

บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง ของการวิเคราะห์ข้อมูลภาคีการนำเข้าและส่งออกของประเทศไทยประจำปี 2558 - 2562 และ การแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการ วิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับ ดังนี้

2.1 แนวคิด

- 2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)
- 2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้การพยากรณ์
- 2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.2 ทฤษฎี

- 2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่
- 2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล
- 2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์
- 2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS
- 2.2.5 ทฤษฎีเกี่ยวกับการ visualization
- 2.2.6 ทฤษฎีเกี่ยวกับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

- 2.3.1 แบบจำลองอนุกรมเวลา Time series
- 2.3.2 การแสดงผลแบบ visualization

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลข้อมูลภาชีนำเข้า-ส่งออกประจำปี 2558 – 2562 และการแสดงผลข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ศึกษา หลักการและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.1.1.1 การวิเคราะห์ข้อมูลเชิงคุณภาพ

การวิเคราะห์ข้อมูลเชิงคุณภาพ นับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัย วิธีการหลักที่ใช้ในการ วิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากการศึกษาจาก ข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช้สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการ วิเคราะห์หลัก แต่จะถือเป็นข้อมูลเสริมในการวิเคราะห์ ข้อมูลเชิงคุณภาพนั้น ผู้วิเคราะห์ข้อมูล นับเป็นผู้ที่มีบทบาทสำคัญยิ่งและควรมีความรอบรู้ในเรื่องแนวคิดทฤษฎีอย่างกว้างขวาง มี ความเป็นสหวิทยาการอยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความและ สร้างข้อสรุปเป็นกรอบแนวคิดและตีความหมายของข้อมูลได้หลาย ๆ แบบ ทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพนั้นอาจไม่จำเป็นต้องใช้เฉพาะการวิจัยเชิงคุณภาพเท่านั้น แต่อาจใช้กับการ วิจัยเชิงปริมาณที่มีการเก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จดบันทึกมาทำการวิเคราะห์เพื่อให้ข้อมูลที่รวบรวมมาตีความหมายและ ตอบคำถามหรือจุดมุ่งหมายของการวิจัยสำหรับสาระในการวิเคราะห์ข้อมูลเชิงคุณภาพ โดย เทคนิคการวิเคราะห์ข้อมูลเชิงคุณภาพที่สำคัญ ๆ และนักวิจัย นิยมใช้ได้แก่ การจำแนกหรือ การจัดกลุ่มข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบ การวิเคราะห์แบบ อุปนัย และการวิเคราะห์ข้อมูลที่เป็นเอกสาร ซึ่งเทคนิคการวิเคราะห์ข้อมูลเชิงคุณภาพแต่ละ เทคนิคนั้นสามารถนำเสนอรายละเอียดพร้อมตัวอย่างประกอบตามลำดับ ดังนี้

2.1.1.2 เทคนิควิเคราะห์ข้อมูลเชิงคุณภาพ

1. การวิเคราะห์โดยการจำแนกชนิดข้อมูล (Typological Analysis) การ วิเคราะห์โดยการจำแนกชนิดข้อมูล คือ การจำแนกข้อมูลเป็นชนิด (Typologies) คำว่า “Typologies” หมายถึง ขั้นตอนของเหตุการณ์ที่เกิดขึ้นอย่างต่อเนื่อง ซึ่งในกรณีที่นักวิจัยทำการ เก็บรวบรวม ข้อมูลแล้วนักวิจัยจำเป็นต้องจัดระบบข้อมูลโดยอาศัยหลักเกณฑ์ ที่ผู้วิจัย กำหนดขึ้น ซึ่งการจำแนกหรือการ จัดกลุ่มข้อมูลนี้ แบ่งได้เป็น 2 ประเภท (สุภางค์ จันทวานิช, 2540: 74-90; รัตนะ บัวสนธ์, 2551: 197-219) ได้แก่

1.1 การจำแนกข้อมูลในระดับจุลภาค

การจำแนกข้อมูลระดับนี้แบ่งเป็น 2 ประเภทย่อย ได้แก่ การ วิเคราะห์คำหลัก (Domain Analysis) กับการวิเคราะห์สารระบบ (Taxonomy Analysis) การ วิเคราะห์คำหลัก หมายถึงการจำแนก จัดกลุ่มคำชุดหนึ่งให้อยู่ภายใต้คำอีกชุดหนึ่ง ซึ่งคำ ดังกล่าวนี้อาจมีความหมายครอบคลุมคำชุดนั้น ๆ หรืออาจกล่าว ได้ว่า การวิเคราะห์คำหลัก เป็น การจัดกลุ่มคำชุดหนึ่งให้อยู่ร่วมกันโดยอาศัยลักษณะความสัมพันธ์บางอย่างของ คำแต่ละคำที่ นำมาใช้จัดกลุ่ม ทั้งนี้ลักษณะความสัมพันธ์บางอย่างที่ว่านี้เป็นความสัมพันธ์เชิงวัฒนธรรมที่ บุคคล แต่ละสังคมเป็นผู้จัดจำแนกเช่นคำว่า “คุณลักษณะครูที่ดี” นั้นมิได้หมายถึงครูที่สอนเท่านั้น แต่คำนี้หมายรวมถึงการเป็นผู้ที่มีความสามารถด้านการสอน มีคุณธรรมจริยธรรมและ บุคลิกที่เป็น

แบบอย่างแก่นักเรียน มีความเข้าใจ ด้านจิตวิทยาการเรียนการสอน สามารถ ทำงานร่วมกับผู้อื่นได้ เป็นต้น ซึ่งถ้าเปรียบเทียบกับ การวิเคราะห์ข้อมูล เชิงปริมาณก็คือการ วิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อจัดกลุ่มตัวแปรย่อยให้อยู่ภายใต้องค์ประกอบ เดียวกันนั่นเอง (รัตนะ บัวสนธ์, 2551. หน้า 198)

1.2 การจำแนกข้อมูลในระดับมหภาค

การจำแนกข้อมูลในระดับมหภาค เป็นการจำแนกข้อมูลตาม เหตุการณ์ (Event) หรือ การวิเคราะห์เหตุการณ์ ตามเรื่องราว (Event Analysis) ที่ปรากฏ ซึ่ง การวิเคราะห์ข้อมูลในระดับมหภาค แบ่งได้เป็น 2 วิธี คือ การวิเคราะห์เหตุการณ์แบบอิงทฤษฎี และแบบไม่อิงทฤษฎี - การวิเคราะห์เหตุการณ์แบบอิงทฤษฎี คือ การแยกชนิดใน เหตุการณ์นั้น ๆ โดย การยึดแนวคิดหรือกรอบการจำแนก เหตุการณ์โดยอาศัยทฤษฎีใดทฤษฎี หนึ่งเป็นกรอบการจำแนก ซึ่งกรอบการจำแนกเหตุการณ์ที่นิยมใช้หรือมีลักษณะกลาง ๆ ที่มัก นำมาใช้ร่วมกันคือ การวิเคราะห์เหตุการณ์ออกเป็น 6 ประเภท คือ ใคร ทำอะไร ที่ไหน เมื่อไร อย่างไรและทำไม - การวิเคราะห์เหตุการณ์แบบไม่อิงทฤษฎี คือ การจำแนกข้อมูล ในเหตุการณ์หนึ่ง ๆ ที่จะวิเคราะห์ตามความเหมาะสมกับข้อมูล ซึ่งอาจใช้สามัญสำนึกหรือ ประสบการณ์ของผู้วิจัย ซึ่งผู้วิจัยจะจำแนก ข้อมูลเป็นชนิดง่าย ๆ ตามประเภทที่ผู้วิจัยสังเกต เมื่อจำแนกข้อมูลเป็นชนิดแล้ว ผู้วิจัยจะพิจารณา ดูความ สม่าเสมอของการเกิดของข้อมูลต่าง ๆ ซึ่งจะเป็นพื้นฐานในการอธิบายสาเหตุของปรากฏการณ์ นอกจากนี้ แม้ว่าผู้วิจัยจะจำแนกเอง แต่ก็ต้องอาศัยมุมมองของบุคคลในสนามวิจัยเป็นพื้นฐานในการจำแนก ดังตัวอย่าง จาก งานวิจัยของรัตนะ บัวสนธ์ (2535) ที่คณะกรรมการศึกษาซึ่งเป็นบุคคลในสนามวิจัยได้จำแนก ครูออกเป็น 3 กลุ่ม

1.3 การวิเคราะห์โดยการเปรียบเทียบเหตุการณ์ (Constant Comparison)

การวิเคราะห์โดยการเปรียบเทียบเหตุการณ์ คือ การใช้วิธีการ เปรียบเทียบ โดยการ นำข้อมูลมาเทียบ เป็นปรากฏการณ์ วิธีการนี้ สามารถทำได้โดยการที่ ผู้วิจัยสังเกต หรือรวบรวมข้อมูลได้หลาย ๆ อย่างแล้วนำมา แยกตามชนิด นำมาเปรียบเทียบ กันโดยทำตารางหาความสัมพันธ์จากสิ่งต่าง ๆ เหล่านั้นและสรุปผลออกมา ผลที่ได้จากการ วิเคราะห์ด้วยวิธีการนี้จะทำให้ได้ข้อสรุปที่มีความเป็นนามธรรมมากขึ้นและครอบคลุมหรือ สามารถใช้อ้างอิงเหตุการณ์ที่เหมาะสม ทั้งนี้โดยทั่ว ๆ ไปการวิเคราะห์ โดยการเปรียบเทียบ เหตุการณ์มักจะกระทำภายหลังจากได้ทำการวิเคราะห์จำแนกหรือจัดกลุ่มข้อมูลแล้ว หลังจากนั้นจึงนำข้อมูลไปใส่ในตาราง ทำการสรุปลักษณะร่วมกันและลักษณะที่แตกต่างกันของข้อมูล เหตุการณ์เหล่านั้น วิธีการวิเคราะห์โดยการเปรียบเทียบเหตุการณ์ ขั้นตอนในการวิเคราะห์ ย่อย ๆ 4 ขั้นตอนดังนี้

- จัดชุดเหตุการณ์ใส่ตาราง ในขั้นตอนนี้เป็นการนำข้อมูล เหตุการณ์แต่ละเหตุการณ์ ซึ่งผู้วิจัยได้ทำการวิเคราะห์จำแนกไว้มาใส่ในตารางเพื่อแยก ประเด็นต่าง ๆ ของแต่ละเหตุการณ์โดยแยก ประเด็น ในแต่ละเหตุการณ์เป็น 6 ประเด็น ดัง ตัวอย่างการวิเคราะห์โดยการเปรียบเทียบเหตุการณ์

