ในรูปของสมการคณิตศาสตร์ ดังนั้นความถูกต้องของการพยากรณ์นี้จะขึ้นอยู่กับความแม่นยำของข้อมูลที่มีอยู่และวิธีการในรูปแบบของสมการทางคณิตศาสตร์ ซึ่งแบ่งได้เป็น 2 ประเภทดังนี้

2.1. รูปแบบปัจจัยสาเหตุ หรือรูปแบบเชิงเหตุผล (Associative Models) เป็นการพยากรณ์ด้วยการวิเคราะห์ปัจจัยต่าง ๆ ที่จะมีผลกระทบต่อสิ่งที่จะพยากรณ์เช่น การวิเคราะห์การถดถอย (Regression Analysis) หรือ ตัวแบบเศรษฐมิติ (Econometric Model)

2.2. รูปแบบอนุกรมเวลา (Time Series Models) ซึ่งได้แก่ วิธีการหาค่าเฉลี่ยแบบตรงตัว (Naive Approach) วิธีหาค่าเฉลี่ยเคลื่อนที่ (Moving Average) วิธีการปรับเรียบแบบเอ็กซ์โปเนนเชียล (Exponential Smoothing) วิธีแยกส่วน (Classical Decomposition) และวิธีการคาดคะเนแนวโน้ม (Trend Projection)

2.1.4.5 การพยากรณ์ข้อมูลเชิงเหตุผล

เทคนิคการพยากรณ์ข้อมูลเชิงเหตุผลที่ได้รับความนิยม คือ การวิเคราะห์การถดถอย (Regression Model) ซึ่งเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) โดยมีวัตถุประสงค์เพื่อรักษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม และศึกษาปัจจัยหรือตัวแปรอิสระ ที่ร่วมกันทำนายหรือพยากรณ์ตัวแปรตาม การวิเคราะห์การถดถอยเพื่อหาความสัมพันธ์หรือสร้างสมการทำนายหรือพยากรณ์ตัวแปรตาม (Y) หนึ่งตัว จากกลุ่มตัวแปรอิสระ (X) ตัวเดียวหรือหลายตัวนั้น ตัวแปรอิสระที่นำมาวิเคราะห์จะต้องมีหลักฐานตามทฤษฎีหรือรายงานการวิจัยที่เกี่ยวข้องว่าเป็นตัวแปรต้นเหตุที่ส่งผลต่อตัวแปรตาม



ภาพที่ 2.1 กรอบแนวคิดในการวิเคราะห์ Regression

ที่มา : https://1th.me/N50jW

การวิเคราะห์ความถดถอยพหุคูณหรือเชิงซ้อน (Multiple Regression Analysis) เป็นการวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป เพื่ออธิบายผลของตัวแปรตามที่เกิดขึ้น โดยมีรูปแบบสมการถดถอยพหุคูณคือ

$$\hat{y} = b0 + b1x1 + b2x2 + \cdots + +bkxk$$

โดยที่ bo คือ จุดตัดแกน Y

bi คือ ค่าความชันของตัวแบบถดถอย

พรสิน (2558) ในการวิเคราะห์การถดถอยจำเป็นต้องมีข้อตกลง (assumption) ในการวิเคราะห์ ซึ่งจำนวนข้อตกลงฐานจะแตกต่างกันไปขึ้นอยู่กับประเภทของตัวแบบการถดถอย ข้อตกลงที่สำคัญ เช่น การแจกแจงแบบปกติของตัวแปรตาม (normality) ความเป็นอิสระของตัวแปรตาม (independency) และความแปรปรวนของตัวแปรตามที่คงที่ (homoscedasticity) เป็นต้น หากไม่ได้ให้ความสำคัญกับการตรวจสอบข้อมูลว่าเป็นไปตามข้อตกลงที่กำหนดไว้ หรือหากข้อมูลไม่เป็นไปตามข้อตกลง อาจทำให้อำนาจการทดสอบ (power of test) ต่ำลงหรือความผิดพลาดในการพยากรณ์มากขึ้น นอกเหนือจากการละเมิดข้อตกลงแล้วการที่ข้อมูลบางค่ามีค่าที่ผิดปกติไปจากข้อมูลอื่น ๆ (outlier) หรือการที่ตัวแปรอิสระบางตัวมีความสัมพันธ์กันเอง (multicollinearity) จะมีผลต่อคุณภาพของตัวแบบการถดถอยหรืออาจทำให้สร้างตัวแบบการถดถอยที่ไม่ถูกต้อง

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization เป็นเครื่องมือสำคัญที่ช่วยให้เห็นภาพของข้อมูลเชิงลึกที่ผ่านการเก็บข้อมูล การจัดการข้อมูล การวิเคราะห์ข้อมูล แล้วรวบรวมได้เป็นชุดข้อมูล

2.1.3.1 การสร้าง Data Visualization

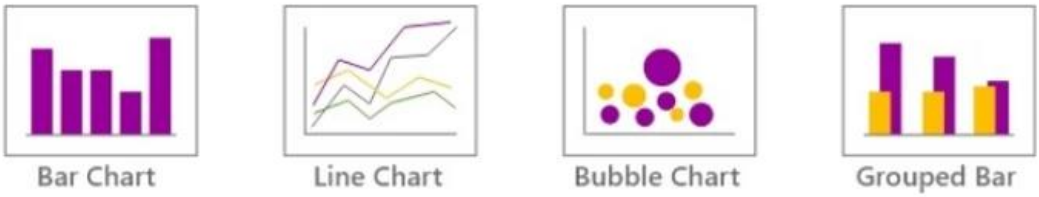
สามารถเริ่มต้นจาก 2 องค์ประกอบหลัก คือ

1) สิ่งที่ต้องการจะสื่อสารจากข้อมูล

2) ลักษณะของข้อมูลที่ต้องการจะสื่อสาร

โดยสิ่งที่ต้องการจะสื่อสารจากข้อมูล คือวัตถุประสงค์หรือความต้องการของผู้ออกแบบ ว่าต้องการจะสร้างกราฟ หรือแผนภูมินี้เพื่อนำเสนออะไร โดยสามารถแบ่งเป็น 4 ประเภทหลัก ๆ คือ

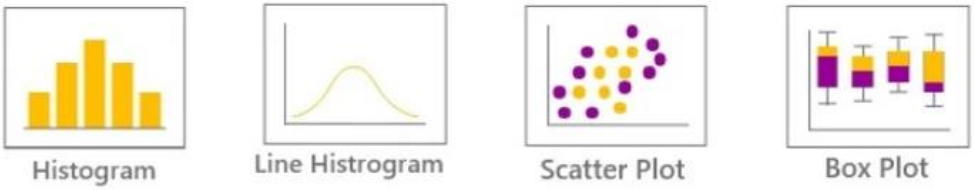
1) การเปรียบเทียบ (Comparison) สามารถใช้เมื่อต้องการเปรียบเทียบข้อมูล โดยจะแสดงให้เห็นความเหมือนหรือความแตกต่างระหว่างข้อมูลที่เราสนใจ โดยรูปแบบของกราฟที่เหมาะสม เช่น



ภาพที่ 2.2 แผนภาพการเปรียบเทียบ

ที่มา : https://www.coraline.co.th/single-post/type-of-chart-in-data-visualization

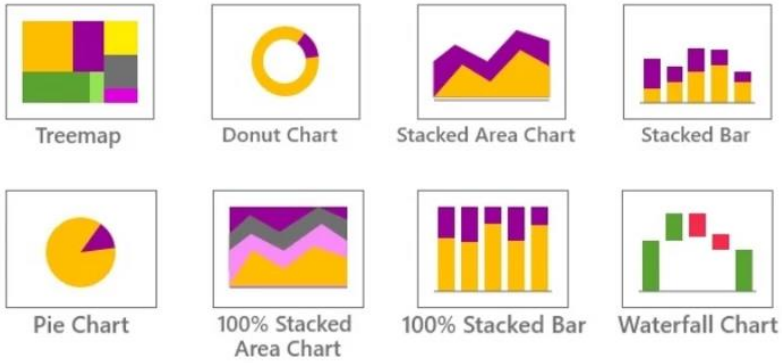
2) การกระจาย (Distribution) สามารถใช้เมื่อต้องการดูความถี่ของข้อมูลว่ามีลักษณะการกระจายตัวอย่างไร เช่น การใช้ Histogram สำหรับการแสดงผลอายุของลูกค้า



ภาพที่ 2.3 แผนภาพการกระจาย

ที่มา : https://www.coraline.co.th/single-post/type-of-chart-in-data-visualization

3) การแบ่งสัดส่วน (Composition) สามารถใช้เมื่อต้องการดูสัดส่วนของข้อมูล ซึ่งมักจะต้องมีข้อมูลตั้งแต่ 2 Category ขึ้นไป โดยกราฟบางรูปแบบอาจจะมีข้อจำกัดในการเลือกใช้ เช่น Pie Chart ที่มักจะเหมาะสมกับข้อมูลที่มีจำนวนไม่เกิน 2-5 Category จึงจะสามารถแสดงข้อมูลออกมาได้ชัดเจน และมีประสิทธิภาพมากพอ



ภาพที่ 2.4 แผนภาพการแบ่งสัดส่วน

ที่มา : https://www.coraline.co.th/single-post/type-of-chart-in-data-visualization

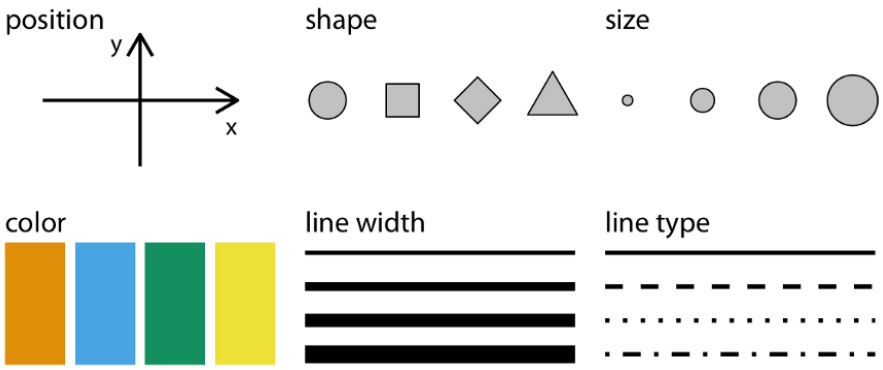
4) ความสัมพันธ์ (Relationship) สามารถใช้เมื่อต้องการดูความสัมพันธ์หรือไม่สัมพันธ์กันของข้อมูล โดยจะต้องมีข้อมูลตั้งแต่ 2 ตัวแปรขึ้นไป

ภาพที่ 2.5 แผนภาพการความสัมพันธ์

ที่มา : https://www.coraline.co.th/single-post/type-of-chart-in-data-visualization

2.1.3.2 ศิลปะและประเภทของข้อมูล

ศิลปะอธิบายทุกแง่มุมขององค์ประกอบกราฟิกที่กำหนด องค์ประกอบที่สำคัญขององค์ประกอบกราฟิกทั้งหมดคือตำแหน่ง ซึ่งจะอธิบายตำแหน่งที่องค์ประกอบมาตรฐานนั้นอยู่ในกราฟิก 2 มิติเราอธิบายตำแหน่งโดยค่า x และ y แต่ระบบพิกัดอื่น ๆ และการสร้างภาพข้อมูลสามมิติเป็นไปได้ อีกส่วนหนึ่งองค์ประกอบกราฟิกทั้งหมดมีรูปร่างขนาดและสี แม้ว่ากำลังเตรียมการวาดภาพขาวดำองค์ประกอบกราฟิกจำเป็นต้องมีสีที่สามารถมองเห็นได้ เช่นสีดำถ้าพื้นหลังเป็นสีขาวหรือสีขาวถ้าพื้นหลังเป็นสีดำ สุดท้ายจะใช้เส้นเพื่อแสดงข้อมูลเส้นเหล่านี้อาจมีความหนาและกว้างที่แตกต่างกันหรือรูปแบบเส้นประจุดนอกเหนือจากตัวอย่างที่แสดงในภาพ แล้วยังมีรูปร่างอื่น ๆ อีกมากมายที่อาจพบในการสร้างภาพข้อมูล ตัวอย่างเช่นหากต้องการแสดงข้อความอาจต้องระบุตระกูลฟอนต์หน้าตัวอักษรและขนาดตัวอักษรและหากวัตถุกราฟิกซ้อนทับกันอาจต้องใช้โปร่งใสหรือทึบเข้ามาช่วย



ภาพที่ 2.6 ศิลปะที่ใช้กันทั่วไปในการแสดงข้อมูล

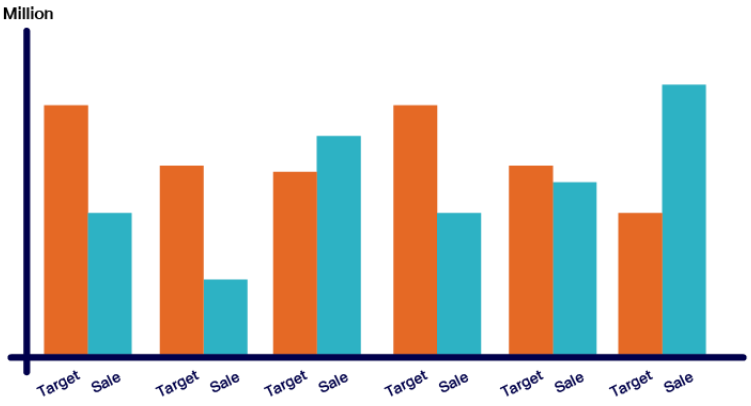
ที่มา : https://1th.me/OzINj

2.1.3.3 การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคที่เทคโนโลยีเข้าถึงผู้คนได้ง่าย ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น การนำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมา Data Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่ง

อาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่น ๆ อีกมากมาย เพื่อให้เข้าใจได้ง่าย และรวดเร็ว การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูลที่มีอยู่ เพื่อให้การนำเสนอข้อมูลน่าสนใจ และมีประสิทธิภาพสูงสุด ซึ่งเลือกแผนภูมิที่นิยมที่ใช้กันบ่อย ๆ ดังนี้

1) แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้ง ที่นิยมแสดงออกมาในรูปแท่งสี่เหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้การเปรียบเทียบจำนวนของข้อมูลในแต่ละชุด เช่น รายรับในแต่ละเดือน ยอดขายที่ขายได้จริงเปรียบเทียบกับเป้าหมายยอดขายที่ตั้งไว้ เป็นต้น ซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท แผนภูมิแท่งแบบจัดกลุ่ม เหมาะสมกับการนำเสนอข้อมูลที่มีข้อมูลย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่ เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะสมกับการนำเสนอข้อมูลที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่ และยังแสดงให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านั้นได้ด้วย ใช้แผนภูมินี้เมื่อมีชุดข้อมูล หลายชุดและต้องการเน้นผลรวมทั้งหมด

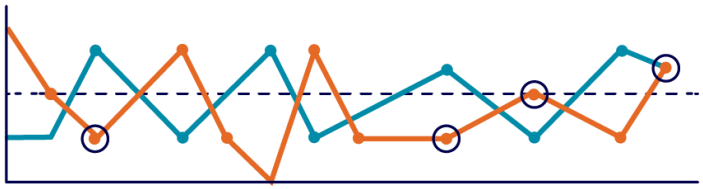


ภาพที่ 2.7 แผนภูมิแท่ง

ที่มา : https://1th.me/9eDde

2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั่น แผนภูมิประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วง

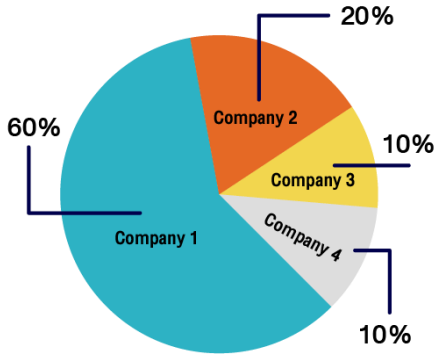
ใช้แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์แนวโน้มในอนาคตได้ เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดูแนวโน้ม เป็นต้น



ภาพที่ 2.8 แผนภูมิเส้น

ที่มา : https://1th.me/9eDde

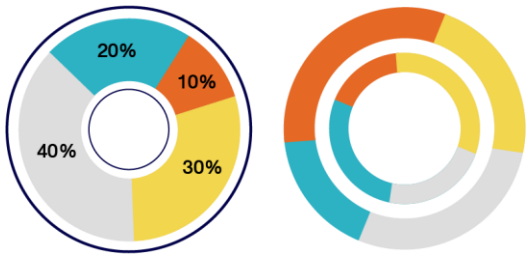
3) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่ายขึ้น และสวยงาม แต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยากเพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share) ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



ภาพที่ 2.9 แผนภูมิวงกลม

ที่มา : https://1th.me/9eDde

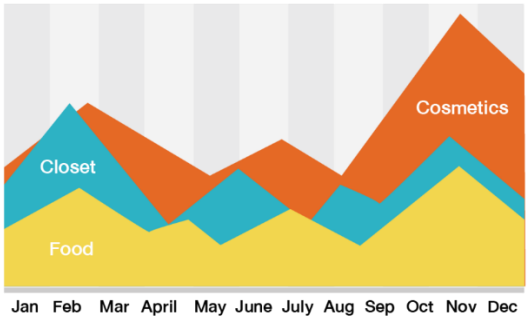
4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลม แต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลาย ๆ ชั้น



ภาพที่ 2.10 แผนภูมิโดนัท

ที่มา : https://1th.me/9eDde

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้เส้นข้อมูล หรือระหว่าง 2 เส้น เพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะสำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลา แสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่าง ๆ เครื่องสำอาง เสื้อผ้าแฟชั่น อาหาร ตามลำดับ

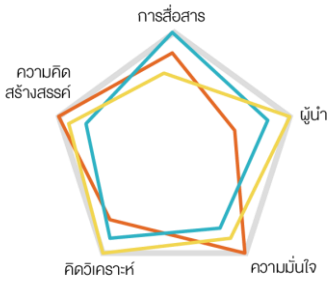


ภาพที่ 2.11 แผนภูมิพื้นที่

ที่มา : https://1th.me/9eDde

6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลม จำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความ

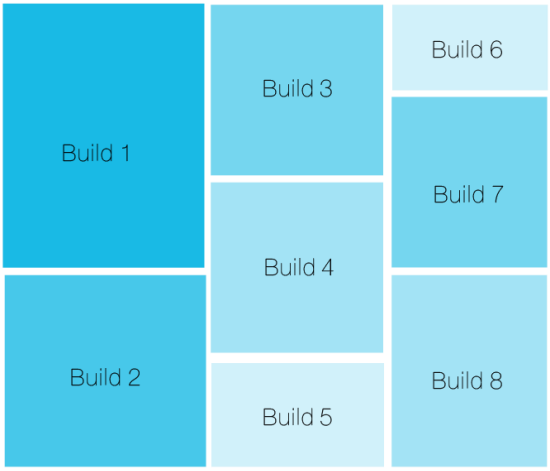
ต่อเนื่องของข้อมูล แต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูล เช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.12 แผนภูมิเรดาร์

ที่มา : https://1th.me/9eDde

7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่แสดงผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึงเขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้

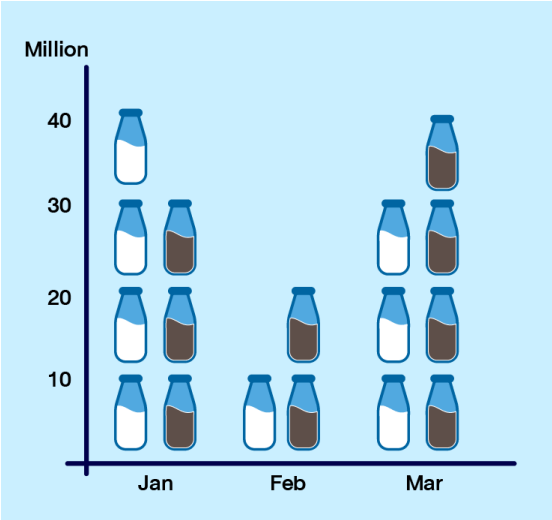


ภาพที่ 2.13 แผนภูมิต้นไม้

ที่มา : https://1th.me/9eDde

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอนและแกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวน

ของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละเดือน ซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปกราฟิกของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้น ซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคน อาจจะแทนด้วยภาพไอคอนคน หรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์แทน ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้ว และยังดึงดูดความสนใจได้มากกว่าการใช้กราฟแท่งสี่เหลี่ยมอีกด้วย



ภาพที่ 2.14 แผนภูมิรูปภาพ

ที่มา : https://1th.me/9eDde

2.1.4 แนวคิดเกี่ยวกับแผงหน้าปัด (Dashboard)

Dashboard คือการนำข้อมูลต่าง ๆ ที่อาจจะเป็นข้อมูลใน report อยู่แล้วหรือข้อมูลใหม่ ๆ มาสรุปให้สามารถเห็นภาพได้ในหน้าเดียว และเป็นข้อมูลที่อัพเดทสม่ำเสมอ เพื่อให้ทางธุรกิจสามารถตัดสินใจได้ทันเวลา ในปัจจุบัน หลายคนก็จะต้องทำ Report โดยที่ส่วนมากมักจะทำแบบ manual โดยการดึงดาต้าออกมา และทำซ้ำเรื่อย ๆ ทุกครั้งที่ต้องทำรีพอร์ต

2.1.4.1 ความหมายของแดชบอร์ด (Dashboard)

หมายถึง หน้ากระดานที่ใช้ในการสรุปข้อมูลแบบ Executive ในมุมมองที่ดูง่าย รวมถึงระบบที่ช่วยสรุปผลการดำเนินงานที่มีการบันทึกข้อมูลทั้งหมด ไม่ว่าจะเป็นการสรุปเอกสาร ใบเสนอราคา ยอดขายสินค้า ยอดขายรอรับชำระ ใบสั่งซื้อที่เปิดอยู่ ยอดซื้อสินค้า ยอดซื้อรอจ่าย ชำระ รายงานวิเคราะห์การขายสุทธิ รายงานเปรียบเทียบยอดขาย และกิจกรรมการนัดหมาย ซึ่งระบบแดชบอร์ดแบ่งออกเป็น 2 ส่วน ได้แก่ My Dashboard คือการรายงานผลการดำเนินการ ตาม Username ที่ Login เข้าสู่ระบบ ซึ่งจะไม่แสดงข้อมูลของผู้ใช้งานท่านอื่น ๆ และ All Dashboard คือ การรายงานผลการดำเนินการของ Username ที่ Login เข้าสู่ระบบทั้งหมด ซึ่งจะแสดงข้อมูล ของผู้ใช้งานท่านอื่น ๆ รวมกัน

2.1.4.2 หลักการในการออกแบบแดชบอร์ด (Dashboard)

1) Users ต้องเข้าใจผู้ใช้เป็นอันดับแรก การเข้าใจผู้ใช้นั้น ให้ยึดหลักการของ Design Thinking หา Insight ของผู้ใช้ หลักของการออกแบบ Dashboard ก็คือ สามารถนำ Dashboard ไปใช้ทำอะไรได้บ้าง เช่น ช่วยตัดสินใจ ติดตามผลการดำเนินงาน ใช้เตือนเมื่อมีสิ่ง ผิดปกติในอดีต

2) Content เมื่อเข้าใจแล้วว่า ผู้ใช้นำ Dashboard ไปใช้อย่างไร ก็สามารถมาดู เนื้อหาที่จะใช้ว่ามีอะไรบ้าง แบ่งเป็น measures หรือ ตัวเลขที่เราสนใจ เช่น ยอดขาย จำนวนลูกค้า dimensions หรือ มุมมองที่เราอยากวิเคราะห์ข้อมูล เช่น ตามช่วงเวลา ตามกลุ่มสินค้า ตามพื้นที่ การเลือก measures นั้น ถ้าเราสามารถคิด ออกแบบให้การวัดผลนั้น มีความน่าสนใจ หรือตรงประเด็นมากขึ้น ก็จะทำให้ Dashboard นั้น มีความน่าสนใจมากขึ้นไปอีก

3) Presentation หลังจากที่รู้แล้วว่า จะใช้ measures และ dimensions แบบ ไหน การเลือกกราฟให้ถูกต้องและเหมาะสมกับข้อมูล

4) Navigation เมื่อได้กราฟแล้ว แสดงข้อมูลทั้งหมดเป็น Dashboard การจัด วางกราฟเป็นส่วนสำคัญ หลัก ๆ คือ กราฟที่เป็นเรื่องเดียวกัน ควรวางไว้ใกล้ ๆ กัน นอกจากนั้น ก็ ควรจัดวางกราฟให้มี visual hierarchy ด้วย คือ จากภาพใหญ่ ไปภาพย่อย เป็นต้น

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับระบบจัดการฐานข้อมูล

2.2.1.1 ความหมายระบบจัดการฐานข้อมูล

ระบบจัดการฐานข้อมูล (Database Management System) หรือที่เรียกว่าดีบีเอ็มเอส (DBMS) คือซอฟต์แวร์สำหรับบริหารและจัดการฐานข้อมูล เปรียบเสมือนสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูล ซึ่งมีหน้าที่ช่วยให้ผู้ใช้เข้าถึงข้อมูลได้ง่ายสะดวกและมีประสิทธิภาพ การเข้าถึงข้อมูลของผู้ใช้ อาจเป็นการสร้างฐานข้อมูล การแก้ไขฐานข้อมูล หรือการตั้งคำถามเพื่อให้ได้ข้อมูลมาโดยผู้ใช้ไม่จำเป็นต้องรับรู้เกี่ยวกับรายละเอียดภายในโครงสร้างของฐานข้อมูล เปรียบเสมือนเป็นสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูลซึ่งต่างจากระบบแฟ้มข้อมูลที่หน้าที่เหล่านี้จะเป็นหน้าที่ของโปรแกรมเมอร์

2.2.1.2 หน้าที่ของระบบการจัดการฐานข้อมูล

1) แปลงคำสั่งที่ใช้จัดการกับข้อมูลภายในฐานข้อมูล ให้อยู่ในรูปแบบที่ฐานข้อมูลเข้าใจ

2) นำคำสั่งต่าง ๆ ซึ่งได้รับการแปลแล้ว ไปสั่งให้ฐานข้อมูลทำงาน เช่น การเรียกใช้ (Retrieve) จัดเก็บ (Update) ลบ (Delete) เพิ่มข้อมูล (Add) เป็นต้น

3) ป้องกันความเสียหายที่จะเกิดขึ้นกับข้อมูลภายในฐานข้อมูล โดยจะคอยตรวจสอบว่าคำสั่งใดที่สามารถทำงานได้ และคำสั่งใดที่ไม่สามารถทำงานได้

4) รักษาความสัมพันธ์ของข้อมูลภายในฐานข้อมูลให้มีความถูกต้องอยู่เสมอ

5) เก็บรายละเอียดต่าง ๆ ที่เกี่ยวข้องกับข้อมูลภายในฐานข้อมูลไว้ในพจนานุกรมข้อมูล (Data Dictionary) ซึ่งรายละเอียดเหล่านี้มักจะถูกเรียกว่า เมทาดาต้า (Metadata) ซึ่งหมายถึง "ข้อมูลของข้อมูล"

6) ดูแลการใช้งานให้กับผู้ใช้ ในการติดต่อกับตัวจัดการระบบแฟ้มข้อมูลได้ โดยจะทำหน้าที่ติดต่อกับระบบแฟ้มข้อมูลซึ่งเสมือนเป็นผู้จัดการแฟ้มข้อมูล (file manager) นำ

ข้อมูลจากหน่วยความจำสำรองเข้าสู่หน่วยความจำหลักเฉพาะส่วนที่ต้องการใช้งาน และทำหน้าที่ประสานกับตัวจัดการระบบแฟ้มข้อมูลในการจัดเก็บ เรียกใช้และแก้ไขข้อมูล

7) ควบคุมการใช้ข้อมูลพร้อมกัน (Concurrency Control) ในระบบคอมพิวเตอร์ที่ใช้อยู่ปัจจุบัน โปรแกรมการทำงานมักจะเป็นแบบผู้ใช้หลายคน (Multi User) จึงทำให้ผู้ใช้แต่ละคนสามารถเรียกใช้ข้อมูลได้พร้อมกัน ระบบจัดการฐานข้อมูลที่มีคุณสมบัติควบคุมการใช้ข้อมูลพร้อมกันนี้ จะทำการควบคุมการใช้ข้อมูลพร้อมกันของผู้ใช้หลายคนในเวลาเดียวกันได้ โดยมีระบบการควบคุมที่ถูกต้องเหมาะสม เช่น ถ้าการแก้ไขข้อมูลนั้นยังไม่เรียบร้อย ผู้ใช้อื่น ๆ ที่ต้องการเรียกใช้ข้อมูลนี้จะไม่สามารถเรียกข้อมูลนั้น ๆ ขึ้นมาทำงานใด ๆ ได้ ต้องรอจนกว่าการแก้ไขข้อมูลของผู้ที่เรียกใช้ข้อมูลนั้นก่อนจะเสร็จเรียบร้อย จึงจะสามารถเรียกข้อมูลนั้นไปใช้งานต่อได้ ทั้งนี้เพื่อป้องกันไม่ให้เกิดปัญหาการเรียกใช้ข้อมูลที่ไม่ถูกต้อง

8) ควบคุมระบบความปลอดภัยของข้อมูลโดยป้องกันไม่ให้ผู้ที่ไม่ได้รับอนุญาตเข้ามาเรียกใช้หรือแก้ไขข้อมูลในส่วนป้องกันเอาไว้ พร้อมทั้งสร้างฟังก์ชันในการจัดทำข้อมูลสำรอง

9) ควบคุมการใช้ข้อมูลในสภาพที่มีผู้ใช้พร้อม ๆ กันหลายคน โดยจัดการเมื่อมีข้อผิดพลาดของข้อมูลเกิดขึ้น

2.2.2 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.2.1 หลักในการออกแบบเว็ปไซต์

เว็บไซต์เป็นสื่อที่ได้รับความนิยมอย่างมากบนอินเตอร์เน็ต ซึ่งเว็บไซต์เป็น สื่อที่อยู่ในความควบคุมของผู้ใช้โดยสมบูรณ์ กล่าวคือ ผู้ใช้สามารถตัดสินใจเลือกได้ว่าจะดูเว็บไซต์ใดและจะไม่เลือกดูเว็บไซต์ใด ได้ตามต้องการ จึงทำให้ผู้ใช้ไม่มีความอดทนต่ออุปสรรค และปัญหาที่เกิดจากการออกแบบเว็บไซต์ผิดพลาดถ้าผู้ใช้เห็นว่าเว็บที่กำลังดูอยู่นั้นไม่มี ประโยชน์ต่อตัวเขาหรือไม่เข้าใจว่าเว็บไซต์นี้จะใช้งานอย่างไร เขาก็สามารถที่จะเปลี่ยนไปดูเว็บไซต์อื่น ๆ ได้อย่างรวดเร็ว เนื่องจากในปัจจุบันมีเว็บไซต์อยู่มากมาย และยังมีเว็บไซต์ที่เกิดขึ้นใหม่ ๆ ทุกวัน ผู้ใช้จึงมีทางเลือกมากขึ้น และสามารถเปรียบเทียบคุณภาพของเว็บไซต์ต่าง ๆ ได้เองเว็บไซต์ที่ได้รับการ

ออกแบบอย่างสวยงาม มีการใช้งานที่สะดวก ย่อมได้รับความสนใจจากผู้ใช้ มากกว่าเว็บไซต์ที่ดูสับสนวุ่นวาย มีข้อมูลมากมายแต่หาอะไรไม่เจอ นอกจากนี้ยังใช้เวลาในการแสดงผลแต่ละหน้านานเกินไป ซึ่งปัญหาเหล่านี้ล้วนเป็นผลมาจากการออกแบบเว็บไซต์ไม่ดีทั้งสิ้น ดังนั้นการออกแบบเว็บไซต์จึงเป็นกระบวนการสำคัญในการสร้างเว็บไซต์ ให้ประทับใจผู้ใช้ ทำให้เขาอยากกลับเข้ามาเว็บไซต์เดิมอีกในอนาคต ซึ่งนอกจากต้องพัฒนาเว็บไซต์ที่ดีมีประโยชน์แล้ว ยังต้องคำนึงถึงการแข่งขันกับเว็บไซต์อื่น ๆ อีกด้วย การออกแบบเว็บไซต์ จึงต้องควรคำนึงถึง

1) ความเรียบง่าย เข้าใจง่าย

การออกแบบเว็บไซต์ที่ดี จะต้องเน้นที่ความเรียบง่ายเป็นหลักโดยเลือกนำเสนอเฉพาะสิ่งที่ต้องการนำเสนอจริง ๆ ในรูปแบบที่หลากหลาย โดยอาจจะเป็นสีสัน กราฟิก ภาพเคลื่อนไหวหรือตัวอักษร ที่สำคัญจะต้องมีการนำเสนอที่ไม่ดูรกหน้าเว็บจนเกินไป เพื่อไม่ให้เกิดความรู้สึกรกสายตา หรือสร้างความเบื่อหน่าย น่ารำคาญให้กับผู้ที่เข้าชมเว็บไซต์

2) ความสม่ำเสมอ ไม่สับสน

ควรออกแบบเว็บไซต์ด้วยความสม่ำเสมอ คือจะต้องมีรูปแบบ กราฟิก โทนสีและการตกแต่งต่าง ๆ ให้แต่ละหน้าบนเว็บไซต์มีความคล้ายคลึงกัน และเป็นแนวเดียวกันไปตลอดทั้งเว็บไซต์ ดังตัวอย่างเว็บไซต์ทั่ว ๆ ไปที่จะสังเกตเห็นได้ว่าทุกหน้าของเว็บไซต์นั้นจะเน้นการตกแต่งในรูปแบบเดียวกันทั้งหมด ต่างก็แค่การนำเสนอของแต่ละหน้าเท่านั้น

3) สร้างความโดดเด่น เป็นเอกลักษณ์

การออกแบบเว็บไซต์เพื่อให้สามารถสื่อถึงจุดประสงค์ในการนำเสนอเว็บได้ดี จะต้องมีการสร้างความเป็นเอกลักษณ์และจุดเด่นให้กับเว็บไซต์ เพื่อให้สามารถสะท้อนถึงลักษณะขององค์กรได้มากที่สุด โดยการสร้างเอกลักษณ์ดังกล่าวนั้น อาจใช้ชุดสี รูปภาพ ตัวอักษร หรือกราฟิก นอกจากนี้ก็ต้องขึ้นอยู่กับว่า เป็นเว็บไซต์แบบทางการหรือไม่ เพื่อจะได้ออกแบบได้อย่างเหมาะสมที่สุด

4) เนื้อหาต้องดี ครบถ้วน

เนื้อหาเป็นสิ่งที่สำคัญที่สุดของการสร้างเว็บไซต์ เพราะสิ่งที่ทำให้ผู้คนเกิดความสนใจ และหมั่นติดตามเว็บไซต์เหล่านั้นอยู่เสมอ ก็คือเนื้อหาที่มีความสมบูรณ์และน่าสนใจ นอกจากนี้จะต้องมีการปรับปรุง พัฒนาเนื้อหาบนเว็บให้มีความทันสมัยอยู่เสมอรวมถึงข้อมูลต้องมีความถูกต้องที่สุด

5) ระบบเนวิเกชัน ใช้ง่าย

ระบบเนวิเกชัน เป็นเสมือนป้ายบอกทางเพื่อให้ผู้ใช้งาน ไม่เกิดความสับสนในขณะใช้งานเว็บไซต์ ซึ่งการออกแบบเนวิเกชันก็จะต้องเน้นที่ความเรียบง่าย ใช้งานสะดวก และมีความเข้าใจได้ง่าย ที่สำคัญจะต้องมีตำแหน่งการวางที่สม่ำเสมอเพื่อให้ดูเป็นแนวทางเดียวกัน ทำให้ผู้ใช้งานหรือผู้ชมรู้สึกประทับใจ และจดจ าเว็บไซต์ได้ง่ายขึ้น ส่วนใครที่มีการนำกราฟิกมาใช้ในระบบเนวิเกชัน ก็จะต้องเลือกกราฟิกที่สามารถสื่อความหมายได้ดีเช่นกัน

6) คุณภาพของเว็บไซต์

เว็บไซต์ที่ดีจะต้องมีคุณภาพ ทั้งสิ่งที่ปรากฏให้เห็นบนเว็บไซต์ ไม่ว่าจะเป็นกราฟิก ชนิดตัวอักษร รูปภาพหรือสีสันที่ใช้ เนื้อหาที่นำมาแสดงผล ซึ่งหากเว็บไซต์มีคุณภาพก็จะสร้างความน่าเชื่อถือ และเป็นจุดเด่นที่ทำให้ผู้คนส่วนใหญ่เกิดความสนใจได้ดี เพราะฉะนั้นห้ามละเลยในส่วนของคุณภาพเด็ดขาด

7) ความสะดวกในการเข้าใช้งาน

เว็บไซต์ควรให้ความสะดวกสบายแก่ผู้ใช้งานได้ดี คือจะต้องมีการแสดงผลได้ในทุกระบบปฏิบัติการ ไม่ว่าจะเป็นเว็บบราว์เซอร์ คอมพิวเตอร์ โน้ตบุ๊ค หรือบนโทรศัพท์มือถือ ที่สำคัญจะต้องมีความละเอียดของการแสดผลและสามารถใช้งานได้โดยไม่มีปัญหาด้วย

8) ความคงที่ของการออกแบบ

การออกแบบเว็บไซต์ควรจะมีความคงที่ในการออกแบบ ด้วยการสร้างเว็บไซต์ด้วยแบบแผนเดียวกัน และมีการเรียบเรียงเนื้อหาอย่างรอบคอบ ทำให้เว็บมีความน่าเชื่อถือ และดูมีคุณภาพ ช่วยสร้างความประทับใจให้กับผู้ใช้งานได้เป็นอย่างดี

9) ความคงที่ของการทำงาน

ระบบการทำงานบนเว็บไซต์จะต้องมีความคงที่ และสามารถใช้งานได้ดี ซึ่งนอกจากการออกแบบระบบการทำงานให้มีความทันสมัยและสร้างสรรค์แล้ว ก็จะต้องหมั่นตรวจสอบอยู่เสมอ เพราะหากระบบการใช้งานมีความผิดปกติก็จะได้แก้ปัญหาได้ทัน นอกจากนี้อาจมีการอัพเดตดีไซน์ให้ทันสมัยขึ้นบ่อย ๆ เพื่อให้ผู้ใช้งานรู้สึกสนุกไปกับการใช้งานเว็บไซต์

2.2.2.2 รูปแบบโครงสร้างของเว็บไซต์

โครงสร้างเว็บไซต์ เป็นส่วนที่สำคัญส่วนหนึ่งที่ผู้ออกแบบเว็บจะต้องคำนึงถึง เพราะโครงสร้างเว็บไซต์เป็นแผนผังของการลำดับเนื้อหาหรือการจัดวางตำแหน่งของเว็บเพจทั้งหมด ซึ่งทำให้ทราบว่าเว็บไซต์นั้นประกอบไปด้วยเนื้อหาอะไรบ้างและมีเว็บเพจไหนที่มีความเกี่ยวข้องเชื่อมโยงกัน การศึกษาด้านโครงสร้างเว็บไซต์จึงเปรียบเสมือนแบบจำลอง ที่ทำให้ผู้ออกแบบเว็บเห็นหน้าตาของเว็บที่อยู่ในรูปธรรมมากขึ้น