- การเปรียบเทียบเหตุการณ์ หลังจากนำเหตุการณ์แต่ละ เหตุการณ์ใส่ตารางแล้ว ดำเนินการเปรียบเทียบระหว่างเหตุการณ์ใหม่ ๆ กับเหตุการณ์เดิมซึ่ง เหตุการณ์เหล่านี้อาจเกิดขึ้นจากการ แสดง พฤติกรรมของกลุ่มบุคคลเดิมหรือกลุ่มบุคคลอื่น ๆ แต่เหตุการณ์ดังกล่าวนี้ได้รับการบันทึกลงตารางโดย

การ จำแนกประเด็นต่าง ๆ เหมือนกันเพื่อ ทำการเปรียบเทียบซึ่งกันและกัน ทั้งนี้เมื่อเปรียบเทียบระหว่าง เหตุการณ์ได้ แล้ว ก็ดำเนินการ สรุปลักษณะที่บันทึกได้ไว้ก่อน ซึ่งข้อสรุปเหล่านี้ก็คือข้อสรุปย่อย ๆ นั่นเอง

- ประมวลข้อมูลเหตุการณ์แต่ละชุด แต่ละประเด็นของข้อมูล เหตุการณ์เข้าด้วยกัน ในขั้นตอนนี้เป็นการนำเหตุการณ์แต่ละชุดที่จัดลงตารางตามประเด็นต่าง ๆ มาเปรียบเทียบกันเพื่อหาลักษณะ ร่วมและลักษณะที่แตกต่างกัน หลังจากนั้นก็เขียนสรุป บรรยายเชื่อมโยงข้อมูลเหตุการณ์แต่ละชุดเข้าด้วยกัน เป็นการสังสมข้อค้นพบ หรือข้อสรุปย่อย ๆ จากการเปรียบเทียบประเภทของข้อมูลและคุณลักษณะของ ประเภท ซึ่งกันและกัน เมื่อ เปรียบเทียบแล้ว ผู้วิจัยก็จะเริ่มเชื่อมโยงข้อมูลเหล่านั้น ทำให้เกิดความสัมพันธ์ และเกิดเป็น แนวคิดย่อย ๆ ขึ้น

- ขยายวงของการเปรียบเทียบแล้วเลือกเป็นเหตุการณ์ที่เป็น กุญแจสำคัญผู้วิจัยจะ ใช้กรอบแนวคิดที่ได้จากการสรุปลักษณะความสัมพันธ์ของข้อมูลใน เหตุการณ์ต่าง ๆ ตามขั้นตอนที่ 3 มา พิจารณา เหตุการณ์อื่น ๆ ที่มีอยู่เมื่อขยายวงของการ เปรียบเทียบออกไป คุณสมบัติที่ได้คล้ายคลึงและที่ แตกต่างกันของ ข้อมูลก็ยิ่งมีความชัดเจน มากขึ้น ซึ่งทำให้ผู้วิจัยพบข้อสรุปได้ ซึ่งข้อสรุปดังกล่าวจะมีลักษณะ เป็นกรอบแนวคิด เชิง นามธรรมเป็นฐานขั้นแรกของการนำไปสู่ทฤษฎีหรือข้อสรุปเชิงนามธรรมที่ใหญ่ขึ้น

1.4 การวิเคราะห์ส่วนประกอบ (Component Analysis)

การวิเคราะห์ส่วนประกอบของข้อมูลเป็นการวิเคราะห์คุณสมบัติ ของส่วนประกอบ ของข้อมูลแต่ละชุด แล้วนำคุณสมบัติของส่วนประกอบของข้อมูล มา เปรียบเทียบเพื่อหาลักษณะร่วมที่ เหมือนกันและแตกต่างกัน หลังจากนั้นจึงทำการสรุปบรรยาย ให้เห็นถึงความหมายของข้อมูลเหล่านั้น โดย การวิเคราะห์ส่วนประกอบจะ กระทำได้ดีก็ต่อเมื่อมี ข้อมูลตั้งแต่สองชุดขึ้นไป แต่ไม่ควรมากเกินไป (สุภางค์ จันทวนิช, 2540 : 100) เพราะ ถ้า หากข้อมูลมากเกินไปจะทำให้ยากแก่การลงสรุปเกี่ยวกับคุณสมบัติของ ส่วนประกอบข้อมูลนั้น นอกจากนั้น แล้วการลงสรุปข้อมูลที่เหมาะสมสำหรับนำมาวิเคราะห์ส่วนประกอบ ควร เป็น ข้อมูลที่มีความละเอียดและได้จาก การเก็บรวบรวมด้วยการวิเคราะห์ที่เจาะลึก หรือเน้นจุด สนใจ ทั้งนี้ เพราะว่าข้อมูลดังกล่าวสามารถนำมาแยกส่วน ประกอบได้หลายส่วน ทั้งนี้การจะ แยกส่วนประกอบของข้อมูล เพื่อพิจารณาคูสมบัตินั้นจะแยกออกเป็นกี่ส่วน นั้นขึ้นอยู่กับ การ ตัดสินใจของผู้วิจัยว่าหากแยกแล้วจะทำให้ ได้ข้อมูลสำหรับการเปรียบเทียบได้ชัดเจนขึ้นก็ควร แยกส่วนประกอบตามนั้น สำหรับการวิเคราะห์ ส่วนประกอบของข้อมูล อาจสรุปได้เป็น 5 ขั้นตอนตามลำดับดังนี้

- เลือกข้อมูล (ที่ทำการวิเคราะห์จัดกลุ่มหรือ กำหนดชื่อข้อมูล แล้ว) ที่จะนำมาแยก ส่วนประกอบเพื่อหาคุณสมบัติที่ต้องการเปรียบเทียบ ซึ่งควรมีข้อมูล อย่างน้อยตั้งแต่สองชุดขึ้นไป แต่ก็ไม่ควร มากเกินไปควรพิจารณาให้เหมาะสมสะดวกต่อการ เปรียบเทียบ

- วิเคราะห์แยกส่วนประกอบข้อมูลแต่ละชุด โดยพิจารณาว่าจะแยก ส่วนประกอบของ ข้อมูลเป็นกี่ส่วนจากคุณสมบัติใดบ้าง ส่วนประกอบที่จะแยกควรพิจารณาว่า ถ้าแยกแล้วสามารถให้คุณสมบัติ ที่จะนำมาเปรียบเทียบกันได้หรือไม่

- จดรายชื่อข้อมูล และส่วนประกอบที่จะแยกข้อมูลชุด นั้น ๆ ไว้ ในกระดาดบันทึก เพื่อป้องกันการลืม

- จัดทำตารางเปรียบเทียบคุณสมบัติข้อมูลแต่ละชุดแยกตาม ส่วนประกอบและใส่คุณสมบัติข้อมูลแต่ละชุดแยกตามส่วนประกอบลงในตาราง ถ้าคุณสมบัติ ของข้อมูลและส่วนประกอบใด ขาดหายไปอาจต้องเก็บข้อมูลเพิ่มเติม

- เปรียบเทียบคุณสมบัติของข้อมูลทั้งหมดตามส่วนประกอบ โดย พิจารณาความเหมือนและความแตกต่างและสร้างข้อสรุปที่ได้จากการเปรียบเทียบโดย บรรยายโยงให้เห็นคุณสมบัติของข้อมูลตามส่วนประกอบนั้น ๆ

1.5 การวิเคราะห์สรุปอุปนัย (Analytic Induction)

การวิเคราะห์แบบอุปนัย คือ การตีความสร้างข้อสรุปข้อมูลจาก สิ่งที่เป็นรูปธรรมหรือปรากฏการณ์ ที่มองเห็นที่เก็บรวบรวมมาได้จากข้อมูลตั้งแต่ 2 ชุดขึ้นไป เช่น การปฏิบัติงาน พฤติกรรม การสอน ตลอดจนการดำเนินชีวิต ความเป็นอยู่ ฯลฯ เมื่อผู้วิจัย ได้เห็นหรือสังเกตหลาย ๆ เหตุการณ์ต่าง ๆ แล้วจึงลงมือสรุปแต่หากข้อสรุปนั้นยังไม่ได้รับการ ตรวจสอบอื่น ๆ ก็ถือว่าผลที่ได้เป็นสมมติฐาน หากได้รับการยืนยันก็ถือว่าเป็นข้อสรุปซึ่งมี ความเป็นนามธรรมในระดับต้น ๆ ซึ่งการวิเคราะห์สรุปอุปนัย จัดได้ว่าเป็นวิธีการวิเคราะห์ ข้อมูลเชิงคุณภาพที่ต้องนำมาใช้สำหรับวิเคราะห์ข้อมูลในการวิจัยเชิงคุณภาพทุกเรื่อง ทั้งนี้ เพราะการวิเคราะห์ สรุปอุปนัย เป็นการพิจารณาลักษณะร่วมกันของข้อมูลรูปธรรมเพื่อสรุป ร่วมลักษณะดังกล่าว ซึ่งเป็นไปตามหลัก ของคำว่า “อุปนัย” (induction) ซึ่งหมายรวมถึงการหา ความจริงจากข้อเท็จจริง (fact) ส่วนย่อยหลาย ๆ ส่วน ที่มีลักษณะเป็นรูปธรรมแล้วสรุปความ จริงชุดใหญ่ที่มีลักษณะเป็นนามธรรมครอบคลุมข้อเท็จจริงส่วนย่อยเหล่านั้น สำหรับวิธีการวิเคราะห์สรุปอุปนัยในการวิเคราะห์ข้อมูลเชิง คุณภาพ นั้น สามารถนำมาใช้ได้ 3 ลักษณะ ได้แก่ การวิเคราะห์สรุปข้อมูลจากบันทึกภาคสนาม ที่เป็นส่วนบันทึกละเอียดหรือบันทึกพรรณนา การวิเคราะห์ สรุปข้อมูลที่ได้จากการวิเคราะห์ จำแนกหรือจัดกลุ่มและการวิเคราะห์ข้อมูลที่ทำกรวิเคราะห์ส่วนประกอบแล้ว

1.6 การวิเคราะห์ข้อมูลเอกสาร

ในการวิเคราะห์ข้อมูลเอกสารนั้น สามารถทำได้โดยวิธีการเชิง ปริมาณและเชิงคุณภาพ วิธีเชิงปริมาณ คือ การทำให้ข้อมูลของเอกสารนั้น ได้แก่ ถ้อยคำ ประโยคหรือใจความที่ปรากฏในเอกสารเป็นจำนวน ที่วัดได้แล้วเจนนับจำนวนของถ้อยคำ ประโยคหรือใจความเหล่านั้น วิธีการวิเคราะห์ข้อมูลแบบนี้ที่รู้จักกันดี คือ การวิเคราะห์เนื้อหา (Content analysis) ซึ่ง โดยปกติการวิเคราะห์เนื้อหาจะทำตามเนื้อหาที่ปรากฏ