การเลือกวิธีการจัดวางโครงสร้างเว็บไซต์นั้น ขึ้นอยู่กับผู้ออกแบบเว็บไซต์ว่าต้องการให้ออกมาในรูปแบบใด โดยคำนึงถึงความสมดุลของโครงสร้าง มีการเชื่อมโยงสัมพันธ์กันระหว่างหน้าเว็บเพจต่างๆ รวมถึงการเชื่อมโยงภายในแต่ละหน้าเว็บเพจด้วย ซึ่งการวางแผนโครงสร้างจะต้องวางให้ดีเพื่อจะเป็นการป้องกันปัญหาที่อาจจะเกิดกับผู้ใช้ เช่น การที่ผู้ใช้ไม่สามารถย้อนกลับมาดูข้อมูลในหน้าหลักได้ เป็นต้น การออกแบบโครงสร้างของเว็บไซต์ที่ได้รับความนิยมมาก ในปัจจุบันสามารถออกแบบได้หลายรูปแบบ โดยสามารถแบ่งออกได้เป็น 4 รูปแบบใหญ่ คือ

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure)

เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ดรรชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินหน้า–ถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อหา

ภาพที่ 2.15 เว็บที่มีโครงสร้างแบบเรียงลำดับ

ที่มา : https://1th.me/vZJuU

2) เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure)

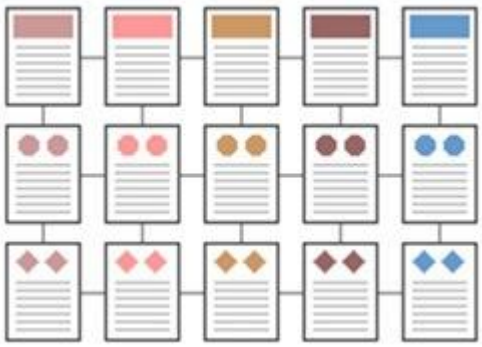
เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่าง ๆ และมีรายละเอียดย่อย ๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิองค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียวนั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

ภาพที่ 2.16 เว็บที่มีโครงสร้างแบบลำดับขั้น

ที่มา : https://1th.me/vZJuU

3) เว็บที่มีโครงสร้างแบบตาราง (Grid Structure)

โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้



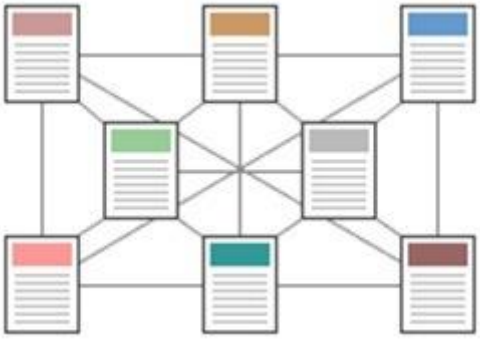
ภาพที่ 2.17 เว็บที่มีโครงสร้างแบบตาราง

ที่มา : https://1th.me/vZJuU

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และ สามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับโครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้นๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจเกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure)

โครงสร้างประเภทนี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้างรูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วย ตนเอง การเชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มีโครงสร้างที่แน่นนอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหา ภายในเว็บนั้นๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.18 เว็บที่มีโครงสร้างแบบใยแมงมุม

ที่มา : https://1th.me/vZJuU

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปสู่หน้าใดๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการ ปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและ เกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3 ทฤษฎีเกี่ยวกับอนุกรมเวลา (Time Series)

อนุกรมเวลาคือเซตของข้อมูลเชิงปริมาณที่มีการจัดเก็บในช่วงเวลาที่ติดต่อกัน ส่วนข้อมูลอนุกรมเวลา (Time Series Data) คือ ชุดของข้อมูลที่มีการเก็บรวบรวมตามระยะเวลาที่ติดต่อกันอย่างเป็นระบบโดยทั่ว ไปข้อมูลอนุกรมเวลาประกอบด้วยองค์ประกอบ 4 ส่วน คือ แนวโน้ม (Trend : T), ฤดูกาล (Seasonal : S), วัฏจักร (Cycle : C) และเหตุการณ์ที่ผิดปกติหรือเหตุการณ์ความไม่แน่นอน (Irregular: I) สำหรับรูปแบบของอนุกรมเวลาโดยทั่วไปนั้น มีอยู่ 2 รูปแบบคือ

1) รูปแบบบวก $Y = T + S + C + I$

2) รูปแบบคูณ $Y = T \times S \times C \times I$

โดยที่ Y คือ อนุกรมเวลา

T คือ อิทธิพลของแนวโน้ม

S คือ อิทธิพลของฤดูกาล

C คือ อิทธิพลของวัฏจักร

I คือ อิทธิพลของความไม่แน่นอน

การวิเคราะห์อนุกรมเวลาเป็นวิธีที่ใช้ในการวิเคราะห์ข้อมูลหรือค่าสังเกตที่มีการเปลี่ยนแปลงไปตามลำดับ เวลาที่เกิดขึ้น หรือการเปลี่ยนแปลงของตัวแปรในช่วงเวลาที่ผ่านไป ลักษณะ ของการเปลี่ยนแปลงอาจมีหรือไม่มีรูปแบบก็ได้ แต่ถ้าอนุกรมเวลาแสดงให้เห็นรูปแบบการเปลี่ยนแปลงในช่วงเวลาที่ผ่านมาในอดีตก็จะทำ ให้สามารถคาดการณ์ได้ว่าในอนาคต ลักษณะการเปลี่ยนแปลงควรอยู่ในรูปแบบใด และสามารถพยากรณ์การเปลี่ยนแปลงข้อมูลในอนาคตได้ การวิเคราะห์ข้อมูลอนุกรมเวลานี้ จะขึ้นอยู่กับการเปลี่ยนแปลงของเวลาในอดีตเป็นพื้นฐาน

2.2.3.1 องค์ประกอบของอนุกรมเวลา

ในการวิเคราะห์อนุกรมเวลา ผู้วิเคราะห์จะแยกองค์ประกอบต่าง ๆ ที่ประกอบกันขึ้นเป็นอนุกรมเวลา โดยจะมีการเปลี่ยนแปลงไปตามอิทธิพลต่าง ๆ เช่น การเปลี่ยนแปลงการผลิต เทคโนโลยี สภาวะอากาศ เป็นต้น ในการหาคุณลักษณะของอนุกรมเวลาเรา

สามารถใช้แบบจำลองได้หลายแบบ แบบจำลองที่ใช้โดยนักเศรษฐศาสตร์แบบหนึ่ง คือ แบบจำลองแบบคลาสสิก (classical model) เป็นการอธิบายถึงองค์ประกอบของการแปรผันของอนุกรมเวลา 4 ส่วน ดังนี้

1) ค่าแนวโน้ม (Secular trend) แทนด้วย Tt เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะราบเรียบ แนวโน้ม อาจมีลักษณะเป็นเส้นตรงหรือเส้นโค้งในทางเพิ่มขึ้นหรือลดลง ค่าแนวโน้มของข้อมูลเป็นการเคลื่อนไหวในช่วงระยะเวลาที่ค่อนข้างนานพอสมควร ควรเป็นข้อมูลรายปี และควรมีข้อมูลอย่างน้อย 15 ปี ซึ่งจะแสดงทิศทางของอนุกรมเวลาตัวอย่าง กราฟแสดงค่าแนวโน้มของผลผลิตเคมีภัณฑ์ชนิดหนึ่ง

2) การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) แทนด้วย St เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะการเพิ่มขึ้น หรือลดลงในลักษณะเดียวกันของรอบระยะเวลาหนึ่งที่แน่นอน เรียกว่า การเปลี่ยนแปลงตามฤดูกาล หน่วยของระยะเวลาสำหรับข้อมูลอาจเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือน รายไตรมาส สำหรับข้อมูลรายปีไม่มีการแปรผันตามฤดูกาล การเปลี่ยนแปลงตามฤดูกาลนั้นกำหนดระยะเวลาการเกิดซ้ำในรอบหนึ่ง ๆ ได้ค่อนข้างแน่นอน ตัวอย่างเช่น ยอดขายรายเดือนของห้างสรรพสินค้าแห่งหนึ่ง

3) การเปลี่ยนแปลงหรือความแปรผันตามวัฏจักร (Cyclical Variation) แทนด้วย Ct การเปลี่ยนแปลงตามวัฏจักร มีการเปลี่ยนแปลงเคลื่อนไหวในลักษณะซ้ำ ๆ กันและจะมีลักษณะคล้ายคลึงกับการเปลี่ยนแปลงตามฤดูกาล จะต่างกันก็ตรงที่การเปลี่ยนแปลงตามวัฏจักรแต่ละรอบจะใช้ระยะเวลาที่นานกว่า คือ ตั้งแต่ 5 ปีขึ้นไป ข้อมูลที่มีการเปลี่ยนแปลงตามวัฏจักรในทางธุรกิจ เรียกว่า "วัฏจักรธุรกิจ" (Business Cyclical) โดยทั่วไปประกอบด้วย ระยะเจริญรุ่งเรือง (prosperity) ระยะฝืดเคือง (recession) ระยะตกต่ำ (depression) และระยะขยายตัว (recovery)

การเปลี่ยนแปลงหรือความแปรผัน เนื่องจากเหตุการณ์ผิดปกติ (Irregularly Variation) แทนด้วย It เป็นการเปลี่ยนแปลงของข้อมูลอนุกรมเวลาที่เกิดจากเหตุการณ์ที่เราไม่สามารถคาดการณ์ได้ล่วงหน้า เช่น การเกิดไฟไหม้ในโรงงาน การเกิดอุทกภัย การนัดหยุดงานของ

คนงาน แผ่นดินไหว เป็นต้น ซึ่งเหตุการณ์เหล่านี้เป็นสิ่งที่เกิดขึ้นโดยบังเอิญไม่คาดคิดมาก่อน เป็นการเปลี่ยนแปลงที่เป็นเชิงสุ่ม (random variation) เพราะไม่ได้อยู่ภายใต้เงื่อนไขที่เรากำหนด จากองค์ประกอบของอนุกรมเวลาทั้ง 4 อย่าง คือ T S C และ I ในข้อมูลอนุกรมชุดหนึ่ง ไม่จำเป็นต้องครบองค์ประกอบข้างต้นก็ได้ ทั้งนี้ ขึ้นอยู่กับชนิดของมูลของเรา

2.2.3.2 รูปแบบของอนุกรมเวลา

จากปัจจัยทั้ง 4 ข้างต้น ถ้า Y แทนข้อมูลอนุกรมเวลาชุดหนึ่ง ๆ เราสามารถกำหนดแบบจำลองได้ 2 แบบ ดังนี้

1) แบบจำลองผลบวก (Additive model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลบวกขององค์ประกอบทั้ง 4 อย่าง

$$Yt = Tt + St + Ct + It$$

2) แบบจำลองผลคูณ (Multiplicative model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลคูณขององค์ประกอบทั้ง 4 อย่าง

$$Y\,t = Tt * St * Ct * It$$

โดยทั่วไปข้อมูลอนุกรมเวลา ในทางธุรกิจจะมีความสัมพันธ์ในรูปแบบจำลองผลคูณ เนื่องจากเป็นการพิจารณาการเปลี่ยนแปลงในรูปอัตราร้อยละ ซึ่งจะทำให้ผลการวิเคราะห์ใกล้เคียงความเป็นจริงมากกว่าการใช้แบบจำลองผลบวก

2.2.4 ทฤษฎีเกี่ยวกับกระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization)

นอร์มัลไลเซชัน (Normalization) การทำนอร์มัลไลเซชัน เป็นวิธีการในการกำหนดแอตทริบิวต์ให้กับแต่ละเอนทิตี เพื่อให้ได้โครงสร้างของตารางที่ดี สามารถควบคุมความซ้ำซ้อนของข้อมูลหลีกเลี่ยงความผิดปกติของข้อมูล โดยทั่วไปผลลัพธ์ของการนอร์มัลไลเซชัน จะได้ตารางที่มีโครงสร้างซับซ้อนน้อยลง แต่จำนวนของตารางจะมากขึ้น

การทำนอร์มัลไลเซชัน จะประกอบด้วยนอร์มัลฟอร์ม (Normal Form) แบบต่าง ๆ ที่มีเงื่อนไขของการทำให้อยู่ในรูปของนอร์มัลฟอร์มที่แตกต่างกันไป ขึ้นอยู่กับผู้ออกแบบฐานข้อมูลว่า

ต้องการลดความซ้ำซ้อนในฐานข้อมูลให้อยู่ในระดับใด ซึ่งประกอบด้วยนอร์มัลฟอร์มแบบต่าง ๆ ดังต่อไปนี้

นอร์มัลฟอร์มที่ 1 (First Normal Form : 1NF)

นอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

นอร์มัลฟอร์มที่ 3 (Third Normal Form : 3NF)

บอยซ์คอดด์นอร์มัลฟอร์ม (Boyce–Codd Normal Form : BCNF)

นอร์มัลฟอร์มที่ 4 (Fourth Normal Form : 4NF)

นอร์มัลฟอร์มที่ 5 (Fifth Normal Form : 5NF)

ถึงแม้ว่าการนอร์มัลไลเซชัน จะเป็นสิ่งสำคัญและจำเป็นที่สุดสำหรับการออกแบบฐานข้อมูล แต่ก็ไม่ได้หมายความว่าจะต้องทำการนอร์มัลไลเซชันจนถึงระดับนอร์มัลฟอร์มที่ 5 โดยทั่วไปการแสดงผลข้อมูลจากตารางที่อยู่ในนอร์มัลฟอร์มที่ 5 จะมีการเชื่อมต่อตารางเป็นจำนวนมาก ทำให้การแสดงผลและการโต้ตอบระหว่างระบบฐานข้อมูลกับผู้ใช้กระทำได้ช้า การออกแบบฐานข้อมูลที่ดีจึงต้องพิจารณาถึงความต้องการของผู้ใช้และต้องสามารถตอบสนองได้อย่างรวดเร็ว เพราะฉะนั้นในบางกรณีจึงมีการลดระดับการนอร์มัลไลเซชันในบางส่วนของการออกแบบฐานข้อมูล เพื่อให้ระบบสามารถตอบสนองได้ตามความต้องการของผู้ใช้ การลดระดับการนอร์มัลไลเซชัน (Denormalization) เป็นวิธีการลดระดับของนอร์มัลฟอร์มลงมา เช่น การแปลงจาก 3NF มาเป็น 2NF อย่างไรก็ตาม สิ่งที่จะได้รับเพิ่มขึ้นมาจากการลดระดับการนอร์มัลไลเซชันนอกจากความเร็วที่ดีขึ้นแล้ว ความซ้ำซ้อนของข้อมูลก็เพิ่มสูงขึ้นด้วย ซึ่งเป็นสิ่งที่ควรนำมาพิจารณาอย่างระมัดระวัง

1) การแปลงให้อยู่ในรูปนอร์มัลฟอร์มที่ 1 (First Normal Form : 1NF)

คุณสมบัติของรีเลชันของแบบจำลองข้อมูลเชิงสัมพันธ์ ก็คือ ข้อมูลในแต่ละทัปเพิลจะต้องไม่ซ้ำกัน และค่าในแต่ละแอตทริบิวต์จะต้องไม่สามารถถูกแบ่งแยกย่อยลงไปได้อีก หรือมีความเป็นอะตอมมิค (Atomic) รวมถึงจะต้องมีค่าเพียงค่าเดียวที่อยู่ในแต่ละแอตทริบิวต์หรือ

มีความเป็นซิงเกิลแวลู (Single Value) ซึ่งในการทำนอร์มัลไลเซชันให้อยู่ในนอร์มัลฟอร์ที่ 1 ก็อาศัยคุณสมบัติดังที่กล่าวไว้ข้างต้น

1.1 รีพีทติ้งกรุ๊ป (Repeating Group) การที่ข้อมูลใน 1 ทัปเพิล สามารถมีค่าในแต่ละแอตทริบิวต์ได้มากกว่าหนึ่งค่า (Multivalued) จะทำให้เกิดรีพีทติ้งกรุ๊ป ดังตารางที่แสดงในภาพข้างล่าง ซึ่งเลขที่โครงการหนึ่งหมายเลขประกอบด้วยกลุ่มข้อมูลหลายกลุ่ม ซึ่งทำให้รีเลชันดังกล่าว ขาดคุณสมบัติซิงเกิลแวลู

การทำงานของพนักงานในโครงการ

เลขที่โครงการ	ชื่อโครงการ	รหัสนักงาน	ชื่อพนักงาน	ตำแหน่งงาน	ค่าแรง/ ชม.	จำนวน ชม.
11	RFID	103	สมชาย	Engineer	500	23.8
		101	วิชา	Programmer	500	19.4
		102	สุรชัย	Administrator	200	12.6

ภาพแสดงการเกิดรีพีทติ้งกรุ๊ป

ภาพที่ 2.19 รีพีทติ้งกรุ๊ป (Repeating Group)

ที่มา : http://1ab.in/Cyg

1.2 นิยามของนอร์มัลฟอร์มที่ 1 รีเลชันจะอยู่ในรูปของนอร์มัลฟอร์มที่ 1 ก็ต่อเมื่อมีคุณสมบัติตามเงื่อนไขดังต่อไปนี้

1. มีการกำหนดแอตทริบิวต์ที่เป็นคีย์
2. ต้องไม่มีรีพีทติ้งกรุ๊ป แต่ละแถวหรือคอลัมน์จะมีค่าได้เพียง 1 ค่าเท่านั้น
3. แอตทริบิวต์ทุกตัวต้องขึ้นอยู่กับคีย์หลัก

จากภาพข้างบน เมื่อการการนอร์มัลไลเซชันให้อยู่ในรูปนอร์มัลฟอร์มที่ 1 จะได้ตารางที่แตกย่อยออกมาเป็น 2 ตาราง ดังภาพข้างล่าง ซึ่งมีคุณสมบัติตามนอร์มัลฟอร์มที่ 1 แล้ว

การทำงานของพนักงาน

รหัสนักงาน	ชื่อพนักงาน	ตำแหน่งงาน	ค่าแรง/ ชม.	จำนวน ชม.
103	สมชาย	Engineer	500	23.8
101	วิชา	Programmer	500	19.4
102	สุรชัย	Administrator	200	12.6

ชื่อโครงการ

เลขที่โครงการ	ชื่อโครงการ
11	RFID

รีเลชันที่อยู่ในรูปนอร์มัลฟอร์มที่ 1

ภาพที่ 2.20 นิยามของนอร์มัลฟอร์มที่ 1

ที่มา : http://1ab.in/Cyg

2) การแปลงให้อยู่ในรูปนอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

ในหนึ่งรีเลชันจะประกอบด้วยแอตทริบิวต์ต่าง ๆ ที่มีความสัมพันธ์ที่ขึ้นต่อกัน ซึ่งความสัมพันธ์ดังกล่าวจะเป็นตัวกำหนดว่าแอตทริบิวต์ใดเป็นตัวกำหนดข้อมูล หรือคีย์แอตทริบิวต์ (Key Attribute) และและแอตทริบิวต์ใดเป็นข้อมูลที่ถูกกำหนดหรือนอนคีย์แอตทริบิวต์ (Non key Attribute)

2.1) ฟังก์ชันนัลดีเพนเดนซี (Functional Dependency: FD) ในการทำนอร์มัลไลเซชัน จะต้องมีความเข้าใจหลักการของฟังก์ชันดีเพนเดนซี (Function Dependency: FD) เสียก่อน โดยมีคำจำกัดความคือ B ขึ้นอยู่กับ A ถ้าทราบค่าของ A ก็จะทำให้รู้ค่าของ B ได้ ฟังก์ชันนัลดีเพนเดนซี สามารถแสดงด้วยการใช้เครื่องหมายลูกศร (–>) ตัวอย่างเช่น A–>B แสดง B เป็นฟังก์ชันนัลดีเพนเดนต์กับ A กล่าวคือ ถ้ารู้ค่า A ก็จะทำให้ทราบค่าของ B ด้วย ทุกค่าของ A ที่มีค่าเท่ากัน จะได้ค่าเท่ากันเสมอ

2.2) พาเชียลดีเพนเดนซี (Partial Dependency) พาร์เชียลดีเพนเดนซี หมายถึงการที่มีแอตทริบิวต์บางแอตทริบิวต์ที่ขึ้นอยู่กับเพียงบางส่วนของคีย์หลักเท่านั้น ตัวอย่างเช่น จากตารางในภาพข้างล่าง แอตทริบิวต์ชื่อพนักงานจะขึ้นอยู่กับคีย์รหัสพนักงาน ในขณะที่แอตทริบิวต์ชื่อแผนก จะขึ้นอยู่กับคีย์รหัสแผนก จะเห็นว่าข้อมูลที่อยู่ในรีเลชันเดียวกัน แต่ไม่ได้ขึ้นอยู่กับคีย์ใดคีย์หนึ่งทั้งหมด แต่จะขึ้นอยู่กับคีย์ใดคีย์หนึ่งเพียงบางส่วนเท่านั้น

ภาพที่ 2.21 พาเชียลดีเพนเดนซี (Partial Dependency)

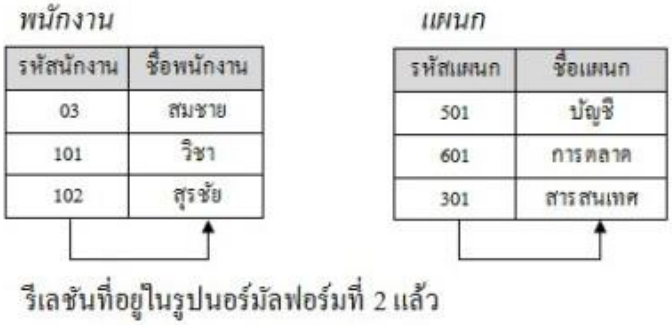
ที่มา : http://1ab.in/Cyg

2.3) นิยามของนอร์มัลฟอร์มที่ 2 รีเลชันจะอยู่ในรูปของนอร์มัลฟอร์มที่ 2 ก็ต่อเมื่อมีคุณสมบัติตามเงื่อนไขดังต่อไปนี้

1) รีเลชันนั้นเป็นนอร์มัลฟอร์มที่ 1 อยู่แล้ว

2) รีเลชันนั้นไม่มีพาร์เชียลดีเพนเดนซี

ตัวอย่างรีเลชันพนักงานในแผนกในภาพข้างบน เมื่อทำการแตกออกเป็นรีเลชันย่อยที่ไม่มีพาร์เชียลดีเพนเดนซีแล้ว จะได้เป็นรีเลชันสองรีเลชัน คือ รีเลชันพนักงานและ รีเลชันแผนก ซึ่งอยู่ในรูปของนอร์มัลฟอร์มที่ 2 แล้ว ดังภาพข้างล่าง



ภาพที่ 2.22 รีเลชันที่อยู่ในรูปนอร์มัลฟอร์มที่ 2 แล้ว

ที่มา : http://1ab.in/Cyg

3) การแปลงให้อยู่ในรูปนอร์มัลฟอร์มที่ 3 (Third Normal Form: 3NF)

ในหนึ่งรีเลชันจะประกอบคีย์แอตทริบิวต์และนอนคีย์แอตทริบิวต์ คีย์แอตทริบิวต์จะต้องเป็นตัวกำหนดความหมายหรือการมีอยู่ของแอตทริบิวต์อื่น ๆ ที่อยู่ในรีเลชันเสมอ

3.1) ทรานซิทีฟดีเพนเดนซี (Transitive Dependency)

ทรานซิทีฟดีเพนเดนซี หมายถึง การที่มีฟังก์ชันนัลดีเพนเดนซี ระหว่างแอตทริบิวต์ที่ไม่ได้เป็นส่วนของคีย์ใด ๆ แต่มีแอตทริบิวต์อื่น ๆ มาขึ้นกับแอตทริบิวต์นั้นตัวอย่างเช่น จากตารางในภาพข้างล่าง แอตทริบิวต์ชื่อพนักงาน และรหัสตำแหน่งงานจะขึ้นอยู่กับคีย์รหัสพนักงาน ในขณะที่แอตทริบิวต์ค่าแรงต่อชั่วโมงของพนักงาน จะขึ้นอยู่กับแอตทริบิวต์รหัสตำแหน่งงานซึ่งไม่ใช่คีย์อีกต่อหนึ่งทำให้มีทรานซิทีฟดีเพนเดนซีเกิดขึ้นในรีเลชันนี้

ภาพที่ 2.23 ตารางที่มีทรานซิทีฟดีเพนเดนซี

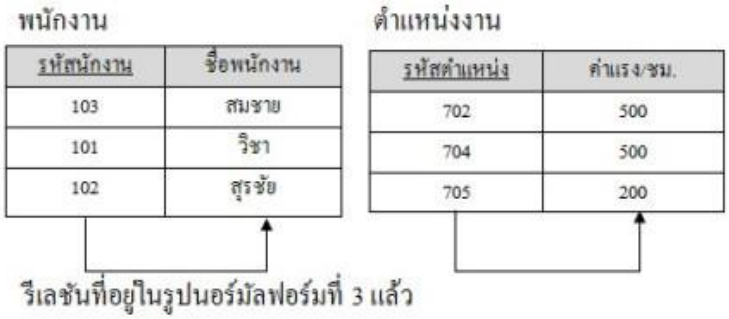
ที่มา : http://1ab.in/Cyg

3.2) นิยามของนอร์มัลฟอร์มที่ 3 รีเลชันจะอยู่ในรูปของนอร์มัลฟอร์มที่ 3 ก็ต่อเมื่อมีคุณสมบัติตามเงื่อนไขดังต่อไปนี้

1) รีเลชันนั้นเป็นนอร์มัลฟอร์มที่ 2 อยู่แล้ว

2) รีเลชันนั้นไม่มีทรานซิทีฟดีเพนเดนซี

ตัวอย่างรีเลชัน การทำงานของพนักงาน ในภาพข้างบน เมื่อทำการแตกออกเป็นรีเลชันย่อยที่ไม่มีทรานซิทีฟดีเพนเดนซีแล้ว จะได้เป็นรีเลชันสองรีเลชัน คือรีเลชันพนักงาน และรีเลชันตำแหน่งงาน ซึ่งอยู่ในรูปของนอร์มัลฟอร์มที่ 3 แล้ว ดังภาพข้างล่าง



ภาพที่ 2.24 รีเลชันที่อยู่ในรูปนอร์มัลฟอร์มที่ 3

ที่มา : http://1ab.in/Cyg

4) การแปลงให้อยู่ในรูปบอยซ์คอดด์นอร์มัลฟอร์ม (Boyce–Codd Normal Form : BCNF) ในหนึ่งรีเลชันอาจจะประกอบด้วยหลายแคนดิเดตคีย์ (Candidate Key) ทุกแอตทริบิวต์ในรีเลชันจะต้องขึ้นอยู่กับแคนดิเดตคีย์เสมอ เราสามารถกำหนดนิยามของรีเลชันที่อยู่ในรูปของบอยซ์คอดด์นอร์มัลฟอร์ม ก็ต่อเมื่อรีเลชันมีคุณสมบัติตามเงื่อนไขดังต่อไปนี้

1) รีเลชันนั้นเป็นนอร์มัลฟอร์มที่ 3 อยู่แล้ว

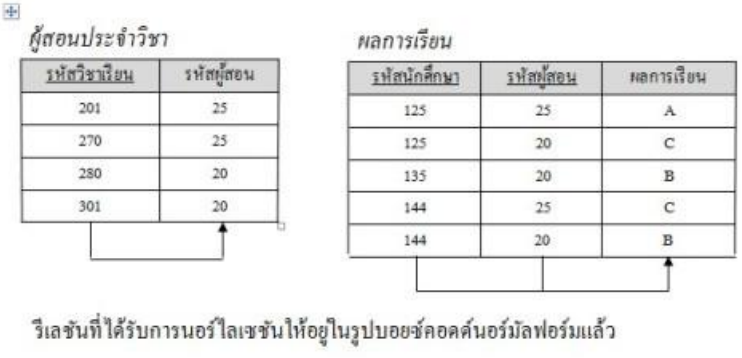
2) ทุกแอตทริบิวต์ในรีเลชันจะต้องขึ้นกับแคนดิเดตคีย์

รีเลชันจะอยู่ในรูปบอยซ์คอดด์นอร์มัลฟอร์ม ถ้าทุกแอตทริบิวต์ขึ้นอยู่กับแคนดิเดตคีย์ (Candidate Key) ดังนั้นถ้าใน 1 รีเลชันมีแคนดิเดตคีย์เพียงตัวเดียวแล้ว นอร์มัลฟอร์มที่ 3 และบอยซ์คอดด์นอร์มัลฟอร์มจะเหมือนกัน โอกาสที่คุณสมบัติของบอยซ์คอดด์นอร์มัลฟอร์มจะถูกละเมิดนั้น เกิดขึ้นได้น้อย และจะเกิดได้กับรีเลชันที่มีแคนดิเดตคีย์มากกว่าหนึ่งเท่านั้น ดังตัวอย่างในภาพข้างล่าง รีเลชันการลงทะเบียนเรียน รีเลชันดังกล่าวอยู่ในรูปนอร์มัลฟอร์มที่ 3 แล้วแต่ก็ยังมีบางส่วนมีปัญหาอยู่ ตรงจุดที่แอตทริบิวต์รหัสวิชาเรียน และผลการเรียนขึ้นอยู่กับคีย์นักศึกษา และคีย์ผู้สอน แต่ในขณะเดียวกันรหัสผู้สอนก็ขึ้นอยู่กับรหัสวิชาเรียน ทำให้ถ้าต้องการเปลี่ยนแปลงผู้สอนในวิชา 301 จะต้องมีการเปลี่ยนแปลงถึง 2 ทัปเพิล ซึ่งผลลัพธ์ที่ได้อาจจะทำให้เกิดความผิดพลาดหากทำการแก้ไขไม่ครบถ้วน และถ้านักศึกษารหัส 135 ถอนการลงทะเบียนวิชา 280 ข้อมูลของผู้ที่สอนวิชานี้จะหายไปจากระบบเลย ถ้าเราลบข้อมูลนี้

ภาพที่ 2.25 รีเลชันที่อยู่ในรูปนอร์มัลฟอร์มที่ 3 แต่ไม่อยู่ในบอยซ์คอดด์นอร์มัลฟอร์ม

ที่มา : http://1ab.in/Cyg

เราสามารถทำการแตกตารางออกมาให้อยู่ในรูปของบอยซ์คอดด์นอร์มัลฟอร์มได้ โดยการแยกแอตทริบิวต์รหัสวิชาเรียนและรหัสผู้สอนซึ่งขึ้นอยู่กับ แอตทริบิวต์รหัสวิชาเรียน ออกมาเป็นอีกหนึ่งรีเลชัน และแยกแอตทริบิวต์ รหัสนักศึกษา รหัสผู้สอน และผลการเรียนออกมาเป็นอีกหนึ่งรีเลชัน ดังแสดงในภาพข้างล่าง



ภาพที่ 2.26 รีเลชันที่ได้รับการนอร์มอไลเซวันให้อยู่ในรูปบอยซ์คอดด์นอร์มัลฟอร์มแล้ว

ที่มา : http://1ab.in/Cyg

5) การแปลงให้อยู่ในรูปนอร์มัลฟอร์มที่ 4 (Fourth Normal Form : 4NF)

ในขณะที่การทำให้อยู่ในรูปของนอร์มัลฟอร์มต่าง ๆ ที่ผ่านมา จะเกี่ยวข้องกับการขึ้นตรงต่อกันของข้อมูลในแต่ละแอตทริบิวต์หรือฟังก์ชันนัลดีเพนเดนซี แต่การทำให้อยู่ในรูปของนอร์มัลฟอร์มที่ 4 จะเกี่ยวข้องกับรูปแบบของการขึ้นตรงต่อกันของข้อมูลในระดับที่ซับซ้อนกว่า

5.1) มัลติแวลูดีเพนเดนซี (Multivalued Dependency)

ถ้าแต่ละแอตทริบิวต์ในหนึ่งรีเลชัน แบ่งออกเป็นกลุ่มของข้อมูลอิสระ เช่นแอตทริบิวต์ X, Y และ Z แบ่งออกเป็นกลุ่มข้อมูลของ X, Y และ Z ที่เป็นอิสระต่อกัน มัลติแวลูดีเพนเดนซี X –>> Y หมายถึงว่าค่า X หนึ่งค่าสามารถที่จะบอกค่า Y ได้หลาย ๆ (X Multi–Determine Y) ไม่ว่า Z จะมีค่าเป็นอะไรก็ตาม

โดยปกติ ถ้า R ประกอบด้วย Attribute X, Y และ Z (Z = R – {XY}) ดังนั้น ถ้า X –>> Y แล้ว X –>> Z เสมอ สามารถเขียนใหม่เป็น X –>> Y | Z ถ้า Y เป็นสับเซทของ X หรือ X ยูเนี่ยน Y = R แล้ว เราเรียก X –>> Y ว่า ทริเวียลมัลติแวลูดีเพนเดนซี (Trivial Multivalued Dependency) ซึ่งจะต่างจากฟังก์ชันนัลดีเพนเดนซี X –> Y ที่ X จะสามารถบอกค่า Y ได้แค่เพียง

ค่าเดียว ดังตัวอย่างภาพข้างล่าง เนื่องจากแอตทริบิวต์ รหัสโครงการ รหัสบริษัท และที่ตั้งโครงการล้วนเป็นคีย์แอตทริบิวต์ ดังนั้นรีเลชันในภาพ จึงถือว่าอยู่ในรูป BCNF แล้ว แต่ยังไม่อยู่ในรูปของ 4NF เนื่องจากรีเลชันดังกล่าวยังมีทริเวียลมัลติแวลูดีเพนเดนซีอยู่ในรีเลชัน ตัวอย่างเช่น รหัสโครงการ A001 สามารถบอกค่าของรหัสบริษัทที่เป็นผู้รับผิดชอบได้มากกว่าหนึ่งบริษัท คือ รหัสบริษัท B001 และ B002 ในขณะเดียวกันรหัสโครงการ A001 ก็บอกถึงที่ตั้งของโครงการสองแห่งคือ จันทบุรี และระยอง ซึ่งถ้ามีการเพิ่มบริษัทที่รับผิดชอบโครงการเข้าไปในโครงการ A001 อีกหนึ่งบริษัทก็จะต้องมีการเพิ่มข้อมูลถึงสองทัปเพิลเนื่องจากโครงการดังกล่าวมีที่ตั้งอยู่ถึงสองแห่งคือ ระยอง และจันทบุรี ส่งผลให้เกิดความซ้ำซ้อนของข้อมูลขึ้นในรีเลชันดังกล่าว และอาจจะเกิดความผิดพลาดในการเพิ่มข้อมูลได้ เนื่องจากที่ตั้งโครงการไม่ได้ขึ้นอยู่กับรหัสบริษัทที่เป็นผู้รับผิดชอบแต่ขึ้นอยู่กับรหัสโครงการ ดังนั้น ถ้าหากมีการเพิ่มบริษัทผู้รับผิดชอบเพิ่มขึ้นอีกหนึ่งบริษัท เราจำเป็นที่จะต้องทำการเพิ่มข้อมูลที่ตั้งโครงการเข้าไปอีกสองแห่งด้วยเสมอ ซึ่งเป็นผลจากความสัมพันธ์ในรูปแบบของ ทริเวียลมัลติแวลูดีเพนเดนซี

โครงการก่อสร้าง

รหัสโครงการ	รหัสบริษัท	ที่ตั้งโครงการ
A001	B001	จันทบุรี
A001	B001	ระยอง
A001	B002	จันทบุรี
A001	B002	ระยอง

รีเลชันที่อยู่ในรูปบอยซ์คอดด์นอร์มัลฟอร์มแล้ว แต่ยังไม่อยู่ในรูปของนอร์มัลฟอร์มที่ 4

ภาพที่ 2.27 รีเลชันอยู่ในรูปบอยซ์คอดด์นอร์มัลฟอร์มแล้ว แต่ยังไม่อยู่ในรูปของนอร์มัลฟอร์มที่ 4

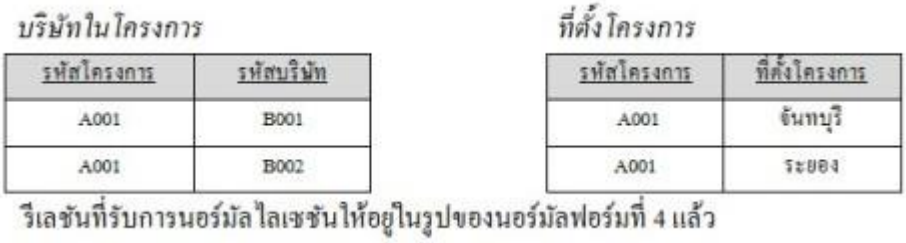
ที่มา : http://1ab.in/Cyg

5.2) นิยามของนอร์มัลฟอร์มที่ 4 รีเลชันจะอยู่ในรูปของนอร์มัลฟอร์มที่ 4 ก็ต่อเมื่อมีคุณสมบัติตามเงื่อนไขดังต่อไปนี้

1) รีเลชันนั้นเป็นบอยซ์คอดด์นอร์มัลฟอร์มอยู่แล้ว

2) รีเลชันนั้นไม่มีทริเวียลมัลติแวลูดีเพนเดนซี

จากรีเลชันในภาพข้างบน เราสามารถขจัดทริเวียลมัลติแวลูดีเพนเดนซี โดยการแตกรีเลชันดังกล่าวออกเป็นรีเลชันย่อย 2 รีเลชัน ซึ่งจะทำให้ทั้งสองรีเลชันอยู่ในรูปของนอร์มัลฟอร์มที่ 4 ดังภาพข้างล่าง



ภาพที่ 2.28 รีเลชันที่รับการนอร์มัลไลเซชันให้อยู่ในรูปของนอร์มัลฟอร์มที่ 4 แล้ว

ที่มา : http://1ab.in/Cyg

6) การแปลงให้อยู่ในรูปนอร์มัลฟอร์มที่ 5 (Fifth Normal Form: 5NF)

การแปลงให้อยู่ในรูปของนอร์มัลฟอร์มที่ 5 จะพิจารณาถึงการขึ้นต่อกันของข้อมูลในการแยกข้อมูลในรีเลชันออกเป็นรีเลชันย่อย และประกอบรีเลชันย่อยกลับเป็นรีเลชันใหญ่เช่นเดิม ซึ่งเป็นการตรวจสอบว่าเมื่อรวมกันใหม่ด้วยวิธีการจอยน์แล้ว จะได้รีเลชันกลับมาเหมือนเดิมทุกประการหรือไม่

6.1) จอยน์โอเปอรชัน (Join Operation)

ถ้ามี R1(X,Y) และ R2(Y,Z) R1 JOIN R2 = R3(X, Y, Z) โดยที่ t(x, y, z) อยู่ใน R3 ก็ต่อเมื่อมี t1(x, y) อยู่ใน R1 และ t2(y, z) อยู่ใน R2

6.2) จอยน์ดีเพนเดนซี (Join Dependency)