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ข้อมูลที่เก็บรวบรวมมาได้บางครั้งยังมีรูปแบบที่กระจัดกระจายเป็นรายบุคคลไม่เป็น ระบบจำเป็นต้องมีกระบวนการจัดกระทำข้อมูลเหล่านั้นให้เป็นระบบหรือเป็นหมวดหมู่เกิดเป็น สารสนเทศที่สามารถนำไปใช้ประโยชน์ เพื่อสรุปอ้างอิงไปยังประชากรต่อไป ศาสตร์ที่ถูกนำเข้ามา ช่วยในขั้นตอนของการตรวจสอบคุณภาพของเครื่องมือไปจนถึงการอ้างอิงเหล่านี้ เรียกว่า สถิติ

ความหมายของสถิติ คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยความสะดวกต่อการบริหารประเทศในด้านต่าง ๆ เช่น การทำ

สำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมด ในสมัยต่อมา คำว่า สถิติ ได้หมายถึง ตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistical data)

อีกความหมายหนึ่ง สถิติ หมายถึง วิธีการที่ว่าด้วยการเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการตีความหมายข้อมูล สถิติในความหมายนี้เป็นทั้ง วิทยาศาสตร์และศิลปศาสตร์ เรียกว่า สถิติศาสตร์

ประเภทของสถิติ สถิติแบ่งเป็น 2 ประเภท คือ

1. สถิติพรรณนา (Descriptive Statistics) เป็นสถิติที่ใช้อธิบายคุณลักษณะของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่งไม่ สามารถอ้างอิงไปยังกลุ่มอื่น ๆ ได้ สถิติที่อยู่ในประเภทนี้ เช่น ค่าเฉลี่ย ค่ามัธยฐาน ค่าฐานนิยม ส่วนเบี่ยงเบนมาตรฐาน ค่าพิสัย ฯลฯ

2. สถิติอ้างอิง (Inferential statistics) เป็นสถิติที่ใช้อธิบายคุณลักษณะของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่งหรือ หลายกลุ่ม แล้วสามารถอ้างอิงไปยังกลุ่มประชากรได้ โดยกลุ่มที่นำมาศึกษาจะต้องเป็นตัวแทนที่ดีของประชากร ตัวแทนที่ดีของประชากรได้มาโดยวิธีการสุ่มตัวอย่าง และตัวแทนที่ดีของประชากร เรียกว่า กลุ่มตัวอย่าง สถิติอ้างอิงแบ่งได้เป็น 2 ประเภท คือ

2.1 สถิติพารามิเตอร์ (Parametric Statistics) เป็นวิธีการทางสถิติที่จะต้อง เป็นไปตามข้อตกลงเบื้องต้น 3 ประการ ดังนี้

- ข้อมูลที่เก็บรวบรวมได้จะต้องอยู่ในระดับช่วงขึ้นไป (Interval Scale)
- ข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่างจะต้องมีการแจกแจงเป็นโค้งปกติ
- กลุ่มประชากรแต่ละกลุ่มที่นำมาศึกษาจะต้องมีความแปรปรวนเท่ากัน สถิติที่อยู่ในประเภท

นี้ เช่น t-test, Z-test, ANOVA, Regression ฯลฯ

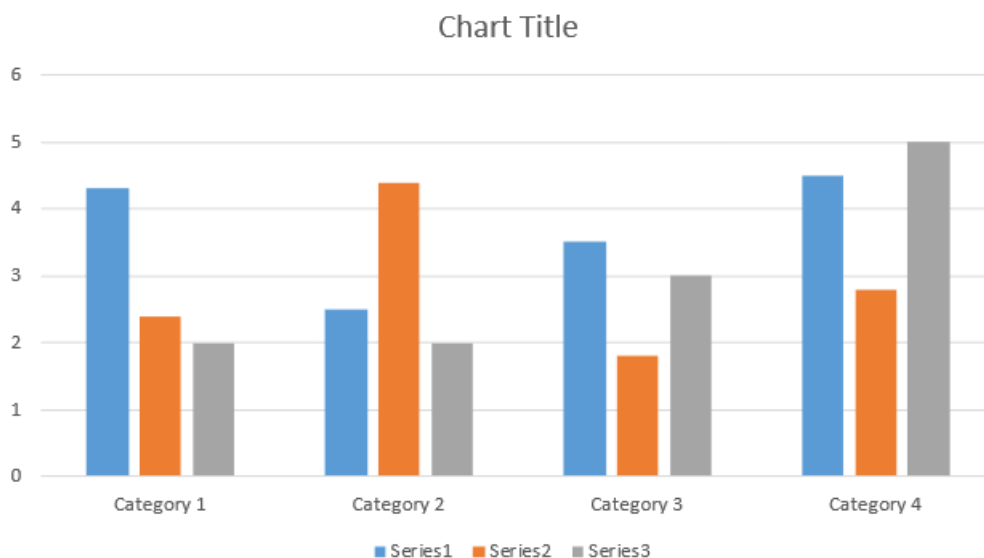
2.2 สถิติไร้พารามิเตอร์ (Nonparametric Statistics) เป็นวิธีการทางสถิติที่สามารถนำมาใช้ได้โดยปราศจากข้อตกลงเบื้องต้นทั้ง 3 ประการข้างต้น สถิติที่อยู่ในประเภทนี้ เช่น ไคสแควร์, Median Test, Sign test ฯลฯ

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณ ทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมหาศาล (Big Data) มาประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิก หรือ แม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภคสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้น ๆ ได้ ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมี ประสิทธิภาพ ซึ่งคำว่า “ประสิทธิภาพ” ในที่นี้หมายถึงมีความชัดเจน (Clarity) , มีความแม่นยำ (Precision) , และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้ เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม , รูปแบบพฤติกรรม , และความสัมพันธ์ เชื่อมโยงได้

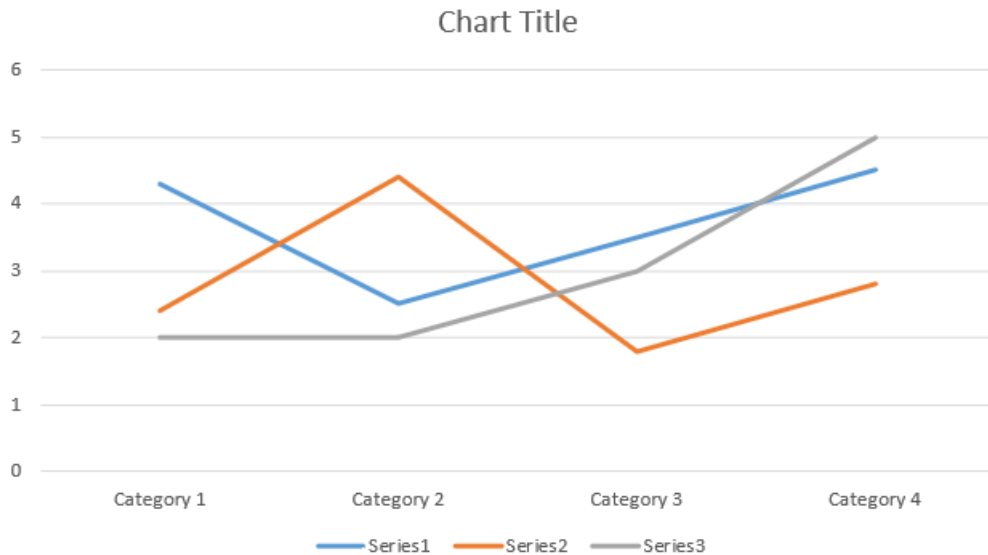
การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคเทคโนโลยี เข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น คนที่ นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมา Data Visualization เป็นการ ใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่ง อาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่น ๆ อีกมากมาย เพื่อให้เข้าใจได้ ง่าย และรวดเร็ว การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูลที่มีอยู่ เพื่อให้การ นำเสนอข้อมูลน่าสนใจและมีประสิทธิภาพสูงสุด ซึ่งเลือกแผนภูมิตามความต้องการที่ต่างกันอย่างไร แพร่หลาย ดังนี้

1. แผนภูมิแท่ง (Bar Charts) เป็นแผนภูมิที่ประกอบด้วยแกนนอน แกนตั้งที่นิยม แสดงออกมาในรูปแบบแท่งสีเหลี่ยมที่สามารถบอกความสูงได้ เหมาะสำหรับใช้การเปรียบเทียบ จำนวนของข้อมูลในแต่ละชุดเช่น รายรับในแต่ละเดือน, ยอดขายที่ขายได้จริงเปรียบเทียบกับ เป้าหมายยอดขายที่ตั้งไว้ เป็นต้นซึ่งแผนภูมิแท่งยังสามารถแบ่งออกเป็นหลายประเภท • แผนภูมิแท่งแบบจัดกลุ่ม แผนภูมิแท่งแบบจัดกลุ่ม เหมาะสมกับการ นำเสนอข้อมูลที่มีข้อมูล ย่อย ๆ อยู่ภายใต้ข้อมูลใหญ่เป็นการเน้นให้เห็นข้อมูลย่อยนั้น ๆ แผนภูมิแท่งแบบวางซ้อนกัน แผนภูมิแท่งแบบวางซ้อนกัน เหมาะสมกับการนำเสนอข้อมูลที่ต้องการให้เห็นข้อมูลย่อยในแต่ละข้อมูลใหญ่และยังแสดงให้เห็นสัดส่วนของข้อมูลย่อยต่าง ๆ เหล่านี้ได้ด้วย ใช้แผนภูมินี้เมื่อ มีชุดข้อมูลหลายชุดและต้องการเน้นผลรวมทั้งหมด



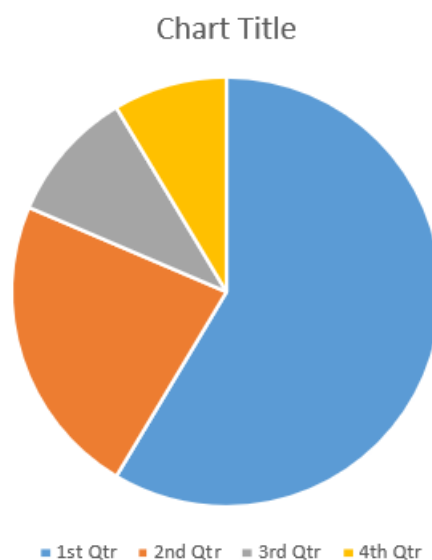
ภาพที่ 2.1 แผนภูมิแท่ง (Bar Charts)

2. แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่งประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั้นเองแผนภูมิ ประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้ แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์ แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดู แนวโน้ม เป็นต้น



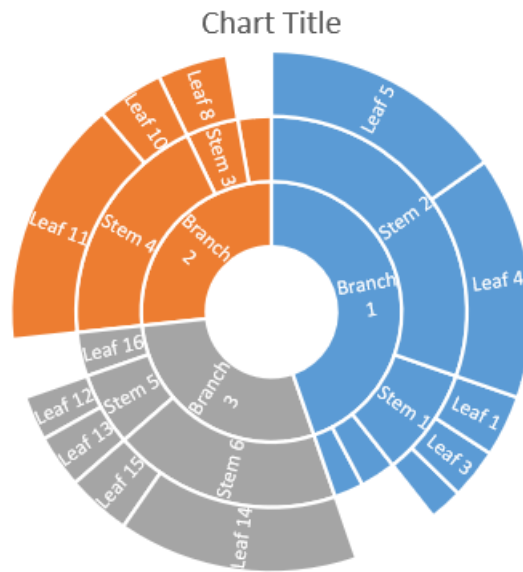
ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

3. แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มี ส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกัน อาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูล แสดงส่วนผสมต่าง ๆ เป็นต้น



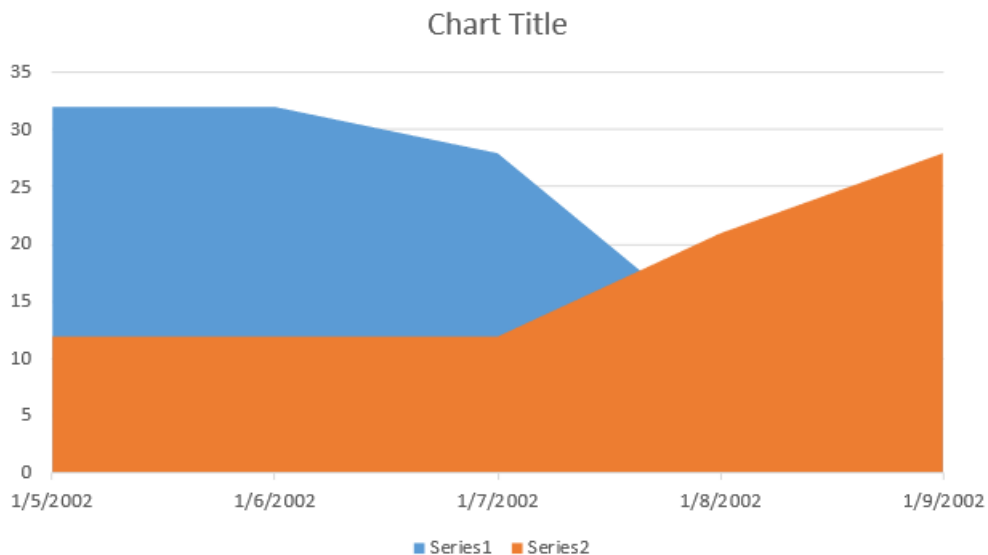
ภาพที่ 2.3 แผนภูมิวงกลม (Pie Charts)

4. แผนภูมิโดนัท (Donut Charts) แผนภูมิโดนัทที่มีหลักการออกแบบเช่นเดียวกับ แผนภูมिवงกลม แต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อน กันหลาย ๆ ชั้นนั่นเอง



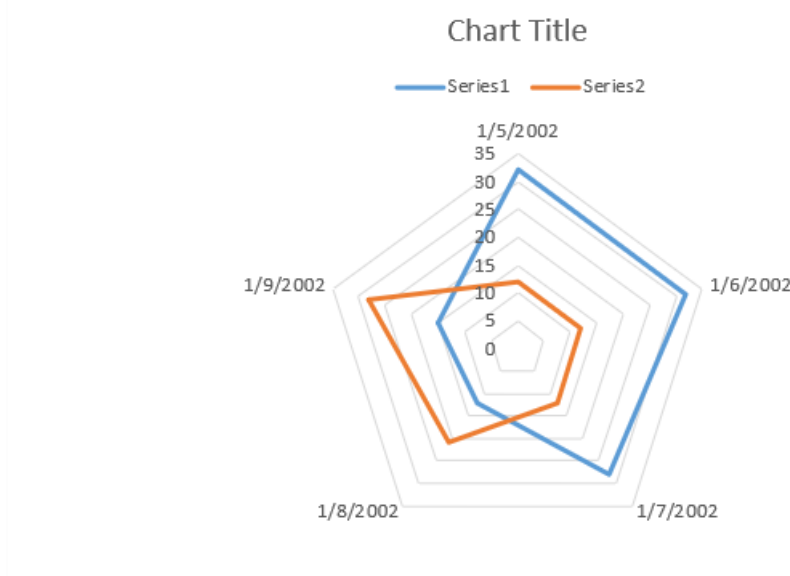
ภาพที่ 2.4 แผนภูมิโดนัท (Donut Charts)

5. แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้ เส้นข้อมูล หรือ ระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะ สำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่าง ระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าใน ห้างสรรพสินค้าในแต่ละเดือน ตามหมวดหมู่ต่าง ๆ ไล่ไป เครื่องสำอาง เสื้อผ้าแฟชั่น อาหาร ตามลำดับ



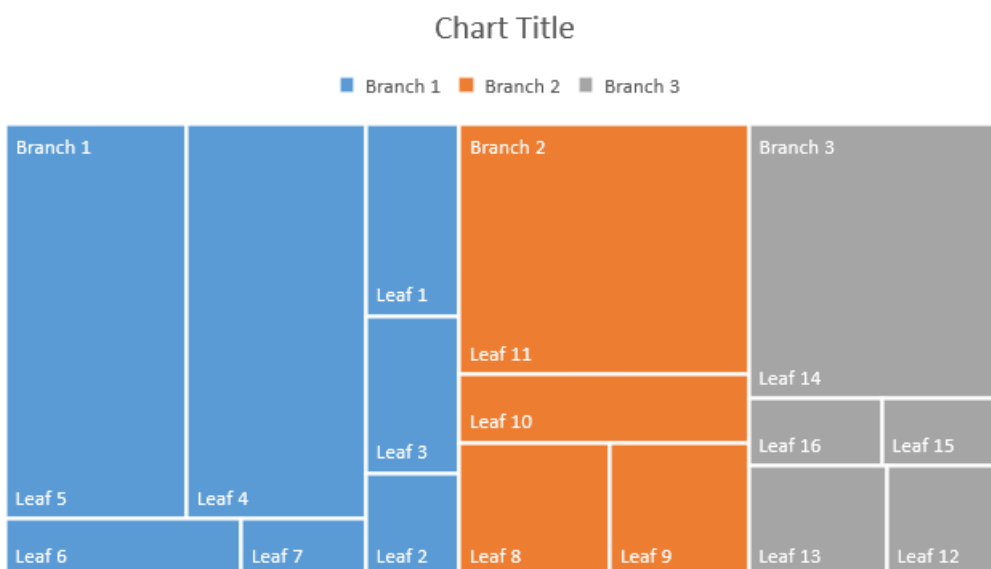
ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

6. แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบ วงกลมจำนวน เหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความ ต่อเนื่องของข้อมูลแต่เหมาะ สำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อน จุดแข็งของข้อมูลเช่น นำเสนอข้อมูล เกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อน จุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

7. แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดง ผลได้ในแบบลำดับ ชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึง เขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกัน ได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยาก หรือเป็นไปได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมาก เพื่อทำการวิเคราะห์ที่มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วง ต้นปีค.ศ. 2000 เมื่อ Dick Korth นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันใน ขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วย 3Vs ดังนี้

Volume (ปริมาณ) องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่น ๆ ในอดีต การจัดเก็บข้อมูลถือเป็นปัญหาใหญ่ – แต่เมื่อค่าใช้จ่ายในการจัดเก็บบน แพลตฟอร์มต่าง ๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้ จึงบรรเทาลง

Velocity (ความเร็ว) ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูก ส่งไปยังธุรกิจต่าง ๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ตมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับ กระแสข้อมูลเหล่านี้แบบเรียลไทม์

Variety (ความหลากหลาย) ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มี โครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูล หุ่น และธุรกรรมทางการเงิน

การวิเคราะห์ Big Data ช่วยให้องค์กรควบคุมข้อมูลของพวกเขาและใช้เพื่อ ระบุโอกาสใหม่ ๆ ในทางกลับกันนำไปสู่การเคลื่อนไหวทางธุรกิจที่ชาญฉลาดเพื่อการ ดำเนินงานที่มีประสิทธิภาพมากขึ้น ผลกำไรที่สูงขึ้นและลูกค้าที่มีความสุขมากขึ้น ใน รายงาน Big Data ใน บริษัท ขนาดใหญ่โดยผู้อำนวยการฝ่ายวิจัยของ Tom Davenport ให้ สัมภาษณ์ว่า มากกว่า 50 ธุรกิจใช้และทำความเข้าใจว่าพวกเขาใช้ Big Data อย่างไร และ พบว่าสามารถช่วยเหลือธุรกิจได้ดังต่อไปนี้

1. ลดต้นทุน: เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์ บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูล จำนวนมารวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

2. เร็วกว่าและตัดสินใจดีกว่า: ด้วยความเร็วของ Hadoop และการ วิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะ สามารถสร้างข้อมูลได้ทันที และสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ

3. ผลิตภัณฑ์และบริการใหม่: ด้วยความสามารถในการวัดความต้องการ และความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดย ดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะ สามารถสร้างผลิตภัณฑ์ใหม่ ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

2.2.1.2 การจัดการข้อมูลขนาดใหญ่

การรวบรวมข้อมูล การรวบรวมข้อมูลของ Big Data เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มาและการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (extract, transform, and load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิควิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาดใหญ่ เทราไบต์ และ อาจจะเป็นระดับเพตาไบต์เลยก็มี ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

การจัดการข้อมูล ข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาดใหญ่
การจัดการเก็บข้อมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือ แบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้ เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูล หรือข้อมูลบางอัน ต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยม กันเป็นอย่างมาก

การวิเคราะห์ การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่า ก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุด ข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แสร้งสิ่งที่ค้นพบใหม่ ๆ ต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูล เหล่านั้นไปใช้งาน

2.2.1.3 เทคโนโลยีการประมวลผลข้อมูล

การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็น business intelligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analytics โดยใช้หลักการของ Data Science ความยากของการประมวลผลคือต้องการความเร็วในการ ประมวลผลข้อมูลที้นอกจากมีขนาดใหญ่แล้วบางครั้งยังเป็นข้อมูลที่ไม่มีโครงสร้าง ดังนั้นต้อง จึงมีการนำเทคโนโลยีหรือภาษาต่าง ๆ มาเพื่อให้สามารถประมวลผลข้อมูลได้ ซึ่งในบางครั้ง หน่วยงานอาจต้องพิจารณาต้องเลือกใช้

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและ ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ใน งานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และ การแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บ และตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบ ฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบ ความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บ ข้อมูลที่เกี่ยวข้อง

- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือ การนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการ สังเกตการณ์ที่ทันสมัย

- การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือ ฐานข้อมูล

- การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและ ตรรกะของข้อมูลขนาดใหญ่ เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้ ขั้นตอนการทำเหมืองข้อมูล ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็น ความรู้ ประกอบด้วยขั้นตอนดังนี้

1. Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาส และหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุ ผลลัพธ์ที่มีได้

2. Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่ เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

3. Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูล ที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลาามากที่สุดในทุกขั้นตอน เพราะ คุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูล ประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4. Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับ การเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5. Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับ วัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6. Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาใช้ประโยชน์ ตามวัตถุประสงค์ และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตามวัตถุประสงค์ที่ตั้ง ไว้หรือไม่

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่ง แรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถ ใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดย ปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชัน และองค์ประกอบอื่น ๆ ที่ ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการ ใช้งานหลักสำคัญในการ ออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อ ความหมาย เกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบ เว็บไซต์ ต้อง คำนึงถึง

1. ความเรียบง่าย ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้ สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มาก จนเกินไปทำให้วุ่นวาย

2. ความสม่ำเสมอ ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่น รูปแบบของหน้าสไตล์ของกราฟิก ระบบเนวิเกชันและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้ง เว็บไซต์

3. ความเป็นเอกลักษณ์ การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะของ องค์กร เพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ

4. เนื้อหาที่มีประโยชน์ เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควร จัดเตรียมเนื้อหาและข้อมูลที่ใช้ต้องการให้ถูกต้อง และสมบูรณ์ มีการปรับปรุงและเพิ่มเติมให้ ทันเหตุการณ์อยู่เสมอ เนื้อหาไม่ควรซ้ำกับเว็บไซต์อื่น จึงจะดึงดูดความสนใจ

5. ระบบเนวิเกชันที่ใช้งานง่าย ต้องออกแบบให้ผู้ใช้เข้าใจง่ายและใช้งาน สะดวก ใช้กราฟิกที่สื่อความหมายร่วมกับคำอธิบายที่ชัดเจน มีรูปแบบและลำดับของรายการที่ สม่ำเสมอ เช่น วางไว้ ตำแหน่งเดียวกันของทุกหน้า

6. ลักษณะที่น่าสนใจ หน้าตาของเว็บไซต์จะต้องมีความสัมพันธ์กับ คุณภาพขององค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่จะต้องสมบูรณ์ การใช้สี การใช้ ตัวอักษรที่อ่านง่าย สบายตา การใช้โทนสีที่เข้ากันลักษณะหน้าตาที่น่าสนใจนั้นขึ้นอยู่กับ ความชอบของแต่ละบุคคล

7. การใช้งานอย่างไม่จำกัด ผู้ใช้ส่วนใหญ่สามารถเข้าถึงได้มากที่สุด เลือกใช้เบราว์เซอร์ชนิดใดก็ได้ในการเข้าถึงเนื้อหาสามารถแสดงผลได้ทุกระบบปฏิบัติการและ ความละเอียดหน้าจอต่าง ๆ กันอย่างไม่มีปัญหาเป็นลักษณะสำคัญสำหรับผู้ใช้ที่มีจำนวนมาก

8. คุณภาพในการออกแบบ การออกแบบและเรียบเรียงเนื้อหาอย่างรอบคอบสร้างความรู้สึก ว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้

9. ลิงค์ต่าง ๆ จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการ ทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

2.2.3.2 องค์ประกอบของการออกแบบเว็บไซต์

1. โครงสร้างที่ชัดเจน ผู้ออกแบบเว็บไซต์ควรจัดโครงสร้างหรือจัด ระเบียบของข้อมูลที่ชัดเจนแยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์กันและให้อยู่ใน มาตรฐานเดียวกัน จะช่วยให้การใช้งาน และง่ายต่อการอ่านเนื้อหาของผู้ใช้

2 การใช้งานที่ง่าย ลักษณะของเว็บที่มีการใช้งานง่ายจะช่วยให้ผู้ใช้รู้สึก สบายใจต่อการอ่าน และสามารถทำความเข้าใจกับเนื้อหาได้อย่างเต็มที่ โดยไม่ต้องมาเสียเวลา อยู่กับการทำความเข้าใจ การใช้งานที่สับสนด้วยเหตุนี้ผู้ออกแบบจึงควรกำหนดปุ่มการใช้งานที่ ชัดเจน เหมาะสม โดยเฉพาะปุ่มควบคุมเส้นทางการเข้าสู่เนื้อหา (Navigation) ไม่ว่าจะเป็น เดินหน้า ถอยหลัง หากเป็นเว็บไซต์ที่มีเว็บเพจจำนวนมาก ควรจะจัดทำแผนผังของเว็บไซต์ (Site Map) ที่ช่วยให้ผู้ใช้ทราบว่า ตอนนี้อยู่ ณ จุดใด หรือเครื่องมือสืบค้น (Search Engine) ที่ ช่วยในการค้นหาหน้าที่ที่ต้องการ

3. การเชื่อมโยงที่ดี ลักษณะไฮเปอร์เท็กซ์ที่ใช้ในการเชื่อมโยง ควรอยู่ใน รูปแบบที่เป็นมาตรฐาน ทั่วไปและต้องระวังเรื่องของตำแหน่งในการเชื่อมโยง การที่จำนวนการ เชื่อมโยงมากและกระจัดกระจายอยู่ทั่วไปในหน้าอาจก่อให้เกิดความสับสน นอกจากนี้คำที่ใช้ สำหรับการเชื่อมโยงต้องเข้าใจง่ายมีความชัดเจนและไม่สั้นจนเกินไป นอกจากนี้ในแต่ละเว็บเพจที่สร้างขึ้นมามี จุดเชื่อมโยงกลับมายังหน้าแรกของเว็บไซต์ที่กำลังใช้งานอยู่ด้วย ทั้งนี้ เพื่อว่าผู้ใช้เกิดหลงทาง และไม่ทราบว่าจะทำอย่างไรต่อไปจะได้มีหนทางกลับมาสู่จุดเริ่มต้นใหม่ ระวังอย่าให้มีหน้าที่ไม่มีการเชื่อมโยง (Orphan Page) เพราะจะทำให้ผู้ใช้ไม่รู้จะทำอย่างไรต่อไป

4. ความเหมาะสมในหน้าจอ เนื้อหาที่น่าเสนอในแต่ละหน้าจอควรสั้น กระชับ และทันสมัย หลีกเลี่ยงการใช้หน้าจอที่มีลักษณะการเลื่อนขึ้นลง (Scrolling) แต่ถ้า จำเป็นต้องมี ควรจะให้ข้อมูลที่มีความสำคัญอยู่บริเวณด้านบนสุดของหน้าจอ หลีกเลี่ยงการใช้กราฟิกด้านบนของหน้าจอ เพราะถึงแม้จะดูสวยงาม แต่จะทำให้ผู้ใช้เสียเวลาในการได้รับ ข้อมูลที่ต้องการ แต่หากต้องมีการใช้ภาพประกอบก็ควรใช้เฉพาะที่มีความสัมพันธ์กับเนื้อหา เท่านั้น นอกจากนี้การใช้รูปภาพเพื่อเป็นพื้นหลัง (Background) ไม่ควรเน้นสีที่ฉูดฉาดมากนัก เพราะอาจจะไปลดความเด่นชัดของเนื้อหา ควรใช้ภาพที่มีสีอ่อน ๆ ไม่สว่างจนเกินไป รวมไปถึงการใช้เทคนิคต่าง ๆ เช่น ภาพเคลื่อนไหว หรือตัวอักษรวิ่ง (Marquees) ซึ่งอาจจะเกิด การรบกวนการอ่านได้ ควรใช้เฉพาะที่จำเป็นจริง ๆ เท่านั้นตัวอักษรที่นำมาแสดงบนจอภาพ ควรเลือกขนาดที่อ่านง่าย ไม่มีสีสั่นและลวดลายมากเกินไป

5. ความรวดเร็ว ความรวดเร็วเป็นสิ่งสำคัญประการหนึ่งที่ส่งผลต่อการ เรียนรู้ ผู้ใช้จะเกิดอาการเบื่อหน่ายและหมดความสนใจกับเว็บที่ใช้เวลาในการแสดงผลนาน สาเหตุสำคัญที่จะทำให้การแสดงผลนานคือการใช้ภาพกราฟิกหรือภาพเคลื่อนไหว ซึ่งแม้ว่าจะ ช่วยดึงดูดความสนใจได้ดี ฉะนั้นในการออกแบบจึงควรหลีกเลี่ยงการใช้ภาพขนาดใหญ่ หรือ ภาพเคลื่อนไหวที่ไม่จำเป็น และพยายามใช้กราฟิกแทนตัวอักษรธรรมดาให้น้อยที่สุด โดยไม่ ควรใช้มากเกินไปกว่า 2 – 3 บรรทัดในแต่ละหน้าจอ

2.2.3.3 โครงสร้างเว็บไซต์

1. เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็น โครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัด ด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การ เรียงลำดับตามตัวอักษร ดรชณี สารานุกรม หรืออภิธาน ศัพท์ โครงสร้างแบบนี้ เหมาะกับ เว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปที่ละหน้าทิศทางของการเข้าสู่ เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินหน้าถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อ

2. เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure) เป็นวิธีที่ดี ที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วน ต่าง ๆ และมีรายละเอียดย่อย ๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิ องค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับ

โครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของเว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียว นั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

3. เว็บที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มี ความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหา ของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้ เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่ใช่เป็นลักษณะเชิงเส้นตรง เนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้ ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมี ลักษณะที่เหมือนกัน และสามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมา บรรจลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับ โครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะไปสู่หน้าเนื้อหา (Topic Page) ที่แสดง รายละเอียดของหัวข้อนั้น ๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของ หัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับ ขึ้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจสร้างความยุ่งยากในการเข้าใจได้ และ อาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นต้องมีการ วางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจ เกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4. เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภท นี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้าง รูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วย ตนเอง การ เชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มี โครงสร้างที่แน่นอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหา ภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือ ไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้ จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการ เพื่อเลือกที่จะเข้าไปสู่หน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการ ท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือ ถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการ ปรับปรุง นอกจากนี้การเชื่อมโยง ระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและ เกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3.4 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่าง ชัดเจน การเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่าง

ระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการ ตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบน จอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดย มีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกละ ผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุดหรือพิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อ สร้างภาพบนจอขึ้นเรียกว่า บิตเดป (Bit-depth) ในภาษา HTML มีการกำหนดสีด้วยระบบ เลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6 หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรก แสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงิน

สีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการ รับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึก อารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึก สงบ บางสีให้ความรู้สึกตื่นตัวรุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทาง อารมณ์ มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็นสามารถแบ่ง ออกเป็น 3 กลุ่ม คือ

1. สีโทนร้อน (Warm Colors) เป็นกลุ่มสีที่แสดงถึงความสุข ความ ปลอดภัย ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มี ชีวิตชีวามากยิ่งขึ้น
2. สีโทนเย็น (Cool Colors) แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อย เป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มน้าวในระยะไกลได้
3. สีโทนกลาง (Neutral Colors) สีที่เป็นกลาง ประกอบด้วย สีดำ สีขาว สี เทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีกลาง ขึ้นมา