ในการแยกรีเลชันออกเป็นส่วนย่อย(Decomposition) R1, R2, R3, Rn มีคุณสมบัติจอยน์ดีเพนเดนซี ก็ต่อเมื่อ R1 JOIN R2 JOIN R3 ... JOIN Rn = R นั่นคือเมื่อเอารีเลชันย่อยมารวมกันก็ต้องได้รีเลชันเดิม ที่ไม่มีข้อมูลสูญหาย และไม่มีทัปเพิลที่เกินมา ที่เรียกว่าสพิวเรียสทัปเพิล (Spurious Tuple)

6.3) นิยามของ 5NF รีเลชันจะเป็น 5NF ถ้า

1) รีเลชันนั้นเป็นนอร์มัลฟอร์มที่ 4 อยู่แล้ว

2) การแบ่งแยกรีเลชันมีคุณสมบัติจอยน์ดีเพนเดนซี

จากตัวอย่างในภาพข้างล่าง รีเลชัน วิชาเรียนประจำภาคอยู่ในรูปของนอร์มัลฟอร์มที่ 4 แล้ว เนื่องจากแอตทริบิวต์ภาคการศึกษาเป็นตัวกำหนดแอตทริบิวต์รหัสวิชาหลายค่าในขณะที่แอตทริบิวต์รหัสวิชา ก็เป็นตัวกำหนดแอตทริบิวต์รหัสชั้นเรียนหลายค่า รีเลชันนี้จึงไม่มีทริเวียลมัลติแวลูดีเพนเดนซี ต่อไปเราจึงทำการทดสอบคุณสมบัตินอร์มัลฟอร์มที่ 5 ของรีเลชันวิชาเรียนประจำภาค โดยเมื่อนำรีเลชันดังกล่าวมาทำการแตกย่อยออกเป็นสามรีเลชันคือ รีเลชันภาคการศึกษา รีเลชันวิชาเรียนของชั้นเรียน และ รีเลชันชั้นเรียนประจำภาค และทำการจอยน์ทั้งสามรีเลชันรวมกลับเป็นหนึ่งรีเลชันอีกครั้ง จะได้จำนวนข้อมูลเท่ากันกับรีเลชันก่อนที่จะมีแตกเป็นรีเลชันย่อยทุกประการ ซึ่งก็คือรีเลชันดังกล่าวมีคุณสมบัติจอยน์ดีเพนเดนซีและอยู่ในรูปของนอร์มัลฟอร์มที่ 5 แล้ว

ภาพที่ 2.29 นอร์มัลฟอร์มที่ 5

ที่มา : http://1ab.in/Cyg

2.3 เครื่องมือในการออกแบบ และวิเคราะห์ข้อมูล

2.3.1 แบบจำลองอนุกรมเวลา (Time series analysis)

2.3.1.1 ส่วนประกอบของอนุกรมเวลา

1) ค่าแนวโน้ม (Long Term Trend: T) ค่าแนวโน้มเป็นการแสดงถึงการเคลื่อนไหวหรือเปลี่ยนแปลงของข้อมูลในระยะยาว เช่น ปริมาณการใช้ไฟฟ้าของประเทศไทย, ปริมาณการนำเข้าน้ำมันดิบ เป็นต้น

2) ค่าการผันแปรตามฤดูกาล (Seasonal Variation: S) หมายถึงการเปลี่ยนแปลงตามฤดูกาล โดยเกิดขึ้นซ้ำ ๆ กันในรอบ 1 ปี จนกลายเป็นแบบแผนเดียวกัน เช่น ผลผลิตข้าวจะสูงในช่วงไตรมาสแรกของปี ยอดขายของห้างสรรพสินค้าจะสูงในช่วงปลายปี เป็นต้น ในการวิเคราะห์การผันแปรตามฤดูกาลนี้จะวัดออกมาในรูปของดัชนีฤดูกาล (Seasonal Index)

3) ค่าการผันแปรตามวัฏจักร (Cyclical Variation: C) หมายถึงการเคลื่อนไหวที่เป็นไปตามวัฏจักร ซึ่งการเคลื่อนไหวตามวัฏจักรนี้จะมีลักษณะคล้ายกับการผันแปรตามฤดูกาล แต่จะมีระยะเวลาที่ยาวนานกว่า

4) การผันแปรเนื่องจากเหตุการณ์ไม่ปกติ (Irregular Variation: I) การผันแปรชนิดนี้ไม่แน่นอน ไม่สามารถคาดการณ์ได้ล่วงหน้า เช่น ภัยธรรมชาติ สงคราม การนัดหยุดงาน เป็นต้น

ข้อมูลอนุกรมเวลา อาจได้รับอิทธิพลของปัจจัยที่เป็นส่วนประกอบของอนุกรมเวลาทั้ง 4 ปัจจัยหรือเพียงปัจจัยใดปัจจัยหนึ่งเท่านั้น การวิเคราะห์จึงควรแยกวิเคราะห์ทีละปัจจัย ซึ่งในที่นี้จะกล่าวถึงการวิเคราะห์ ปัจจัยค่าแนวโน้ม และค่าผันแปรตามฤดูกาลเท่านั้น

2.3.1.2 วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (Exponential smoothing)

วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (Exponential smoothing) เป็นวิธีพยากรณ์ที่เหมาะสมกบัการพยากรณ์ในระยะสั้นและปานกลางวิธีการนี้ได้ให้ความสำคัญกับข้อมูลล่าสุด และความสำคัญของข้อมูลที่ห่างออกไปจะลดลง วิธีการ Exponential smoothing ที่นิยมในปัจจุบันมีอยู่ 3 วิธีดังนี้

1) Single Exponential Smoothing (SES)

เป็นวิธีการทำให้เรียบอย่างง่าย โดยใช้การหาค่าเฉลี่ยแบบถ่วงน้ำหนัก และสมมติให้ค่าน้ำหนักหรือค่าความสำคัญของข้อมูล คือ α (Alpha) วิธีการนี้มีเงื่อนไขว่า ข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์จะต้องไม่มีแนวโน้ม (Trend) และอิทธิพลของฤดูกาล (Seasonality) หมายความว่าข้อมูลอนุกรมเวลาที่นำมาใช้การวิเคราะห์ต้องมีลักษณะคงที่ สำหรับสมการที่ใช้ในการพยากรณ์สามารถแสดงได้ดังนี้

$$\hat{y}_t = \alpha\, y_1 + (1+\alpha)\,\hat{y}_{t-1}$$

หรือ

$$\hat{y}_t = \alpha \sum_{n=0}^{i-1}(1+\alpha)^n y_{t-n}$$

โดยที่ y_t = ข้อมูล ณ เวลาที่ t; t = 1, 2,..., n

α = ค่าถ่วงน้ำหนักความสำคัญที่ให้แก่ข้อมูล ณ เวลาที่ t ($0 \le \alpha \le 1$)

$\hat{y}_t$ = ค่าประมาณหรือพยากรณ์ของขอ้มูล ณ เวลาที่ t

$\hat{y}_{t-1}$ = ค่าประมาณหรือค่าพยากรณ์ของขอ้มูล ณ เวลา t–1

2) Holt's Two–Parameter Method

วิธี Holt's Two–Parameter Method เป็นวิธีที่ Holt (1957) ได้ปรับปรุงวิธี Single Exponential Smoothing (SES) –7hoใหม่เพื่อให้สามารถใช้ขอ้ มูลอนุกรมเวลาที่มีแนวโน้ม (trend) ของเวลา เรียกวิธีการนี้ว่า "Holt's Two–Parameter Method" วิธีการนี้ให้ความสำคัญกับข้อมูลล่าสุดและแนวโน้มเวลา ดังนั้นจึงมีค่าคงที่ในการทำให้เรียบ 2 ค่าคือ α (Alpha) และ β (Beta) โดยมีสมการที่ใช้ในการพยากรณ์ดังนี้

$\hat{y}_{t+k} = \alpha + bK$; $\hat{y}_{t+k}$ = ค่าพยากรณ์ณ เวลาที่ t +k

ค่า a และ b คำนวณจาก

$$a_t = \alpha\, y_t + (1+\alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta\,(a_t - a_{t-1}) + 1 - Bb_{t-1}$$

โดยที่

α = ค่าคงที่ที่ทำให้เรียบระหว่างขอ้มูลกับพยากรณ์ ($0 \le \alpha \le 1$)

β = ค่าคงที่ที่ทา ให้เรียบระหวา่ งแนวโนม้จริงกบัค่าประมาณของแนวโน้ม

($0 \le \beta \le 1$)

3) Holt–Winters–Trend and Seasonal (Three–Parameter) วิธีนี้ถูกพัฒนาเพิ่ม ขึ้นจากวิธีการของ Holtโดย Winters (1960) ได้พัฒนาให้วิธีการนี้สามารถวิเคราะห์ข้อมูลอนุกรมเวลาที่มีทั้งแนวโน้ม (Trend) และฤดูกาล (Seasonality) วิธีการนี้ให้ความสำคัญกับข้อมูลล่าสุดแนวโน้มเวลาและฤดูกาล ดังนั้นจึงมีค่าคงที่ในการทำให้เรียบ 3 ค่าคือ (Alpha), (Beta) และ (Gamma) โดยมีสมการที่ใช้ในการพยากรณ์ดังนี้

Holt – Wintes – Mulitplicative

$\hat{y}_{t+k} = (a + bk)c_{t+k}$; $\hat{y}_{t+k}$ = ค่าพยากรณ์ณ เวลาที่ t + k

ค่า a,b และ c คำนวณจาก

$$a_t = \alpha \frac{y_t}{c_{t-s}} + (1 + \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta\,(a_t - a_{t-1}) + (1 - B)b_{t-1}$$

$$c_t = \gamma \frac{y_t}{a_t} + (1 - \gamma)\, c_{t-1}$$

โดยที่ α = ค่าคงที่ที่ทำให้เรียบระหว่างข้อมูลกับพยากรณ์ ($0 \le \alpha \le 1$)

β = ค่าคงที่ที่ทำให้เรียบระหว่างแนวโนม้จริงกับค่าประมาณของแนวโน้ม ($0 \le \beta \le 1$)

γ = ค่าคงที่ที่ทำให้เรียบระหว่างฤดูกาลจริงกับค่าประมาณของฤดูกาล

($0 \le \gamma \le 1$)

Holt – Wintes – Additive

$\hat{y}_{t+k} = (a+ bk)c_{t+k}$; $\hat{y}_{t+k}$ = ค่าพยากรณ์ณ เวลาที่ t +k

ค่า a,b และ c คำนวณจาก

$a_t = \alpha (y_{t-1} - c_{t-1}) + (1-\alpha)(a_{t-1} + b_{t-1})$

$b_t = \beta (a_t - a_{t-1}) +1 - Bb_{t-1}$

$c_t = \gamma (y_t - a_{t+1}) - \gamma c_{t-1}$

โดยที่ α = ค่าคงที่ที่ทำให้เรียบระหว่างข้อมูลกับพยากรณ์ ($0 \le \alpha \le 1$)

β = ค่าคงที่ที่ทำให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้ม ($0 \le \beta \le 1$)

γ = ค่าคงที่ที่ทำให้เรียบระหว่างฤดูกาลจริงกับค่าประมาณของฤดูกาล ($0 \le \gamma \le 1$)

2.3.1.3 การประเมินค่าการพยากรณ์

การศึกษาครั้งนี้ใช้ค่าสถิติ RMSE (Root Mean Squared Error) และ MAPE (Mean Absolute Percentage Error) ในการประเมินค่าการพยากรณ์เพื่อทดสอบความแม่นยำของผลการพยากรณ์

Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\sum_{t=T+1}^{T+h} \left(\hat{Y}_t - Y_t\right)^2 / h}$$

Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = 100 \sum_{t=T+1}^{T+h} \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right| / h$$

ภาพที่ 2.30 การประเมินค่าการพยากรณ์

ที่มา : https://1th.me/sMm8I

2.3.2 การแสดงผลแบบ (visualization)

รูปแบบของจินตทัศน์ (Visualization) รูปแบบของจินตทัศน์ที่ใช้ในการแสดงผลสารสนเทศนั้นมีจำนวนมาก ซึ่งการเลือกใช้ก็ขึ้นอยู่กับสารสนเทศที่ต้องการแสดง รวมถึงผลลัพธ์ที่ต้องการให้เกิดการตอบสนองต่อ สารสนเทศนั้น ๆ ในงานวิจัยนี้มุ่งศึกษาแนวทางการ ออกแบบแผงหน้าปัดโดยใช้ซอฟต์แวร์ IBM Cognos ดังนั้นจึงมีรูปแบบแผนภูมิที่สามารถสร้างบนซอฟต์แวร์ IBM Cognos ตามที่ได้อธิบายไว้ในคู่มือการใช้งาน IBM Cognos Business Intelligence V10.1 Handbook ดังนี้

1) แผนภูมิแท่ง (Column and Bar Chart) แผนภูมิแท่งใช้ในการแสดงข้อมูลตามช่วงเวลาหรือข้อมูลที่ไม่ต่อเนื่องกัน โดยมีทั้งแผนภูมิแท่งแนวตั้งและแนวนอน ซึ่งสามารถเปรียบเทียบข้อมูลได้ทั้งข้อมูลเดียว หรือหลายข้อมูล เหมาะต่อการดูข้อมูลที่เรียงลำดับตาม ค่าต่ำสุดหรือสูงสุด นอกจากนี้แผนภูมิแท่ง สามารถแสดงข้อมูลหลาย ๆ ข้อมูลเรียงกันเป็นชั้น ภายในแผนภูมิแท่งเดียวกันเป็นแผนภูมิแท่งแบบชั้น (Stack Charts) (Browne et al.,2010) ซึ่งเป็นรูปแบบที่เหมาะกับการแสดงข้อมูลที่มีความสัมพันธ์เป็นชุด เพื่อให้ค่าที่แสดงสามารถ เปรียบเทียบกันได้ง่ายเมื่อมีค่าข้อมูลที่ต้องเปรียบเทียบแสดงอยู่ด้วยกัน อาจใช้ในการแสดง ข้อมูลตามช่วงเวลาสั้น ๆ ข้อมูลการขายตามหมวดหมู่สินค้า (Kerzner , 2011) สามารถใช้การวัดหลายอย่างได้ภายในแผนภูมิ เดียวกันโดยที่ไม่ทำให้ข้อมูลดูหนาแน่นเกินไปสามารถใช้ในการ แจ้งเตือนผู้ใช้แผงหน้าปัดถึงค่าที่ผิดปกติไปจากเป้าหมายที่วางไว้ ข้อควรระวังคือ ต้องมั่นใจในความถูกต้องของข้อมูล เนื่องจากถ้าค่าข้อมูลผิดก็จะทำให้แผนภูมิแสดงความสัมพันธ์ และแนวโน้มที่ผิด (Smietana, 2010)

2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้นมีความคล้ายคลึงกับแผนภูมิแท่งเพียงแต่จะใช้จุดเป็นตัวบอกค่าของข้อมูลแทนตัวแท่ง แล้วมีการลากเส้นเชื่อมแต่ละจุด จึงเหมาะสำหรับการแสดงข้อมูล แนวโน้มหลาย ๆ แนวโน้มเปรียบเทียบกัน แต่ก็ไม่ควรใช้ในการเปรียบเทียบข้อมูลหลาย ๆ ค่า เพราะจะแยกความแตกต่างระหว่างข้อมูลได้ยากกว่าการแสดงแบบเป็นชั้นในแผนภูมิแท่ง (Browne et al., 2010) เหมาะสำหรับการวิเคราะห์อนุกรมเวลาที่ต้องการที่จะเห็นการเปลี่ยนแปลงที่เกิดขึ้นของข้อมูล หนึ่งหรือมากกว่าหนึ่งข้อมูลการวัด ค่าในช่วงเวลา แผนภูมิเส้น

ยังให้การเปรียบเทียบการวิเคราะห์แนวโน้มใช้ข้อมูลหลายชุดซ้อนกันในหนึ่งแผนภูมิ (Kerzner, 2011) ในแผนภูมิเส้นสามารถใช้จุดบอกตำแหน่งของข้อมูลที่ชัดเจนได้ เนื่องจากถ้าใช้แต่เส้นก็จะทราบเฉพาะแนวโน้มของข้อมูล หรือการใช้จุดอย่างเดียว ก็จะเห็นแต่ข้อมูล ณ ตำแหน่งนั้น ๆ แต่จะไม่เห็นแนวโน้ม ปัจจุบันจึงมีการใช้แผนภูมิเส้น และ จุด (Spline) ด้วยกัน เพื่อให้เห็นทั้งค่าข้อมูลและแนวโน้ม (Smietana, 2010) ควรทำเส้นให้จุด ของข้อมูลเด่นว่ามีข้อมูลที่ตำแหน่งใด (Juice Inc., 2009)

3) แผนภูมิพื้นที่ (Area Chart) แผนภูมิพื้นที่สามารถใช้สีในบริเวณพื้นที่ใต้กราฟได้ แทนการลากเส้น เพื่อแสดงแนวโน้มของข้อมูลเพื่อใช้ในการเปรียบเทียบ (Browne et al.,2010) การใช้สีเพื่อแสดงพื้นที่ใต้หรือเหนือกราฟของแผนภูมิเส้นใช้เมื่อต้องการเปรียบเทียบแนวโน้ม ตามช่วงเวลาของข้อมูล (Kerzner, 2011)

4) แผนภูมิจุด (Point Charts) แผนภูมิจุดมีลักษณะคล้ายแผนภูมิเส้นเพียงแต่แสดงเฉพาะตำแหน่งค่าข้อมูล โดยไม่มีการลากเส้นเพื่อแสดงแนวโน้ม แผนภูมิจุดมีประโยชน์สำหรับการแสดงข้อมูลเชิงปริมาณ โดยใช้จุดแทนข้อมูลตามแกนโดยวางตามลำดับ (Browne et al., 2010)

5) แผนภูมิแบบผสม (Combination Charts) แผนภูมิแบบผสมใช้ในการแสดงค่าข้อมูลหลายค่าโดยใช้แผนภูมิแท่ง แผนภูมิพื้นที่ หรือแผนภูมิเส้น รวมไว้ในแผนภูมิเดียวกัน มีประโยชน์ในการเน้นความสัมพันธ์ของข้อมูลแต่ละชุด (Browne et al., 2010)

6) แผนภูมิแบบกระจาย (Scatter Plot) แผนภูมิแบบกระจายมักใช้ในการเปรียบเทียบความสัมพันธ์ของข้อมูลระหว่างชุดเพื่อดูความแตกต่างตามตำแหน่งของข้อมูลที่ปรากฏบนแผนภูมิ (Browne et al., 13 2010) ใช้ในกรณีที่ต้องการดูความสัมพันธ์ของข้อมูล มากกว่าค่าของข้อมูลจะเป็นการดูแนวโน้มใน ภาพรวมว่าข้อมูลชุดนี้มีค่าไปในทิศทางใด (Smietana, 2010)

7) แผนภูมิฟอง (Bubble Charts) แผนภูมิฟองมีลักษณะคล้ายแผนภูมิแบบกระจายเพียงแต่จะมีตัววัดเพิ่มขึ้นโดยเป็นขนาดของฟองตามค่าของข้อมูล โดยมักใช้ในการแสดงข้อมูลด้านการเงิน (Browne et al., 2010) แสดงค่าเป็นกลุ่มของข้อมูลด้วยขนาดของฟอง (วงกลม หรือ ทรงกลม) ที่แตกต่างกันตามค่าของข้อมูล ใช้ในการเปรียบเทียบระหว่างข้อมูลเดียวกันแต่

ต่างช่วงเวลา หรือเปรียบเทียบ ระหว่างข้อมูลที่แตกต่างกันด้วย (Kerzner,2011) การเลือกใช้ควรแน่ใจในเรื่องการอธิบายขนาดของฟองสบู่แต่ละฟอง ว่ามีการกำหนดค่าได้ถูกต้องและ ชัดเจน (Smietana, 2010)