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง

CSS ย่อมาจาก Cascading Style Sheet มักเรียกโดยย่อว่า "สไตลชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการ ระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภท ตัวอักษร และการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยก เนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของ การแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผล ลัทธิของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือ ต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามา

ครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดยองค์กร World Wide Web Consortium หรือ W3C

ประโยชน์ของ CSS

1. CSS มีคุณสมบัติมากกว่า tag ของ html เช่น การกำหนดกรอบให้ข้อความ รวมทั้งสี รูปแบบของข้อความที่กล่าวมาแล้ว

2. CSS นั้นกำหนดที่ต้นของไฟล์ html หรือตำแหน่งอื่น ๆ ก็ได้ และสามารถมีผล กับเอกสารทั้งหมด หมายถึงกำหนด ครั้งเดียวจุดเดียวก็มีผลกับการแสดงผลทั้งหมด ทำให้ เวลาแก้ไขหรือปรับปรุงทำได้สะดวก ไม่ต้องไล่ตามแก้ tag ต่าง ๆ ทั่วทั้งเอกสาร

3. CSS สามารถกำหนดแยกไว้ต่างหากจาก ไฟล์เอกสาร html และสามารถ นำมาใช้ร่วม กับเอกสารหลายไฟล์ได้ การแก้ไขก็แก้เพียง จุดเดียวก็มีผลกับเอกสารทั้งหมด

CSS กับ HTML / XHTML นั้นทำหน้าที่คนละอย่างกัน โดย HTML / XHTML จะทำ หน้าที่ในการวางโครงสร้างเอกสารอย่างเป็นรูปแบบ ถูกต้อง เข้าใจง่าย ไม่เกี่ยวข้องกับการ แสดงผล ส่วน CSS จะทำหน้าที่ในการตกแต่งเอกสารให้สวยงาม เรียกได้ว่า HTML /XHTML คือส่วน coding ส่วน CSS คือส่วน design

2.2.5 ทฤษฎีเกี่ยวกับการ visualization

Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการ ทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือ การสร้าง ภาพ ให้เป็นบันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณ ข้อมูล ช่วยลดภาระการ คำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อ นำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมา เพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจ อย่าง ถูกต้อง รวดเร็ว และได้ผลเป็นอย่างดี

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพใน ความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้ สามารถเข้าใจได้ง่าย ในแบบของกราฟ แผนภูมิ

2.2.5.1 Tools Data Visualization

Tools ที่ นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts, Data wrapper และอื่น ๆ อีกมากมาย



ภาพที่ 2.9 Tools Data Visualization

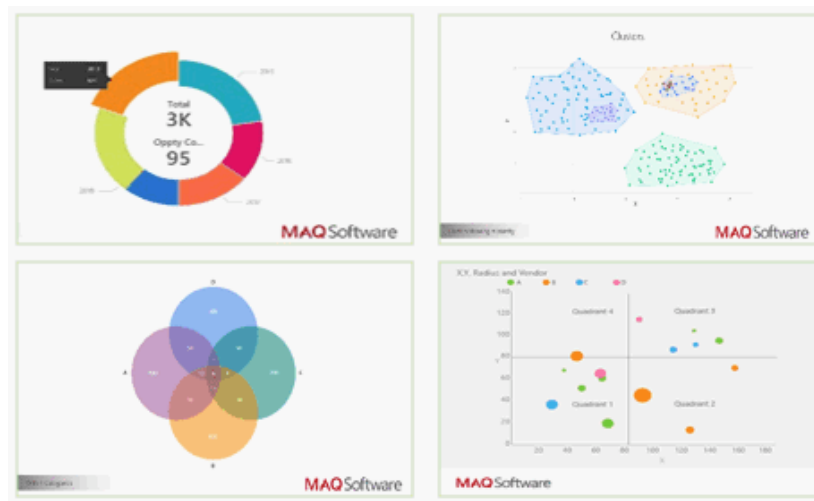
2.2.5.2 รูปแบบในการใช้ Data Visualization

การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) เราใช้กราฟที่แสดงผล แบบทิศทางหรือแนวโน้ม เพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) รวมถึงเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart, Radar Chart, Area Chart เป็นต้น



ภาพที่ 2.10 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

การนำเสนอแบบกลุ่มข้อมูล (Classification) เป็นการนำเสนอโดยนำข้อมูล มาจัดเป็นกลุ่ม ๆ เช่น Donut Chart, Ring Chart, Pie Chart,



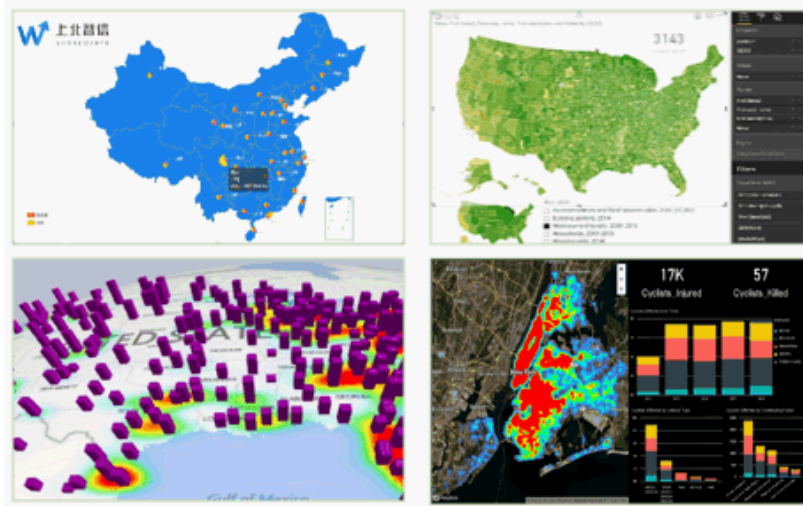
ภาพที่ 2.11 การนำเสนอแบบกลุ่มข้อมูล (Classification)

การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการ นำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่นเทียบกับปีที่แล้ว(YoY) เปรียบเทียบกับเป้า ที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสมและมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น



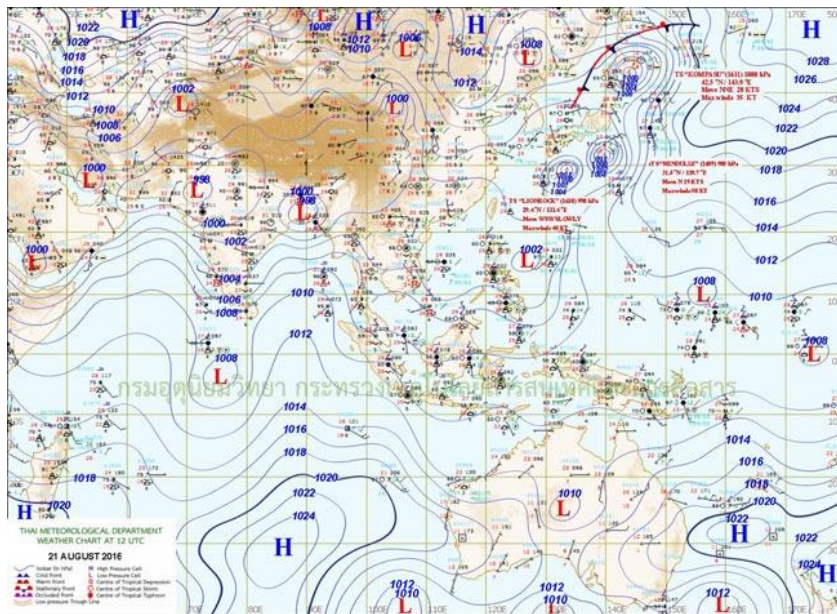
ภาพที่ 2.12 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับนำเสนอข้อมูลบน แผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อFocus กลุ่ม ลูกค้าในแต่ละพื้นที่ที่เราสนใจ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ 2.13 การนำเสนอรูปแบบแผนที่ (Geographical)

กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics) เราสามารถที่จะใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคต และนำเสนอข้อมูลที่ได้ในรูปแบบของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.14 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics)

ข้อดีของ Data Visualization

- ข้อมูลที่มีปริมาณมากมาหายมหาศาลยากต่อการเข้าใจ เราสามารถทำให้ เข้าใจง่ายได้ด้วยรูปภาพ
- ช่วยจัดระเบียบความคิดวิเคราะห์ข้อมูลให้มีความน่าเชื่อถือ
- ประหยัดเวลาในการนำเสนอ

2.2.6 ทฤษฎีเกี่ยวกับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

ความสัมพันธ์ระหว่างตัวแปร การศึกษาความสัมพันธ์ระหว่างตัวแปรอันเป็นเป้าหมายของข้อสรุปงานวิจัยนั้นก็เพื่อให้เกิดความรู้ ความเข้าใจที่จะสามารถบรรยาย อธิบาย ตลอดจนควบคุมสิ่งต่าง ๆ ได้นั้น ค่าสถิติที่นำมาใช้บ่อยมาก คือ ค่าสัมประสิทธิ์สหสัมพันธ์ (r_{xy}) ซึ่งใช้ได้กับตัวแปร x และ y ที่มีมาตรการวัดแบบอันตรภาคขึ้นไป แต่ยังมีค่าสถิติอีกหลายตัวที่ใช้หาความสัมพันธ์ระหว่างตัวแปรที่อยู่ในเงื่อนไขที่ต่างออกไป ก่อนที่จะอธิบายรายละเอียดของสถิติที่ใช้หาความสัมพันธ์ระหว่างตัวแปรนั้น เพื่อให้เกิดความเข้าใจในการเลือกใช้สถิติเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรที่ชัดเจนขึ้น จึงมีความจำเป็นต้องเข้าใจในเรื่องมาตรการวัดของตัวแปร ซึ่งสรุปได้ ดังนี้

การแบ่งประเภทของข้อมูลตามมาตรการวัด แบ่งเป็น

1. มาตรการวัดแบบนามบัญญัติ (Nominal data) เป็นการจำแนกลักษณะของข้อมูลที่ได้ ออกเป็นประเภทต่าง ๆ หรือเป็นพวกๆ โดยจัดลักษณะที่เหมือนกันไว้ด้วยกัน เช่น ตัวแปร เพศ เชื้อชาติ สถานภาพสมรส เป็นต้น การจำแนกลักษณะของข้อมูลของตัวแปรเป็น 2 ลักษณะ เรียกว่าตัวแปรทวิภาค (Dichotomous Variable) มีรูปแบบในการจำแนกที่แตกต่างกันได้ 2 ลักษณะ คือ ตัวแปรทวิภาคแท้ (True dichotomous Variable) และตัวแปรทวิภาคจำแนกตามเกณฑ์ (Artificially dichotomous Variable) โดยพิจารณาจากเกณฑ์การจำแนกที่มีอยู่แล้ว กับเกณฑ์ที่ต้องสร้างขึ้น ถ้าเกณฑ์ในการแบ่งตัวแปร ออกเป็น 2 ลักษณะ เป็นเกณฑ์ที่มีอยู่แล้วเช่น ตัวแปรเพศ แบ่งเป็น หญิงและชาย ก็จัดว่าเป็นทวิภาคแท้ แต่ถ้าเป็นเกณฑ์ที่ต้องสร้างขึ้นเช่นการสอบได้ - ตกของนักเรียนก็จัดว่าเป็นทวิภาคจำแนกตามเกณฑ์