8) แผนภูมิสี่ส่วน (Quadrant Charts) โดยหลักแล้วจะใช้แผนภูมิสี่ส่วนเป็นแกนพื้น หลังให้แผนภูมิฟอง ซึ่งจะแบ่งพื้นที่เป็นสี่ส่วนเท่า ๆ กัน ในการแบ่งส่วนของแผนภูมิอาจใช้สีในการแยกความแตกต่างของแต่ละส่วน ซึ่งขนาดของส่วนที่แบ่งอาจแบ่งย่อยลงไปได้อีกเสมือน เป็นมาตรวัดในแผนภูมิ สามารถใช้แผนภูมิสี่ส่วนในการแบ่งหมวดหมู่เพื่อวางข้อมูลตามหมวดหมู่เช่นการวิเคราะห์จุดอ่อนจุดแข็งโอกาส และอุปสรรค (SWOT Analysis) เป็นต้น (Browne et al., 2010)

9) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมใช้ในการแสดงความสัมพันธ์ของ ข้อมูลในส่วนนั้น เปรียบเทียบกับสัดส่วนทั้งหมดของแผนภูมิ ซึ่งหากต้องการแสดงข้อมูลจริง เป็นแผนภูมิแท่งแบบเป็น ชั้นมากกว่าแผนภูมิวงกลมไม่ใช่ตัวเลือกที่ดีในการแสดงข้อมูลที่มีค่า เป็นศูนย์หรือมีค่าติดลบ (Browne et al., 2010) เป็นแผนภูมิที่มีองค์ประกอบมากกว่าแผนภูมิ แบบอื่น โดยเป็นเรื่องยากมากที่จะแยกแยะความแตกต่างของสัดส่วนของข้อมูลกับการแบ่ง พื้นที่ตามค่าข้อมูล นอกจากจะเป็นกรณีของชุดข้อมูลขนาดเล็กที่มีความแตกต่างของค่าข้อมูลอยู่มาก และแผนภูมิวงกลมที่ยังก่อให้เกิดปัญหาสำหรับการใส่คำอธิบายเพราะมีทั้งสีของ แผนภูมิและรายละเอียดต่าง ๆ ภายใน ซึ่งการใส่คำอธิบายอื่นเพิ่มอาจก่อให้เกิดการสับสนในการดูข้อมูล (Kerzner, 2011) แผนภูมวิงกลมเป็นแผนภูมิสำหรับการแสดงองค์ประกอบข้อมูลที่ มีส่วนประกอบค่อนข้างน้อยและมีเพียงหนึ่งตัวชี้วัดเชิงปริมาณ และการแสดงแผนภูมิวงกลม ต้องคิดค่าข้อมูลรวมเป็น 100% และข้อมูลที่ไม่เป็นค่าลบ ใช้การแสดงค่าข้อมูลทั้งหมดด้านนอกแผนภูมิ โดยเรียงค่าข้อมูลตามลำดับที่แสดงในแผนภูมิ (Smietana, 2010)

10) แผนภูมิจุดนำ (Bullet Charts) แผนภูมิจุดนำเป็นรูปแบบหนึ่งของแผนภูมิแท่ง เป็นการแสดงตัว วัดที่สำคัญ ใช้ในการเปรียบเทียบตัววัดตั้งแต่ 1 ตัวหรือมากกว่านั้นโดยสามารถวัดโดยเปรียบเทียบ กับสีของพื้นหลังที่จะแสดงตัววัดที่มากขึ้น เช่น ระดับความพึงพอใจที่ดี จนถึงระดับที่ไม่ดี และเนื่องจากเป็นแสดงสารสนเทศที่ไม่จำเป็นต้องใช้พื้นที่มากจึงสามารถใส่แผนภูมิจุด

นำเป็นส่วนหนึ่งในการแสดงรายงานบนแผงหน้าปัด (Browne et al.,2010) จะแสดงการวัดมูลค่าในแกนแนวนอน ซึ่งอาจมีการเพิ่มบริบทข้องข้อมูลเข้าไปโดยการใช้สีเป็นตัวบอกระดับความพึงพอใจ ทำให้มองข้อมูลเข้าใจได้ในครั้งเดียว (Smetana,2010)

11) แผนภูมิมาตรวัด (Gauge Charts) แผนภูมิมาตรวัดหรืออาจเรียกว่าแผนภูมิหน้าปัด (Dial Charts) หรือแผนภูมิมาตรอัตราเร็ว (Speedometer Charts) มีความคล้ายกับแผนภูมิจุดนำในแง่การเปรียบเทียบตัววัดหลายค่าเพียงแต่ใช้เข็มในการแสดงค่า ซึ่งการอ่านค่า สามารถทำได้ง่ายเหมือนการอ่านค่าจากหน้าปัดและแต่ละค่าที่แสดงจะเปรียบเทียบด้วยช่วงสีที่อยู่บนแผนภูมิ โดยแผนภูมิมาตรวัดนั้นเป็นทางเลือกที่ดีกว่าแผนภูมิจุดนำเมื่อต้องเปรียบเทียบ ค่ามากกว่าสองค่าขึ้นไป โดยแผนภูมินี้มักถูกใช้ในการแสดงตัวชี้วัดผลการดำเนินงานหลักบน แผงหน้าปัดสำหรับผู้บริหารระดับสูง สำหรับการออกรายงานโดยใช้ PDF และ HTML จะจำกัดอยู่ที่แผนภูมิ และใน Microsoft Excel จะไม่สนับสนุนการแสดงผลแผนภูมิประเภทนี้ (Browne et al., 2010) ใช้ในการบอกสถานะปัจจุบันโดยจะใช้สีแดง ส้มเหลือง เขียว ในการบอกถึงระดับของปัญหาตั้งแต่มาก (สีแดง) จนถึงระดับปกติ (สีเขียว) สามารถช่วยให้เห็นความสัมพันธ์ ระหว่างค่าที่วัด แต่ว่าจะค่อนข้างสิ้นเปลืองเนื้อที่ จึงควรเลือกใช้เมื่อมีพื้นที่มากเพียงพอ (Smetana, 2010)

12) แผนภูมิพาเรโต (Pareto Charts) แผนภูมิพาเรโตเป็นลำดับของหมวดหมู่ที่เรียง จากความถี่มากที่สุดไปจนถึงน้อยที่สุด โดยจะมีเส้นการสั่งสม (Cumulation Line) ซึ่งเป็นการแสดงอัตราส่วนร้อยละยอดรวมสะสมทั้งหมดของแผนภูมิแท่ง โดยแผนภูมิชนิดนี้จะใช้ในการ ควบคุมคุณภาพ ทำให้สามารถระบุและลดสาเหตุของปัญหาที่เกิดขึ้นตามลำดับความสำคัญ ของปัญหา และสามารถที่จะสร้างแผนภูมิพาเรโตได้ทั้งก่อนและหลังการแก้ปัญหาเพื่อเป็นการ เปรียบเทียบให้เห็นถึงส่งที่เปลี่ยนแปลงไป (Browne et al., 2010)

13) แผนภูมิแท่งแบบก้าวหน้าหรือแผนภูมิน้ำตก (Progressive Column Charts or Waterfall Charts) เป็นแผนภูมิที่แสดงการเปลี่ยนแปลงที่เกิดขึ้นต่อจากแผนภูมิแท่ง หรือแผนภูมิแท่งแบบเป็นชั้น โดยมักใช้ในการเปรียบเทียบค่าข้อมูล 1 ข้อมูลต่อข้อมูลทั้งหมด โดยใน Microsoft Excel ไม่สนับสนุนการแสดงผลแผนภูมิประเภทนี้ (Browne et al., 2010)

14) แผนภูมิมาริเมกโก (Marimekko Charts) เป็นแผนภูมิชั้นซึ่งความกว้างของชั้น ของแท่งแผนภูมิจะเป็น สัดส่วนของข้อมูล 1 ข้อมูลต่อข้อมูลทั้งหมด โดยส่วนที่เป็นความสูงของ แต่ละแนวแท่งแผนภูมินั้นจะเป็นอัตราร้อยละของมูลค่ารวมแนวตามลำดับ (Browneet al., 2010)

15) แผนภูมิเรดาร์หรือแผนภูมิแมงมุม (Radar or Spider Charts) แผนภูมิเรดาร์เป็นแผนภูมิที่ใช้ในการเปรียบเทียบข้อมูลหลายค่าตามแกนหลายแกนโดยจะใส่ค่าข้อมูลโดยเริ่มจากตรงกลางที่มีแกนวัดค่าเป็นรัศมี มีประโยชน์ในการเปรียบเทียบข้อมูลหลายชุดและหลายตัววัดและมีประโยชน์ในการดูค่าผิดปกติ (Browne et al., 2010) ข้อมูลจะถูกใส่ค่าตามระยะห่างจากจุดศูนย์กลาง แต่ละจุดข้อมูลที่มีส่วนประกอบในแนวตั้ง คือระยะห่างจากจุดศูนย์กลางของแกนแผนภูมิเรดาร์มีประโยชน์สำหรับการแสดงข้อมูลวัฏจักร เช่นระดับการรับพนักงานรายวันหรือผลรวมรายได้รายเดือนในการออกแบบสามารถวางซ้อนเป้าหมายและเกณฑ์การให้บริบทสำหรับตัวชี้วัดเชิงปริมาณ (Smetana, 2010)

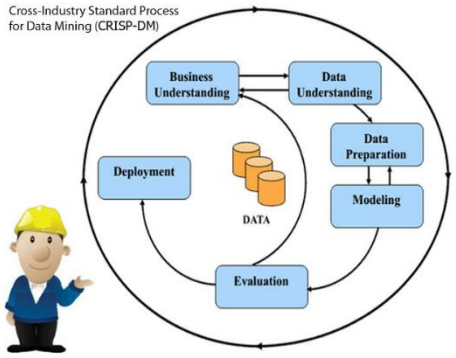
16) แผนภูมิผลได้เสีย (Win–Loss Charts) เป็นแผนภูมิขนาดเล็ก (Micro charts) ที่ใช้ในการวัดค่าที่ตั้งอัตโนมัติ (Default Measure) และค่าได้หรือเสียจากค่าที่ตั้งไว้ ซึ่งแผนภูมิผลได้เสียนั้นเป็นการแสดงผลของเรื่องที่กำหนด และสามารถกำหนดค่าที่ตั้งอัตโนมัติได้เช่นกัน ซึ่งสามารถใช้แผนภูมินี้ในการแสดงจินตทัศน์สำหรับแนวโน้มผลการดำเนินงาน (Browne et al., 2010)

17) แผนภูมิขั้วโลก (Polar Charts) เป็นแผนภูมิวงกลมที่ใช้ค่าข้อมูลและขนาดของมุมเพื่อแสดงสารสนเทศเป็นลักษณะขั้วพิกัด (Browne et al., 2010)

18) เส้นฐานและเส้นแนวโน้ม (Baselines and Trend lines) เส้นฐานและเส้นแนวโน้มนั้นจะแสดงเป็นข้อมูลรายละเอียดเพิ่มเติมบนแผนภูมิเส้นฐานนั้นสามารถอยู่ได้ทั้งแนวตั้ง และแนวนอนที่ตัดผ่านแผนภูมิเพื่อระบุถึงส่วนที่สำคัญของข้อมูลสำหรับเส้นแนวโน้มนั้นใช้เป็น แนวโน้มของข้อมูลเมื่อต้องการพยากรณ์ค่า โดยส่วนมากเส้นแนวโน้มจะเป็นเส้นตรงหรือเส้นโค้ง ที่ลากผ่านจุดอย่างน้อย 2 จุด เพื่อแสดงแนวโน้มซึ่งสามารถใช้ในแผนภูมิแท่ง แผนภูมิเส้น แผนภูมิพื้นที่ แผนภูมิฟอง และแผนภูมิแบบกระจาย

2.3.3 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)

งานวิทยาการด้านข้อมูล (Data Science) ซึ่งกำลังมีบทบาทอย่างมากในโลกยุคปัจจุบันและทวีความสำคัญยิ่งขึ้นในอนาคต ก็มี CRISP–DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP–DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่



ภาพที่ 2.31 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP–DM)

ที่มา : https://www.iok2u.com/index.php/article/innovation/444–cross–industry–standard–process–for–data–mining–crisp–dm

1) ทำความเข้าใจระบุโอกาสหรือหาปัญหาที่จะเกิดขึ้นกับธุรกิจ (Business Understanding)กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดเพื่อนำมาแก้ไขปัญหาองค์กร และต้องสามารถระบุผลลัพธ์ที่มีได้

2) ทำความเข้าใจข้อมูล (Data Understander) โดยการรวบรวมข้อมูลที่เกี่ยวข้องให้อยู่รวมกลุ่มกัน คัดเลือกให้เหลือเพียงข้อมูลที่มีความสำคัญต่อการทำงาน

3) เตรียมข้อมูล (Data Preparation) ขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดใน 6 ขั้นตอน เพราะคุณภาพของงานที่ได้นั้นจะดีเพียงใด ต้องขึ้นอยู่กับคุณภาพข้อมูลที่ได้จัดเตรียมในขั้นตอนนี้ การเตรียมข้อมูลประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) สร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ถ่ายทอดข้อมูลหรือทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่นำไปใช้ประโยชน์ได้จริงในทางธุรกิจ

5) การประเมินผล (Evaluation) เพื่อเป็นการทดสอบผลที่ได้โดยการวัดประสิทธิภาพจากผลลัพธ์ที่ได้มาจัดทำงานจริงและติดตามตรวจสอบผลที่ได้ใหม่ ซึ่งสามารถตรวจสอบได้หลายทางเช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค่า หรือ วัดจากกำไรสุทธิ เป็นต้น

6) การปฏิบัติตามผลเสนอแนะ (Deployment) คือ การนำเอาข้อมูลที่เป็นผลลัพธ์มาลองปฏิบัติจริงกับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวมผลที่ได้เพื่อการปรับปรุงต่อไป

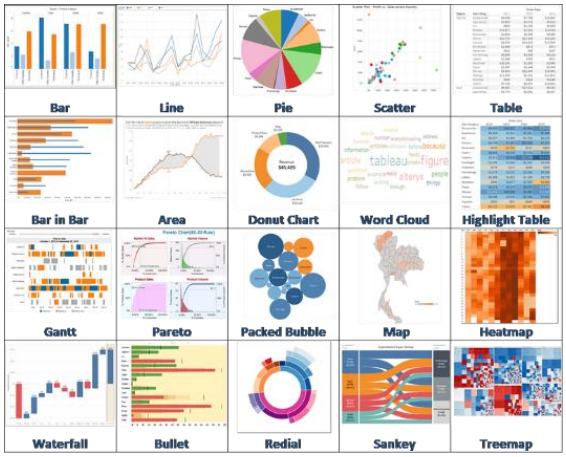
จากขั้นตอนที่กล่าวมาคือการทำเหมืองข้อมูลในงานระบบทางธุรกิจเป็นกระบวนการทางสถิติที่เน้นการทำกับข้อมูลที่มีจำนวนมากในหลากหลายรูปแบบ คัดเลือกเฉพาะข้อมูลที่สำคัญและจำเป็นต้องนำมาใช้งาน จัดทำการกำหนดรูปแบบจัดแบ่งกลุ่มลำดับความสำคัญ จากนั้นจึงจะเริ่มค้นหารูปแบบแนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยแต่ละขั้นตอนจะอาศัยผลลัพธ์จากอีกขั้นตอนหนึ่งกลายเป็นข้อมูลให้ขั้นตอนต่อไป การทำเหมืองข้อมูลจะช่วยเปลี่ยนข้อมูลดิบให้เป็นสารสนเทศที่มีประโยชน์ การระบุแหล่งข้อมูลที่ถูกต้องจึงเป็นสิ่งที่สำคัญต่อผลลัพธ์ที่ได้จากการวิเคราะห์

2.3.4 โปรแกรม Tableau Public

Tableau คือ โปรแกรมที่ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ (Data Visualization) เพื่อทำให้มีความเข้าใจในข้อมูลมากขึ้น ทำให้สามารถนำข้อมูลจำนวนมากมาวิเคราะห์และมองออกได้อย่างง่ายขึ้น ซึ่งตัวโปรแกรม Tableau มีด้วยกันหลายเวอร์ชันด้วยกัน เช่น Tableau Desktop, Tableau Online เป็นต้น แต่ในวันนี้เราจะมาเรียนในโปรแกรม Tableau Public เนื่องจากโปรแกรมนี้เป็นตัวที่ฟรี และสามารถใช้งานได้ดี

2.3.4.1 Tableau ในมุมมองด้าน Visualization

ในความเห็นของผู้เขียน ถ้าจะให้คำนิยามว่า Tableau คือซอฟต์แวร์ประเภทไหนที่ตรงประเด็นที่สุดแล้ว ผู้เขียนเห็นว่า Data Visualization น่าจะเป็นคำนิยามที่ตรงประเด็นมากที่สุด กล่าวคือ Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) ตัวอย่างของการแปลงข้อมูลเป็นภาพเป็นสิ่งที่เราคุ้นเคย และมีมาเนิ่นนานแล้วก็คือการสร้างกราฟนั่นเอง จะเห็นได้ว่าเรามีการสร้าง ตาราง กราฟแท่ง กราฟเส้น กราฟวงกลม (pie) มาก่อนแล้ว กราฟที่กล่าวมาถือเป็น Visualization อย่างหนึ่ง แต่ก็ยังมีกราฟรูปแบบอื่น ๆ อีกมากมายที่มีประสิทธิภาพและทรงพลัง ในการสื่อสารมากกว่า รูปด้านล่างคือตัวอย่าง Visualization ที่สร้างโดย Tableau



ภาพที่ 2.32 กระบวนการวิเคราะห์ข้อมูลด้วย (CRISP-DM)

ที่มา : https://www.aiteam.co.th/what-is-tableau/

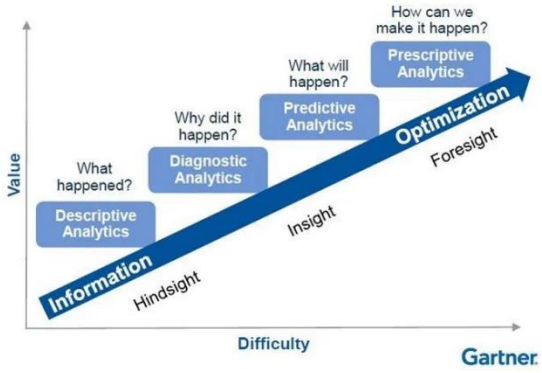
ความท้าทายที่สำคัญไม่ได้อยู่ที่ซอฟต์แวร์สามารถทำกราฟได้หลายรูปแบบ แต่สิ่งที่สำคัญจริง ๆ ที่เป็นแก่นแท้คือ ทำอย่างไรให้ผู้ใช้งานทั่วไปสามารถสร้าง Visualization เหล่านี้ได้ด้วยตนเอง โดยไม่ต้องใช้ Skill ทางด้านภาษาคอมพิวเตอร์ที่ซับซ้อน เพื่อให้เข้าถึงคนหมู่มากได้

2.3.4.2 Tableau ในมุมมองด้าน Analytics

Gartner ได้จำแนกการทำ Analytics ออกเป็น 4 ระดับดังรูป ประเด็นที่ผู้เขียนต้องการจะชี้ให้เห็นคือ Tableau ไม่ได้เป็นซอฟต์แวร์ที่สามารถทำ Analytics ได้ทั้ง 4 ระดับ การทำ Analytics ที่ Tableau สามารถทำได้ดีมี 2 ระดับคือ

1) Descriptive Analytics คือการทำ Dashboard เพื่อสรุปข้อมูลประจำรอบเวลา เช่น รอบสัปดาห์ หรือรอบเดือนเป็นต้น การทำ Dashboard เป็นการรวบรวม Visualization หลาย ๆ แบบเข้าด้วยกัน ซอฟต์แวร์ Tableau, Microsoft Power BI และ Qlik ต่างก็สามารถสร้าง Dashboard ที่มีประสิทธิภาพเหมือน ๆ กัน