2. มาตรการวัดแบบอันดับ (Ordinal data) เป็นการกำหนดลักษณะของข้อมูลที่ได้ ออกเป็นอันดับที่บอกความมากน้อยระหว่างกันได้ เช่นลำดับที่ของนักเรียนมารยาทดี ค่าลำดับที่ 1, 2, 3 สามารถบอกได้ว่าใครมารยาทดีกว่าใคร แต่ไม่สามารถบอกได้ว่าคนที่ได้มารยาทดีลำดับที่ 1 ดีกว่าลำดับที่ 2 อยู่เท่าไร และไม่สามารถบอกได้ว่าความแตกต่างระหว่างคนที่ได้มารยาทดีลำดับที่ 1 และ 2 จะเท่ากับ ความแตกต่างระหว่างคนที่ได้มารยาทดีลำดับที่ 2 และ 3 หรือช่วงความห่างของค่าตัวแปรแต่ละค่าไม่เท่ากัน

3. มาตรการวัดแบบอันตรภาค (Interval data) เป็นการกำหนดตัวเลขให้กับลักษณะของข้อมูลตามความมากน้อย โดยตัวเลขที่กำหนดสามารถบอกความมากน้อยระหว่างกันแล้วยังมีช่วงห่างระหว่างค่าที่เท่ากันด้วย แต่ค่าศูนย์ที่กำหนดตามมาตรการวัดนี้ไม่ใช่ศูนย์แท้ ตัวอย่าง เช่น คะแนน อุณหภูมิ เป็นต้น ค่าของอุณหภูมิ 80°C สูงกว่าอุณหภูมิ 50 °C อยู่ 30°C แต่อุณหภูมิ 0 °C มิได้แปลว่าไม่มีความร้อน ความจริงมีความร้อนระดับหนึ่งแต่ถูกสมมุติให้เป็น 0 °C

4. มาตรการวัดแบบอัตราส่วน (ratio data) เป็นการกำหนดตัวเลขให้กับลักษณะของข้อมูลเดียวกับ มาตรการวัดแบบอันตรภาค แต่มาตรการวัดระดับนี้จะมีค่า 0 ที่แท้จริงด้วย เช่น อายุ รายได้ น้ำหนัก ส่วนสูง เป็นต้น ส่วนสูง 0 เซนติเมตรก็แปลว่าไม่มีความสูงเลย

เพื่อให้เห็นภาพรวมของสถิติที่ใช้ในการหาความสัมพันธ์ จึงขอเสนอตารางสรุประเบียบวิธีวัดความสัมพันธ์จำแนกตามมาตรวัดตัวแปรก่อนแล้วตามด้วยรายละเอียดของแต่ละวิธีต่อไป

สัมประสิทธิ์ พี (Phi correlation)

เมื่อต้องการหาความสัมพันธ์ระหว่างตัวแปรที่มีมาตรการวัดเป็นทวิวิภาคแท้ทั้งคู่ หรือตัวหนึ่งเป็นทวิวิภาคแท้ อีกตัวหนึ่งเป็นทวิวิภาคจำแนกตามเกณฑ์ จะต้องใช้สัมประสิทธิ์พี (ϕ) ซึ่งจะได้ขนาดความสัมพันธ์ว่ามีมากน้อยเพียงใด การหาความสัมพันธ์ของสองตัวแปรเช่นนี้อาจหาได้โดยใช้สถิติ

χ^2 แต่ χ^2 จะบอกได้แต่เพียงว่ามีความสัมพันธ์หรือไม่มีความสัมพันธ์เท่านั้น ไม่ได้บอกขนาดความสัมพันธ์

$$\phi = \frac{bc - ad}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

ทดสอบความมีนัยสำคัญโดยใช้ χ^2 หรือ t-test

$$\chi^2 = N \phi^2$$

$$t - test = r \sqrt{\frac{N - 2}{1 - r^2}}$$

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 แบบจำลองแบบจำลองอนุกรมเวลา Time series

2.3.1.1 การพยากรณ์

1. Simple moving average

ค่าเฉลี่ยเคลื่อนที่ หมายถึง เอนทิตีที่แสดงค่าเฉลี่ยของเส้นกราฟของราคา ที่ไหลไปตามเวลาหนึ่งๆ โดยเราจะใช้ค่าเฉลี่ยเคลื่อนที่นี้ ในการวิเคราะห์เชิงเทคนิคเพื่อวิเคราะห์แนวโน้มในระยะยาวของราคา หรือราคาสินค้าใด ๆ

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}$$

A = ค่าในแต่ละช่วงเวลา n

n = จำนวนช่วงเวลา

2. Weighted moving average

ค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก คือรูปแบบหนึ่ง ที่พัฒนามาจาก ค่าเฉลี่ยเคลื่อนที่ แบบเดิม สำหรับ Weighted Moving Average หรือ ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก จะต่างจากค่าเฉลี่ยเคลื่อนที่ (Moving average) แบบธรรมดา ด้วยการใส่ค่าน้ำหนักของข้อมูลแต่ละตัวเข้าไปด้วยโดยข้อมูลตัวที่เป็นปัจจุบันที่สุดจะได้ค่าน้ำหนักที่มากที่สุดและข้อมูลตัวถัดไปก็จะได้ค่าน้ำหนักที่ลดลงไปตามลำดับ

$$WMA = \frac{A_1 \times n + A_2 \times (n - 1) + \dots + A_n \times (n - n)}{\frac{n \times (n + 1)}{2}}$$

A = ค่าในแต่ละช่วงเวลา n

n = จำนวนช่วงเวลา

3. Exponential smoothing

การปรับให้เรียบแบบเอ็กซ์โปเนนเชียลเป็นกฎของเทคนิคห้วแม่มือสำหรับการทำให้ข้อมูลอนุกรมเวลาให้เรียบโดยใช้ฟังก์ชันหน้าตาต่างเอ็กซ์โปเนนเชียล ในขณะที่ค่าเฉลี่ยเคลื่อนที่อย่างง่ายการสังเกตที่ผ่านมามีน้ำหนักเท่ากันฟังก์ชันเลขชี้กำลังใช้เพื่อกำหนดน้ำหนักที่ลดลงแบบทวีคูณเมื่อเวลาผ่านไป

$$F_t = \alpha A_{t-1} + (1 - \alpha)F_{t-1}$$

F_t = ค่าพยากรณ์ความต้องการใหม่

F_{t-1} = ค่าพยากรณ์ที่ผ่านมา

α = ค่าคงที่ปรับเรียบ (Smoothing constant) ($0 \leq \alpha \leq 1$)

A_{t-1} = ค่าที่เกิดขึ้นจริงในช่วงที่ผ่านมา

2.3.1.2 การวัดผลการพยากรณ์

Mean absolute percentage error

ค่าเฉลี่ยเปอร์เซ็นต์ข้อผิดพลาดสัมบูรณ์หรือที่รู้จักกันในชื่อค่าเบี่ยงเบนเปอร์เซ็นต์สัมบูรณ์เป็นตัวชี้วัดความแม่นยำในการทำนายของวิธีการพยากรณ์ในสถิติเช่นในการประมาณแนวโน้มยังใช้เป็นฟังก์ชันการสูญเสียสำหรับปัญหาการถดถอยในการเรียนรู้ของเครื่อง มันมักจะแสดงความแม่นยำเป็นอัตราส่วนที่กำหนดโดยสูตร

$$MAPE = \frac{\sum \frac{A - F}{A} \times 100}{N}$$

2.3.2 การแสดงผลแบบ visualization

2.3.2.1 การแสดงผลแบบ visualization ด้วย Tableau Public

ทาโบลัว (Tableau) Business Intelligence (BI) Software คือซอฟต์แวร์เพื่อ การวิเคราะห์ข้อมูลอันหลากหลายเป็น Software ในระดับต้นๆของโลกที่ทำเรื่อง Data Visualization Tableau ทำหน้าที่แปลงข้อมูล (Data) ให้เป็นภาพ (Visualization) สามารถนำ ข้อมูลออกมาเป็น Chart แบบต่าง ๆ ได้อย่างง่ายดายรวมถึงการนำหลาย Chart มาสร้างเป็น Dashboard และ Story Teller ก็ได้เช่นกัน รองรับการใช้งานถึง

ข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูลจากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL, Oracle, MySQL เป็นต้น

Tableau นั้นถูกพัฒนาอย่างต่อเนื่องเพื่อให้แสดงผลได้บนทุก Platform อย่างง่ายดาย ไม่ว่าจะ Desktop, Mobile หรือแม้แต่ Browser โดยองค์กรสามารถเลือก ติดตั้งใช้งานภายในองค์กรเองก็ได้ หรือเช่าใช้ผ่านบริการ Tableau Cloud ก็ได้เช่นกัน ทำให้การ ทำ Business Intelligence และ Data Analytics ไม่ใช่เรื่องยากอีกต่อไป และง่ายขึ้นจนถึงขนาด ผู้บริหารก็สามารถใช้งานด้วยตัวเองได้

2.3.2.2 ลักษณะเด่นของ Tableau Public

- รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลง รายงานได้ตามความต้องการ - รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล และสามารถนำข้อมูล จากฐานข้อมูลในองค์กรนั้นมาใช้งาน เช่น Excel, Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot

- รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการ นำเสนอต่อผู้บริหาร

- การเข้าถึงหรือการใช้งาน สามารถทำได้โดยง่าย ตั้งแต่ผู้บริหาร ระดับสูง ถึง พนักงานระดับปฏิบัติการ ทั้งนี้ทั้งนั้น สามารถเพิ่มทักษะและแนวคิดให้กับ พนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น

- สิ่งที่สำคัญ คือ Engine ของ Tableau Software นั้น พัฒนาจาก VizQL Technology เป็นเทคโนโลยีที่ช่วยให้เห็นภาพของข้อมูลขนาดใหญ่ได้เพียงการลากและวาง (Drag and Drop) โดยใช้ข้อมูลจากฐานข้อมูล ตอบสนองในรูปแบบของ กราฟฟิก