2) Diagnostic Analytics คือการหาคำตอบที่สงสัยว่า สิ่งที่เกิดขึ้นนั้นเกิดได้อย่างไร การทำ Analytics ระดับนี้ Tableau จะโดดเด่นมากที่สุด (บางครั้งเรียกว่า Data Exploration) เพราะ Tableau มีความยืดหยุ่น รวดเร็ว เป็นมิตรกับผู้ใช้งานมากที่สุด จะได้เปรียบซอฟต์แวร์อื่น ๆ



ภาพที่ 2.33 รูปแสดง Analytics ในระดับต่าง ๆ

ที่มา : https://www.aiteam.co.th/what-is-tableau/

ส่วนการทำ Analytics อีก 2 ระดับนั้น ขออธิบายคร่าว ๆ ณ ที่นี้ดังนี้

3) Predictive Analytics คือการใช้ Model และทฤษฎีทางด้าน Data Science หรือ Machine Learning มาวิเคราะห์ข้อมูล เพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต ซอฟต์แวร์ที่สามารถทำได้คือซอฟต์แวร์ประเภท Data Science[9] หรือ Machine Learning Platform เช่น Rapid Miner, SAS, Alteryx, Knime และ ai เป็นต้น ถึงแม้ว่า Tableau จะไม่สามารถทำ Predictive Analytics ได้ด้วยตนเอง แต่ Tableau ก็เปิดช่องทางให้เรียกใช้ภาษา R หรือ Python ในการคำนวณ

ขั้นสูงได้ ซึ่งก็จะมีความยุ่งยากเพิ่มขึ้น แต่ก็เป็นทางออกที่ยืดหยุ่น และทำให้ Tableau สามารถทำ Predictive Analytics ได้ในระดับหนึ่งเลยทีเดียว

4) Prescriptive Analytics คือซอฟต์แวร์ประเภท Optimization หรือ Simulation คือการสร้าง Model เพื่อหาจุดที่ดีที่สุด (Optimal Solution) ซอฟต์แวร์ที่อยู่ใน Class เหล่านี้ได้แก่ IBM, Fico และ Lingo เป็นต้น

2.3.5 โปรแกรม Minitab

Minitab เป็นโปรแกรมสำเร็จรูปใช้ประมวลผลข้อมูลทางด้านสถิติ โดยพัฒนาจากกลุ่มนักวิชาการทางด้านสถิติมากว่า 30 ปีแล้ว โดยปัจจุบันได้พัฒนาปรับปรุงมาจนถึงเวอร์ชัน 16 ซึ่งสามารถใช้กับระบบปฏิบัติการวินโดวส์

โดย Minitab เข้ามามีบทบาทสำหรับผู้ใช้สถิติในส่วนของการประมวลผลและการแสดงผลข้อมูลในลักษณะของตัวเลขและผลในลักษณะของกราฟ ประกอบกับเทคโนโลยีทางด้านคอมพิวเตอร์ได้พัฒนาและมีบทบาทต่อชีวิตประจำวัน

ดังนั้น Minitab จึงถูกเลือกใช้ด้วยเหตุผลหลัก 3 ประการ คือ

1) ความซับซ้อนในการประมวลผล (Complexity)

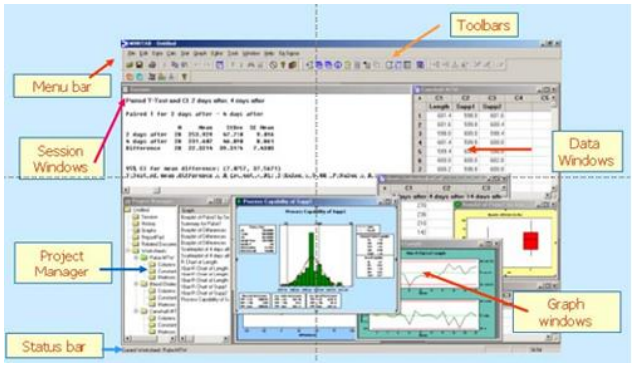
2) ความเที่ยงตรงและแม่นยำในการประมวลผลข้อมูล (Accuracy)

3) ความรวดเร็วและความสามารถในการทำซ้ำ (Repeatability)

Minitab เป็นโปรแกรมที่มีความโดดเด่นในด้านการใช้งานที่ง่ายและมีการพัฒนาปรับปรุงฟังก์ชันต่างๆ ให้สอดคล้องกับความรู้และทฤษฎีใหม่ๆ รวมถึงการประยุกต์ทางด้านสถิติโดยเฉพาะในงานด้านคุณภาพอย่างต่อเนื่อง

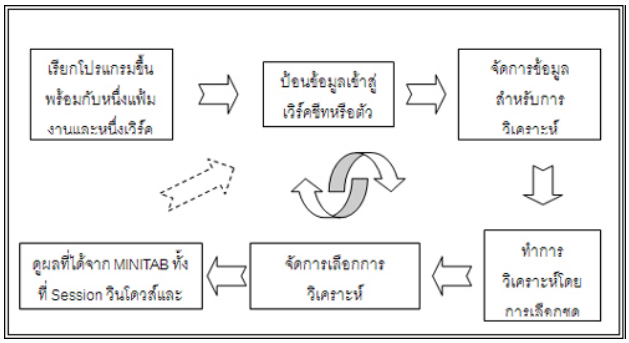
ดังนั้น Minitab จึงเป็นที่รู้จักกันเป็นอย่างดีสำหรับกลุ่มผู้ที่พัฒนาปรับปรุงคุณภาพด้วยหลักการ “ซิกซ์ ซิกม่า” เนื่องจาก Minitab เป็นเครื่องมือที่สำคัญสำหรับกลุ่มผู้ใช้กลุ่มนี้แต่ไม่ได้หมายความว่าจะจำเพาะกลุ่มผู้ใช้กลุ่มนี้เท่านั้น Minitab ยังเป็นโปรแกรมที่นิยมใช้กันอย่างแพร่หลายในกลุ่มนักวิชาการ นักวิจัย นักวิทยาศาสตร์ และผู้ใช้สถิติทั่วไป

ถึงแม้ว่าโปรแกรม Minitab จะเป็นโปรแกรมที่ช่วยให้การทำงานของเราสะดวกขึ้นมาก แต่ความรู้ความเข้าใจในการทำงานของโปรแกรมก็ยังเป็นสิ่งสำคัญสำหรับผู้ใช้เสมอ ในบทนี้จะขออธิบายภาพรวมและการทำงานของโปรแกรมโดยทั่วไป



ภาพที่ 2.34 ลักษณะหน้าตาโปรแกรม Minitab

ที่มา : http://www.thailandindustry.com/indust_newweb/news_preview.php?cid=13107



ภาพที่ 2.35 ภาพรวมการใช้โปรแกรม Minitab

ที่มา : http://www.thailandindustry.com/indust_newweb/news_preview.php?cid=13107

2.3.5.1 ขั้นตอนการใช้งานโปรแกรม Minitab

ขั้นตอนที่ 1 การเปิดโปรแกรมหรือเริ่มต้นเข้าสู่การทำงานบนแฟ้มงาน

ขั้นตอนที่ 2 ป้อนข้อมูลเข้าสู่เวิร์คชีท เนื่องจากโปรแกรม Minitab เป็นโปรแกรมสำหรับการวิเคราะห์และประมวลผลข้อมูล ดังนั้นเมื่อเราเปิดแฟ้มงานแล้วสิ่งที่จำเป็นมากสำหรับการใช้งานโปรแกรมคือ ข้อมูลซึ่งจะเป็นการป้อนจากแป้นพิมพ์ การนำข้อมูลจากฐานข้อมูลชนิดอื่น เช่น Excel, Access หรือ Text เป็นต้นหรือการเปิดจากไฟล์เวิร์คชีทของ Minitab ก็แล้วแต่ซึ่งขึ้นอยู่กับผู้ใช้เอง

ขั้นตอนที่ 3 เป็นส่วนการจัดการข้อมูลเป็นขั้นตอนที่สามารถข้ามไปยังขั้นตอนต่อไปได้ ถ้าหากว่าข้อมูลที่อยู่ในเวิร์คชีทมีความพร้อมและอยู่ในรูปแบบที่ถูกต้องสำหรับการวิเคราะห์โดยใช้โปรแกรม Minitab แต่ถ้าหากไม่เราจำเป็นต้องจัดการข้อมูลดังกล่าวก่อนอาจใช้ฟังก์ชันที่มีใน Minitab หรือการจัดการจากโปรแกรมอื่นก่อนนำเข้าสู่เวิร์คชีทใน Minitab ก็ได้แล้วแต่ซึ่งจะขึ้นอยู่กับปัญหาและประสบการณ์ในการจัดการข้อมูลของผู้ใช้งานโปรแกรม ตัวอย่างขั้นตอนนี้คือ การรวมข้อมูล (Stack) การเปลี่ยนทิศทางการเรียงข้อมูล (Transpose Data) การคำนวณข้อมูล (Calculate) การสร้างชุดข้อมูลย่อย (Subset) เป็นต้น

ขั้นตอนที่ 4 เป็นขั้นตอนการวิเคราะห์ทางสถิติโดยเราจะเลือกฟังก์ชันหรือตัวสถิติที่เราจะทำการวิเคราะห์ เช่น การวิเคราะห์ความแปรปรวน (ANOVA) การออกแบบการทดลอง (DOE) เป็นต้น

ขั้นตอนที่ 5 เป็นส่วนในรายละเอียดของการวิเคราะห์ใดๆที่เราเลือก โดยปกติแล้วถ้าเราใช้การตั้งค่าเริ่มต้นของโปรแกรม (Default) ซึ่งเป็นค่าที่มีการใช้ทั่วไปแล้วเมื่อเราใส่ข้อมูลครบถ้วนตามที่โปรแกรมต้องการแล้ว เราสามารถข้ามขั้นตอนนี้ไปได้ แต่ในบางครั้งเราต้องการปรับเปลี่ยนค่าดังกล่าว เช่นค่าความเชื่อมั่น 95% เป็น 99% การสั่งสร้างกราฟ หรือจะเป็นการสั่งให้โปรแกรมแสดงค่าหรือเก็บค่าใด ๆ ที่เราต้องการ เป็นต้น

ขั้นตอนที่ 6 เป็นการจัดการกับผลการประมวลผลที่โปรแกรม Minitab ดำเนินการให้ซึ่งจะออกมาใน 2 รูปแบบคือ ผลทางตัวเลขหรือตัวอักษรและกราฟ ตัวอย่างของการทำงานในขั้นตอนนี้ เช่น การสร้างรายงาน ตัดข้อมูลผิดปกติ การเปลี่ยนสีกราฟ เป็นต้น

จากขั้นตอนโดยรวมทั้งหมดที่กล่าวมานี้ โดยปกติผู้ใช้งานสามารถที่จะทำซ้ำหรือการวนหรือย้อนกลับขั้นตอนต่าง ๆ ได้ตลอดเวลา ตราบใดที่เรากำลังทำงานบนแฟ้มงาน

2.3.5.2 ความสามารถด้านฟังก์ชันทางสถิติและกราฟ

1) Basic Statistics เป็นฟังก์ชันพื้นฐานทางสถิติซึ่งประกอบด้วยสถิติเชิงพรรณา (descriptive) และสถิติเชิงอนุมาน(inferential) เช่น การหาค่าพารามิเตอร์ทางสถิติ เช่น ค่าเฉลี่ย (mean), ค่าความแปรปรวน (variance), ค่าพิสัย (range) เป็นต้น นอกจากนี้ยังมี

ชุดคำสั่งในการหาช่วงความเชื่อมั่นและการทดสอบสมมติฐาน (confidence interval และ hypothesis testing) โดยผลการคำนวณจะให้ทั้งผลลัพธ์บน Session และกราฟ

2) Regression Analysis เป็นฟังก์ชันการวิเคราะห์การถดถอยเชิงเส้นได้ทั้งรูปแบบสมการเส้นตรง (linear) และสมการกำลัง (polynomial) หรือรูปแบบอื่น ๆ ที่ต้องการ รวมถึงการเก็บค่าเศษเหลือ (residual) และกราฟประกอบการวิเคราะห์

3) ANOVA เป็นฟังก์ชันการวิเคราะห์ความแปรปรวน โดยสามารถวิเคราะห์ปัจจัยตั้งแต่ 1 ปัจจัย (one-way) 2 ปัจจัย (two-way) หรือมากกว่า (genera) รวมถึงการแสดงผลกราฟปัจจัยอิทธิพลหลัก (main effect plot) และปัจจัยอิทธิพลร่วม (interaction effect plot) และอื่น ๆ ที่เกี่ยวข้อง

4) Statistical Quality Tools เป็นฟังก์ชันสนับสนุนงานทางด้านการวิเคราะห์ด้านคุณภาพ ประกอบด้วยหัวเรื่องหลัก 4 เรื่อง คือ Quality Tools สามารถทำการสร้างกราฟพาเรโต (pareto) รันชาร์ต (run chart) เป็นต้น Control Charts ชุดคำสั่งสร้างแผนภูมิควบคุมมากมายหลากหลายครอบคลุมข้อมูลทุกประเภท รวมถึงคำสั่งเพิ่มเติมช่วยในการวิเคราะห์ข้อมูลผิดปกติ

5) Capability Analysis เป็นคำสั่งวิเคราะห์ความสามารถกระบวนการโดยการนำเอาความผันแปรข้อมูลเทียบกับข้อกำหนด (specification) ซึ่งสามารถจัดการได้ทั้งกรณีข้อมูลเป็นปกติ (normal)หรือไม่ปกติ (non-normal) และ Measurement System เป็นคำสั่งวิเคราะห์และประเมินความสามารถกระบวนการวัด เช่น Stability Bias Linearity และ Gage R&R

6) Design of Experiment เป็นชุดคำสั่งในการออกแบบการทดลอง ประกอบด้วย factorial, response surface, mixture และ touchy สำหรับคำสั่งการออกแบบการทดลองในโปรแกรม Minitab จะช่วยเหลือตั้งแต่การออกแบบการทดลอง การจัดเก็บข้อมูลและการวิเคราะห์ผลที่ได้และกราฟประกอบการแปลผล รวมถึงการหาการตั้งค่าเพื่อผลลัพธ์ที่ต้องการ (response optimizer)

7) Reliability เป็นฟังก์ชันสำหรับวิเคราะห์ค่าความน่าเชื่อถือของผลิตภัณฑ์ โดยอาศัยหลักกการทางสถิติ โดยมีฟังก์ชันช่วยในการหาฟังก์ชันความน่าจะเป็น (fit distribution) แบบต่าง ๆ และทำการวิเคราะห์ผ่านฟังก์ชันโดยอาศัยหลักการความน่าจะเป็น

8) Power and Sample Size เป็นฟังก์ชันเพื่อช่วยในการหาขนาดทดสอบ (sample size) หรือความเชื่อมั่นในการทดสอบ (power) สำหรับการทดสอบสมมติฐานแต่ละแบบ

9) Multivariate Analysis เป็นฟังก์ชันสำหรับการวิเคราะห์ตัวแปรหลายตัวแปร เช่น คำสั่ง Principal Component, factor analysis, cluster analysis เป็นต้น

10) Time Series and Forecasting เป็นฟังก์ชันการวิเคราะห์ข้อมูลแปรผันตามเวลาเพื่อดูแนวโน้ม ลักษณะกราฟหรือทิศทาง เพื่อใช้ในการทำนาย

11) Nonparametric เป็นฟังก์ชันการวิเคราะห์โดยไม่ใช้พารามิเตอร์

12) Tables เป็นฟังก์ชันที่จัดการข้อมูลที่เป็นอยู่ในรูปแบบของตารางของข้อมูลนับจำนวนความถี่ พร้อมคำสั่งการวิเคราะห์ Chi-square

13) Simulation and Distribution เป็นคำสั่งช่วยในการสุ่มชุดข้อมูลผ่านฟังก์ชันความน่าจะเป็น (distribution) เพื่อประโยชน์ในการทำการทดสอบโมเดลหรือการวิเคราะห์ข้อมูล

14) EDA (explore data analysis) เป็นหมวดฟังก์ชันในการประเมินข้อมูลเพื่อดูลักษณะการกระจายของข้อมูล เช่นคำสั่ง Stem-and leaf หรือ boxplot เป็นต้น

2.3.6 โปรแกรม IBM SPSS

โปรแกรม IBM SPSS for Windows (Statistical Package for Social Science) เป็นโปรแกรมสำเร็จรูปทางสถิติที่ใช้ในการวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพ มีการจัดการข้อมูลอย่างเป็นระบบ สาารถบันทึกและสร้างไฟล์ข้อมูลได้ และมีการนำเสนอข้อมูลในรูปแบบต่าง ๆ

การใช้โปรแกรม SPSS เบื้องต้น

โปรแกรม IBM SPSS เป็นโปรแกรมสำเร็จรูปที่ใช้สำหรับการวิเคราะห์หาข้อมูลทางสถิติ พัฒนาโดยบริษัท SPSS Inc. ประเทศสหรัฐอเมริกา โดยในระยะเริ่มต้นเรียก โปรแกรม SPSSX สำหรับใช้กับเครื่องคอมพิวเตอร์ ชนิดมินิหรือเมนเฟรมคอมพิวเตอร์ต่อมาเมื่อไมโครคอมพิวเตอร์

ได้มีการใช้กันอย่างแพร่หลายมากขึ้น จึงได้มี การพัฒนาโปรแกรม SPSS เป็น SPSS/PC+ ทำงานภายใต้ระบบปฏิบัติการที่เรียกว่า DOS (Disk Operating System) ซึ่งใช้กับเครื่องไมโครคอมพิวเตอร์หรือที่เรียกว่าคอมพิวเตอร์ส่วนบุคคล

ข้อตกลงเบื้องต้น

1) สำหรับการใช้งานโปรแกรม SPSS for Windows ที่กล่าวถึงในเอกสารชุดนี้จะอธิบายการใช้งานโปรแกรม SPSS for Windows version 11.5 เป็นหลัก โดยจะขอเรียกย่อ ๆ ว่า SPSS และเนื่องจากโปรแกรม SPSS for Windows version 11.5 สามารถใช้งานได้บนระบบปฏิบัติการ Windows 98 ขึ้นไป ดังนั้นจะ ขอเรียกโดยรวมว่า Windows

2) รูปภาพ และผลลัพธ์ต่าง ๆ ที่แสดงในหนังสือเล่มนี้เกิดจากการใช้งานโปรแกรม IBM SPSS for Windows version 11.5 บนระบบปฏิบัติการ Windows XP ซึ่งจะมีหน้าตาคล้ายคลึงกันกับการใช้งานบน Windows อื่น ๆ

3) สำหรับผู้ที่ต้องการเรียนรู้การใช้โปรแกรม SPSS นี้ควรจะมีประสบการณ์ในการใช้งาน Windows มาบ้าง พอสมควร และควรมีความรู้ในเรื่องสถิติและการวิจัยบ้าง เพื่อให้เกิดความเข้าใจยิ่งขึ้น

ความสามารถของโปรแกรม IBM SPSS ชุดพื้นฐาน

โปรแกรม IBM SPSS (Statistical Package for the Social Sciences) เป็นโปรแกรมสำเร็จรูปทางสถิติที่มีชื่อเสียงใช้กันอย่างแพร่หลาย จัดเป็นโปรแกรมทางสถิติขนาดใหญ่ที่มีการพัฒนามาหลายรุ่น โปรแกรม SPSS มีความสามารถที่จัดแบ่งได้ 2 ประเภทใหญ่ ๆ ดังนี้ (ศิริชัย พงษ์วิชัย, 2538, หน้า 4-5)

1) ความสามารถในการวิเคราะห์ข้อมูลด้วยเทคนิคทางสถิติ

การวิเคราะห์ข้อมูลด้วยเทคนิคทางสถิติเบื้องต้น เช่น ค่าเฉลี่ยร้อยละค่าส่วนเบี่ยงเบนมาตรฐาน ค่าฐานนิยม ค่าแสดงตำแหน่งของข้อมูล เช่น ควอไทล์ การนำเสนอข้อมูลในรูปของกราฟ การแจกแจงความถี่ในรูปของตาราง การวิเคราะห์ข้อมูลแบบ Multiple Response การวิเคราะห์ข้อมูลด้วยเทคนิคทางสถิติขั้นต้น ที่ใช้ในการสรุปผลข้อมูลต่อไปนี้ ค่าสถิติที่ใช้ในการทดสอบสมมติฐานด้วยวิธีของ t-test χ2-test F-test ด้วยการวิเคราะห์ความแปรปรวน (Analysis

of variance) ทั้งแบบทางเดียว แบบหลายทาง การทดสอบแบบจับคู่พหุคูณ การตรวจสอบ และหาค่าความสัมพันธ์ ด้วย χ2-test,Pearson Correlation, Spearman Rank Test ฯลฯ การวิเคราะห์การถดถอยแบบเส้นตรง (Linear Regression) และการวิเคราะห์การถดถอยแบบไม่ใช่เส้นตรง (Non-Linear Regression) การทดสอบด้วยวิธีนอนพาราเมตริก (Non-parametric) เช่น Kolmogorov – Smirnov, Friedman test ฯลฯ การวิเคราะห์ข้อมูลด้วยเทคนิคทางสถิติขั้นสูงที่ใช้ในการสรุปผลข้อมูล ซึ่งประกอบเทคนิคต่าง ๆ เช่น Factor Analysis, Discriminant Analysis

2) ความสามารถในการจัดการกับข้อมูล และผลลัพธ์ ในรูปของตาราง และกราฟ เป็นความสามารถที่จะดำเนินการจัดการเกี่ยวกับข้อมูลในลักษณะต่าง ๆ ต่อไปนี้

1) การเปลี่ยนรูปข้อมูล (Data Transformation)
2) การเลือกข้อมูล (Select Case)
3) การจัดการกับไฟล์ข้อมูล
4) การดำเนินการกับข้อมูลในลักษณะอื่น ๆ

2.4 วรรณกรรมที่เกี่ยวข้อง

วรางคณา กีรติวิบูลย์ (2557) วัตถุประสงค์ของการวิจัยครั้งนี้ คือ การสร้างตัวแบบพยากรณ์ที่เหมาะสมที่สุดของมูลค่าการส่งออกข้าวหอมมะลิ โดยใช้ข้อมูลจากเว็บไซต์ของสำนักงานเศรษฐกิจการเกษตร ตั้งแต่เดือนมิถุนายน 2545 ถึงเดือนกรกฎาคม 2556 จำนวน 134 ค่า ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 คือ ข้อมูลตั้งแต่เดือนมิถุนายน 2545 ถึงเดือนกรกฎาคม 2555 จำนวน 122 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติ 3 วิธี ได้แก่ วิธีบอกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก และวิธีการพยากรณ์รวม ชุดที่ 2 คือ ข้อมูลตั้งแต่เดือนสิงหาคม 2555 ถึงเดือนกรกฎาคม 2556 จำนวน 12 ค่า สำหรับการตรวจสอบความแม่นของตัวแบบพยากรณ์ด้วยเกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด ผลการวิจัยพบว่า จากวิธีการพยากรณ์ทั้งหมดที่ได้ศึกษา วิธีการพยากรณ์รวมเป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุด ซึ่งมีตัวแบบพยากรณ์เป็นเมื่อ และแทนค่าพยากรณ์

เดี่ยว ณ เวลา t จากวิธีบอกซ์–เจนกินส์ และวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก ตามลำดับ

ศรัณย์ ทัศกานิเวศน์ (2562) การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาแนวโน้มการแปรผันตามฤดูกาล และการพยากรณ์มูลค่าการส่งออกข้าวหอมมะลิของประเทศไทย ปี 2561 – 2562 ข้อมูลที่นำมาใช้ในการสร้างตัวแบบการพยากรณ์เป็นข้อมูลรายไตรมาส แบ่งออกเป็น 2 ชุด คือ ชุดที่ 1 ข้อมูลสำหรับสร้างตัวแบบ (Training Data) ชุดที่ 2 ข้อมูลสำหรับทดสอบตัวแบบ (Testing Data) ที่ได้มาจากการรวบรวมของสำนักงานเศรษฐกิจการเกษตร ทำการวิเคราะห์อนุกรมเวลาด้วยวิธีแบบคลาสสิค ผลการศึกษาพบว่า ตัวแบบที่ใช้ในการพยากรณ์มีความแม่นยำสูงมาก ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ของข้อมูลที่สร้างตัวแบบมีค่าเท่ากับ 6.42 ส่วนค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ของข้อมูลทดสอบตัวแบบมีค่าเท่ากับ 3.09 มูลค่าการส่งออกข้าวหอมมะลิของประเทศไทยมีแนวโน้มลดลงและมีการแปรผันตามฤดูกาลของการส่งออกข้าวหอมมะลิมูลค่าการส่งออกข้าวหอมมะลิของประเทศไทยมากที่สุดในไตรมาสที่ 4 (มากกว่าปกติร้อยละ 17.63) และในไตรมาสที่ 2 ประเทศไทยมีมูลค่าการส่งออกข้าวหอมมะลิน้อยที่สุด (น้อยกว่าปกติร้อยละ 10.66)

ไกรภพ จิตต์มั่น (2559) การศึกษาครั้งนี้เป็นการศึกษาผลกระทบของมูลค่าการส่งออกข้าวหอมมะลิไทยเพื่อพยากรณ์การส่งออกในช่วงเวลาถัดไปโดยใช้แบบจำลองอารีแมกซ์ (ARIMAX) โดยมีอัตราแลกเปลี่ยน และอัตราเงินเฟ้อเป็นตัวแปรอธิบาย และมีลักษณะเป็นข้อมูลรายเดือนตั้งแต่ มกราคม พ.ศ. 2545 ถึงเดือนธันวาคม พ.ศ. 2558 รวมทั้งสิ้นจำนวน 168 เดือน ผลการศึกษาพบว่าข้อมูลที่นำมาศึกษานั้นไม่นิ่ง ณ ลำดับ (0) จึงต้องทำการหาผลต่างลำดับที่ 1 จากนั้นจึงนำเอามาเลือกแบบจำลองและพบว่าแบบจำลอง ARIMAX ((1,12)1,1) มีความเหมาะสมมากที่สุดเนื่องจากมีค่า BIC ที่ต่ำที่สุด และผลการทดสอบค่าความคลาดเคลื่อนไม่พบปัญหาอัตตสหสัมพันธ์ของตัวแปร จึงสามารถนำแบบจำลองมาอธิบายความสัมพันธ์ของตัวแปรที่นำมาศึกษาได้ ผลการประมาณค่าพบว่าอัตราแลกเปลี่ยน (EXC) มีอิทธิพลต่อมูลค่าการส่งออกข้าวหอมมะลิของไทยอย่างมีนัยสำคัญทางสถิติโดยมีการเปลี่ยนแปลงในทิศทางตรงข้าม เมื่อเงินบาทมีลักษณะ

อ่อนค่าจะทำให้การส่งออกเพิ่มมากขึ้นตามไปด้วย ทางด้านอัตราเงินเฟ้อ (IF) พบว่าไม่มีนัยยะสำคัญทางสถิต จากนั้นจึงพยากรณ์มูลค่าการส่งออกข้าวหอมมะลิล่วงหน้า 8 เดือน พบว่ามูลค่าการส่งออกข้าวหอมมะลิของประเทศไทยที่เกิดขึ้นจริงและมูลค่าที่ได้จากการพยากรณ์มีความใกล้เคียงกัน และมีค่า RMSE ต่ำกว่าแบบจำลองอื่น ๆ โดยเปรียบเทียบ

ศศิธร โกฏสีบ (2559) การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาตัวแบบที่เหมาะสมและเพื่อเปรียบเทียบตัวแบบสำหรับการพยากรณ์ราคาข้าวหอมมะลิ 105 ข้อมูลที่ใช้ในการศึกษาคือ ข้อมูลรายเดือนราคาข้าวหอมมะลิ 105 ตั้งแต่เดือนมกราคม พ.ศ.2540 ถึงเดือนธันวาคม พ.ศ.2558 รวม 18 ปี (หรือ 216 เดือน) ซึ่งเก็บรวบรวมโดยสำนักงานเศรษฐกิจการเกษตร ศึกษาตัวแบบการพยากรณ์โดยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลซ้ำสองครั้ง และวิธีบ็อกซ์และเจนกินส์ และเปรียบเทียบตัวแบบสำหรับการพยากรณ์โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) น้อยที่สุด ผลการศึกษาพบว่า วิธีการพยากรณ์ที่เหมาะสมที่สุด คือวิธีบ็อกซ์และเจนกินส์ซึ่งได้ตัวแบบพยากรณ์คือ ARIMA (1,1,2)

ปิยะพร แซ่ลิ้ม (2562) การวิจัยครั้งนี้เป็นการศึกษาการพยากรณ์ผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา มีวัตถุประสงค์เพื่อหาวิธีการพยากรณ์ที่เหมาะสมสำหรับพยากรณ์ผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมาเทคนิคการพยากรณ์เชิงสถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ วิธีการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis) โดยวิธีการถดถอยแบบลำดับขั้น (Stepwise Regressions) และวิธีการวิเคราะห์อนุกรมเวลา (Timeseries Analysis) โดยวิธีบอกซ์–เจนกินส์ (Box–Jenkins method) และใช้ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เป็นเกณฑ์ในการเปรียบเทียบวิธีการพยากรณ์ทั้ง 2 วิธีในการวิจัยครั้งนี้ข้อมูลที่นำมาศึกษานั้น เป็นข้อมูลทุติยภูมิรายปีจากแหล่งข้อมูล 2 แหล่งคือกรมเศรษฐกิจการเกษตร และกรมอุตุนิยมวิทยา ผลการเปรียบเทียบตัวแบบพยากรณ์ที่ได้จากเทคนิคการพยากรณ์เชิงสถิติทั้ง 2 วิธีพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์พบว่า ตัวแบบการถดถอยพหุคูณ (Multiple Regression Analysis) โดยวิธีการถดถอยแบบลำดับขั้น (Stepwise Regressions) เป็นตัวแบบที่เหมาะสมกับการพยากรณ์ผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา และเมื่อนำตัวแบบมา

พยากรณ์ผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา ล่วงหน้าอีก 1 คาบเวลา คือปี 2560 จากผลการพยากรณ์สรุปได้ว่า ปริมาณผลผลิตข้าวหอมมะลิจังหวัดนครราชสีมา จะมีแนวโน้มเพิ่มสูงขึ้นจากปี 2559 ประมาณ 7.429%

ธนะรัตน์ รัตนกูล (2562) วัตถุประสงค์ของการวิจัยครั้งนี้เพื่อหาตัวแบบการพยากรณ์ความต้องการข้าวไรซ์เบอรี่ด้วยการนำข้อมูลปริมาณยอดการสั่งซื้อรายเดือนที่มีลักษณะข้อมูลแบบอนุกรมเวลา ตั้งแต่ปี พ.ศ.2558-2560 ของกลุ่มนาข้าวอินทรีย์ จำนวน 1 กลุ่ม และทำการพยากรณ์ปริมาณความต้องการข้าวไรซ์เบอรี่ ล่วงหน้า 24 เดือน โดยใช้โปรแกรมสำเร็จรูปทางสถิติในการวิเคราะห์ข้อมูล ลักษณะการกระจายตัวของข้อมูลมีอิทธิพลของฤดูกาลและแนวโน้ม และผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบพยากรณ์ปริมาณความต้องการข้าวไรซ์เบอรี่ของกลุ่มนาข้าวอินทรีย์กรณีศึกษาล่วงหน้า 24 เดือน ตัวแบบพยากรณ์วินเทอร์โดยข้อมูลมีอิทธิพลของฤดูกาลและแนวโน้มเชิงบวก เป็นตัวแบบพยากรณ์ที่เหมาะสมที่สุด โดยมีค่า MAPE ซึ่งหมายถึงค่าความคลาดเคลื่อนในการพยากรณ์ เท่ากับ 7.281 รองลงมา คือตัวแบบพยากรณ์ข้อมูลมีอิทธิพลฤดูกาลเพียงอย่างเดียว มีค่า MAPE เท่ากับ 7.364 และตัวแบบพยากรณ์วินเทอร์โดยข้อมูลมีอิทธิพลของฤดูกาลและแนวโน้มเชิงคูณ มีค่า MAPE เท่ากับ 7.540 ตามลำดับ

น้ำอ้อย นิสัน (2560) วัตถุประสงค์ของการศึกษาครั้งนี้คือ การสร้างตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาปริมาณการส่งออกไก่แปรรูป โดยใช้ข้อมูลจากเว็บไซต์ของ สำนักงานเศรษฐกิจการเกษตร ตั้งแต่เดือนมกราคม 2542 ถึงเดือนตุลาคม 2558 จำนวน 222 ค่า ซึ่งข้อมูลถูกแบ่งออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่เดือนมกราคม 2542 ถึงเดือนธันวาคม 2557 จำนวน 192 ค่า สำหรับการสร้างตัวแบบพยากรณ์ โดยวิธีบอกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณ และวิธีการพยากรณ์รวม ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคมถึงเดือนตุลาคม 2558 จำนวน 12 ค่า นำมาใช้สำหรับการเปรียบเทียบความถูกต้องของค่าพยากรณ์ โดยใช้เกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ที่ต่ำที่สุด ผลการศึกษาพบว่า วิธีบอกซ์-เจนกินส์มีความถูกต้องแม่นยำในการพยากรณ์มากที่สุด (MAPE = 4.9640)

วรางคณา เรียนสุทธิ์ (2562) การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบสำหรับการพยากรณ์ปริมาณการส่งออกกุ้งสดแช่แข็ง โดยใช้ข้อมูลจากเว็บไซต์ของสำนักงานเศรษฐกิจการเกษตร ตั้งแต่เดือนมกราคม 2554 ถึงเดือนสิงหาคม 2562 จำนวน 104 ค่า ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 ตั้งแต่เดือนมกราคม 2554 ถึงเดือนธันวาคม 2561 จำนวน 96 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีบ็อกซ์ – เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลต์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของบราวน์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแดม วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีฤดูกาลอย่างง่าย วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณ และวิธีการพยากรณ์รวม ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคมถึงเดือนสิงหาคม 2562 จำนวน 8 ค่านำใช้สำหรับการเปรียบเทียบความแม่นยำของการพยากรณ์ โดยใช้เกณฑ์ร้อยละค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด ผลการศึกษาพบว่าวิธีที่มีความแม่นยำมากที่สุดคือ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบบวก

คุณานนต์ ตันตินวะชัย (2558) วัตถุประสงค์ของปัญหาพิเศษนี้คือ การหาตัวแบบที่เหมาะสมสำหรับพยากรณ์ปริมาณการส่งออกของทุเรียนสด ทุเรียนแช่แข็ง ทุเรียนกวน และทุเรียนอบแห้ง ที่เก็บรวบรวมข้อมูลโดยสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์โดยใช้ข้อมูลทุติยภูมิ แบบรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2548 ถึง เดือนธันวาคม พ.ศ. 2558 รวมทั้งสิ้น 132 เดือน โดยวิธีที่นำมาใช้ในงานวิจัยนี้ ได้แก่ วิธีแยกส่วนประกอบ วิธีปรับให้เรียบ และวิธีบ็อกซ์และเจนกินส์ ซึ่งใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นเกณฑ์ในการเลือกวิธีที่เหมาะสม จากการเปรียบเทียบผลลัพธ์ของการวิเคราะห์ทั้ง 3 วิธี พบว่าวิธีที่เหมาะสมสำหรับการพยากรณ์ปริมาณการส่งออกของทุเรียนสด และทุเรียนอบแห้ง คือ วิธีแยกส่วนประกอบ วิธีที่เหมาะสมสำหรับการพยากรณ์ปริมาณการส่งออกของทุเรียนแช่แข็ง และทุเรียนกวน คือ วิธีบ๊อกซ์และเจนกินส์ โดยอนุกรมเวลาของปริมาณการส่งออกของทุเรียนแช่แข็งมีรูปแบบเป็น ARIMA (0,0,1) x SARIMA (0,1,1)$_{12}$ อนุกรมเวลาปริมาณการส่งออกของทุเรียนกวนมีรูปแบบเป็น ARIMA(1,1,1) x SARIMA (1,1,1,)10

ภาวินี ไพศร (2019) ยางพาราเป็นพืชเศรษฐกิจและเป็นสินค้าส่งออกที่สำคัญของประเทศไทย การศึกษาและพยากรณ์ปริมาณอุปสงค์ยางพาราสามารถใช้เป็นข้อมูลสำหรับการวางแผนอุปทานเกี่ยวกับปริมาณผลผลิตและเนื้อที่เพาะปลูกได้อย่างเหมาะสม งานวิจัยนี้ได้หาตัวแบบสำหรับการพยากรณ์ปริมาณการส่งออกยางพาราประเภทน้ำยางข้น ยางแผ่นรมควันชั้น 1 และยางก้อนจับตัว ซึ่งเป็นตัวแบบอนุกรมเวลาได้แก่ ตัวแบบ SARIMA (0,0,0) (0,1,1)12 ตัวแบบ SARIMA (0,1,1) (1,1,0)12 และ ตัวแบบ SARIMA (0,1,1) (1,1,0)12 ตามลำดับ หลังจากนั้นได้วิเคราะห์สถิติย้อนหลังเกี่ยวกับผลผลิต และนำมาประมาณการผลผลิตและเนื้อที่เพาะปลูกยางพาราที่สอดคล้องกับผลการพยากรณ์ความต้องการยางพาราทั้งจากภายในประเทศและการส่งออก พบว่า ในปี 2565 ประเทศไทยควรมีผลผลิตยางพารา ประมาณ 1.573 ล้านตัน ซึ่งต้องใช้เนื้อที่เพาะปลูกประมาณ 6.561 ล้านไร่ ผลการวิจัยที่ได้สามารถนำมาวางแผนและกำหนดข้อเสนอเชิงนโยบายที่เหมาะสมในการผลิตยางพาราของไทย ซึ่งพบว่า ประเทศไทยควรลดพื้นที่เพาะปลูกยางพาราลง 16.291 ล้านไร่ เนื่องจากปัจจุบันปลูกมากถึง 22.852 ล้านไร่ โดยรัฐบาลอาจส่งเสริมให้ปลูกพืชเศรษฐกิจอื่นทดแทนเพื่อให้เกิดความสมดุลระหว่างอุปสงค์และอุปทาน

2.5 บทสรุป

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของกับการวิเคราะห์ข้อมูลปริมาณการส่งออกข้าวหอมมะลิ จากกรมการค้าต่างประเทศ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับระบบ ผู้วิเคราะห์ได้เลือกใช้ขั้นตอนกระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) ในการจัดการข้อมูลบนฐานข้อมูล เพื่อลดความซ้ำซ้อนของข้อมูล และกระบวนการวิเคราะห์ CRISP-DM จากเทคนิค Neural network ในรูปแบบอนุกรมเวลา เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยตัวแบบอนุกรมเวลาที่เลือกใช้คือ วิธีการปรับให้เรียบแบบชิงเกิ้ล เอ็กซ์โปเนนเชียล (Single Exponential) และยังมีเอกสารทฤษฎีที่เกี่ยวข้องกับการนำเสนอแผนภาพข้อมูล การออกแบบและพัฒนาเว็บไซต์ ที่ใช้จัดทำสำหรับโครงงานนี้