2.4 วรรณกรรมที่เกี่ยวข้อง

ลักขณา ฤกษ์เกษม (2558) การวิจัยนี้เป็นการศึกษาและเปรียบเทียบวิธีการพยากรณ์ สำหรับการพยากรณ์ความต้องการการใช้ผ้าสำหรับการผลิตชุดปฏิบัติการสำหรับห้องสะอาดโดยใช้ข้อมูลจำนวน 12 เดือนของ พ.ศ. 2556 เพื่อหาวิธีพยากรณ์ล่วงหน้าที่เหมาะสมที่สุด วิธีการพยากรณ์ที่ใช้ในงานวิจัยนี้มี 4 วิธี คือ วิธีถัวเฉลี่ยเคลื่อนที่ (Moving Average) วิธีปรับเรียบเอ็กซ์โพเนนเชียลอย่างง่าย (Simple Exponential Smoothing) วิธีพยากรณ์แบบฤดูกาลของวินเตอร์ (Winter's Linear and Seasonal Exponential Smoothing) และวิธีการพยากรณ์โดยการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Regression Analysis) การเลือกวิธีการพยากรณ์ที่เหมาะสมพิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percent Error, MAPE) ที่ต่ำที่สุด ผลการศึกษาพบว่าวิธีการพยากรณ์ที่ให้รูปแบบที่เหมาะสมที่สุด คือวิธีการพยากรณ์โดยการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายให้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ที่น้อยที่สุด

นรวัดณ์ เหลืองทอง และ นันทชัย กานตานันทะ (2558) งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการพยากรณ์ผลผลิตการเกษตรของพืช 4 ชนิด ได้แก่ ข้าวนาปี ข้าวนาปรัง มันสำปะหลัง และสับปะรด ในจังหวัดที่มีผลผลิตสูงสุด 3 อันดับแรกของประเทศไทย และเลือกวิธีการพยากรณ์ผลผลิตการเกษตรที่เหมาะสม โดยทำการเปรียบเทียบผลการพยากรณ์ของวิธีอนุกรมเวลา 4 วิธี คือ วิธีการปรับเรียบแบบเอกซ์โพเนนเชียลอย่างง่ายวิธีการปรับเรียบแบบเอกซ์โพเนนเชียลสองครั้ง วิธีโฮลท์และวินเทอร์แบบบวก และวิธีโฮลท์และวินเทอร์แบบคูณ ความแม่นยำของการพยากรณ์จะถูกเปรียบเทียบโดยใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย จากผลการศึกษาพบว่า ข้าวนาปีมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ในช่วง 6.74 ถึง 12.59 เปอร์เซ็นต์ ข้าวนาปรังมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ในช่วง 4.65 ถึง 15.27 เปอร์เซ็นต์ มันสำปะหลังมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ในช่วง 7.28 ถึง 15.10 เปอร์เซ็นต์ และสับปะรดมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ในช่วง 8.45 ถึง 14.89 เปอร์เซ็นต์ โดยวิธีการพยากรณ์ที่เหมาะสมกับพืชแต่ละชนิดในจังหวัดที่ศึกษาสามารถสรุปได้ดังตารางที่ 7 ของบทความค่าหลัก การพยากรณ์ วิธีอนุกรมเวลา ผลผลิตการเกษตร

เกาทัณฑ์ทอง, นุชจรินทร์ (2561) การศึกษามีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อมูลค่าการนำเข้าและส่งออกทองคำของประเทศไทย โดยใช้ข้อมูลทุติยภูมิแบบรายเดือน ตั้งแต่เดือนมกราคม 2548 ถึงกันยายน 2557 นำมาวิเคราะห์โดยใช้แบบจำลองทางเศรษฐมิติประเภทตัวแบบสมการต่อเนื่องกำลังสองน้อยที่สุดสองชั้น พบว่า ปัจจัยที่มีผลต่อการนำเข้าทองคำของประเทศไทยในทิศทางเดียวกัน ได้แก่ อัตราเงินเฟ้อพื้นฐาน, การเปิดซื้อขายโกลด์ฟิวเจอร์ส และดัชนีตลาดหลักทรัพย์แห่งประเทศไทย โดยมีระดับนัยสำคัญ 0.05 ตามลำดับ ส่วนราคาทองคำในตลาดโลก และอัตราแลกเปลี่ยนบาทเทียบเหรียญสหรัฐของเดือนก่อนมีความสัมพันธ์ในทิศทางตรงข้าม ที่ระดับนัยสำคัญ 0.10 และ 0.05 ตามลำดับ สำหรับปัจจัยที่มีผลต่อการส่งออกทองคำของประเทศไทย ได้แก่ ราคาทองคำในตลาดโลกมีความสัมพันธ์ในทิศทางเดียวกัน ที่ระดับนัยสำคัญ 0.05 ในขณะที่อัตราดอกเบี้ยนโยบายธนาคารแห่งประเทศไทยดัชนีผลผลิตอุตสาหกรรม และการเปิดซื้อขายโกลด์ฟิวเจอร์สมีความสัมพันธ์ในทิศทางตรงข้าม ที่ระดับนัยสำคัญ 0.10

วิระ จิริกิจอนุสรณ์ , ชุสิทธิ์ จรัสกุลชัย (2553) ภาษีมูลค่าเพิ่มเป็นหนึ่งในแหล่งรายได้ของภาครัฐที่นำมาจัดทำงานงบประมาณในการพัฒนาประเทศ ประสิทธิภาพของการพยากรณ์รายได้จากภาษีมูลค่าเพิ่มจะช่วยให้ภาครัฐสามารถวางแผนงบประมาณอย่างมีประสิทธิภาพด้วย เทคนิคของเหมืองข้อมูลได้ถูกนำมาใช้ในงานประยุกต์หลายแขนงรวมทั้งการศึกษาวิจัยคาดการณ์ภาษีมูลค่าเพิ่ม อย่างไรก็ตามปัจจัยหรือตัวแปรที่ใช้ในการศึกษายังไม่ครอบคลุมทั้งหมดงานวิจัยฉบับนี้ศึกษาเพิ่มเติมปัจจัย อื่น ๆ ที่อาจส่งผลกระทบต่อคาดการณ์ และงานวิจัยนี้สร้างแบบจำลองด้วยเงื่อนไขที่หลากหลาย จากการศึกษาวิจัยพบว่าแบบจำลอง Multilayer perceptron ที่ใช้ตัวแปร ค่าดัชนีราคาผู้บริโภค ดัชนีสินค้านำเข้า ดัชนีการอุปโภคบริโภคภาคเอกชน ดัชนีการลงทุนภาคเอกชน รายจ่ายภาครัฐ จะให้ค่าการคาดการณ์ผลการจัดเก็บภาษีมูลค่าเพิ่ม ที่มีค่าผิดพลาดในการคาดการณ์เฉลี่ยน้อยที่สุด และมีการคาดการณ์ถูกต้องเฉลี่ยคิดเป็น 98 %

รัตนา สุวรรณวิชณี (2560) ระบบบริหารจัดการงานวิจัยแห่งชาติ (National Research Management System : NRMS) ของสำนักงานคณะกรรมการ-วิจัยแห่งชาติ (วช.) นำมาใช้งานในการบริหารจัดการข้อมูลงานวิจัยและงบประมาณวิจัยของหน่วยงานภาครัฐและแหล่งทุนของหน่วยงานต่าง ๆ แต่ทาง วช. ยังขาดเครื่องมือในการวิเคราะห์ข้อมูลจำนวนมากในระบบ NRMS ให้ได้ทันเวลาและนำเสนอข้อมูลในหลากหลายมิติที่ตรงตามความต้องการใช้ประโยชน์ของผู้ใช้ เพื่อนำเสนอรัฐบาลหรือหน่วยงานนโยบายสำหรับการตัดสินใจเชิงนโยบาย บทความนี้จึงนำเสนอการพัฒนาระบบรายงานรูปแบบหลายมิติเพื่อสนับสนุนการตัดสินใจเชิงนโยบายของสำนักงานคณะกรรมการวิจัย-แห่งชาติ โดยใช้ระบบข่าวกรองทางธุรกิจ (Business Intelligence : BI) พัฒนารายงานรูปแบบหลายมิติตามฟิลด์ เช่น จำแนกตามปี แหล่งทุน กระทรวง หน่วยงาน นโยบายและยุทธศาสตร์การวิจัย และประเด็นการวิจัย เป็นต้น อีกทั้งการนำเสนอรายงานลักษณะแดชบอร์ด (Dashboard) สนับสนุนการตัดสินใจเชิงนโยบายของเจ้าหน้าที่และผู้บริหาร วช. เพื่อให้มองเห็นภาพรวมข้อมูลวิจัยของประเทศได้ การพัฒนาระบบรายงานงานวิจัยนี้ใช้ซอฟต์แวร์ทาบิล (Tableau) ในการพัฒนาระบบรายงาน ผลการประเมินผู้ใช้งานระบบที่พัฒนาขึ้น พบว่าเจ้าหน้าที่ วช.ที่มีความพึงพอใจต่อระบบอยู่ในระดับพอใจ ทำให้เจ้าหน้าที่ วช. สามารถนำเสนอรายงานต่อผู้บริหารได้รวดเร็วขึ้น และหลากหลายมิติมากขึ้น

พูลผล สือเสาวลักษณ์ (2561) การผนวกเทคโนโลยีสารสนเทศและการวิวัฒนาการทางวิทยาศาสตร์ข้อมูล (Data science) ได้พัฒนาขึ้นโดยใช้ระบบคอมพิวเตอร์สำหรับการพยากรณ์ราคาหุ้นที่ซื้อขายในตลาดหลักทรัพย์แห่งประเทศไทย ตามหลักนั้นลักษณะเฉพาะของสภาวะแวดล้อมมักมีอิทธิพลต่อราคาหุ้นในแต่ละประเภทธุรกิจโดยมีนัยสำคัญที่แตกต่างกันดังนั้นจึงมีความจำเป็นที่จะต้องใช่วิธีการที่แตกต่างกันในการพยากรณ์ราคาหุ้น งานวิจัยชิ้นนี้นำเสนอระบบการพยากรณ์ราคาหุ้น (Stock forecasting system) โดยการใช้เทคนิค 3 แบบของวิธีการปรับให้เรียบ (Smoothing methods) คือ เทคนิคค่าเฉลี่ยเคลื่อนที่ (Moving average) เทคนิคค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weighted moving average) และเทคนิคปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (Exponential smoothing) ซึ่งทุกเทคนิคที่ใช้นำเสนอกระบวนการพยากรณ์ราคาซื้อขายทั้งหมด 5 รูปแบบ ได้แก่ ราคาเปิด ราคาปิดราคาสูงสุด ราคาต่ำสุด และราคาปัจจุบัน (Real time) ระบบนี้ได้ออกแบบให้มีความยืดหยุ่น เพื่อให้นักลงทุนสามารถกำหนดเงื่อนไขที่จำเป็นต่อระบบได้ด้วยตนเองสำหรับแต่ละเทคนิคของการพยากรณ์ แทนที่ผู้เขียนโปรแกรมจะเป็นผู้กำหนดเงื่อนไขเหล่านั้นแต่เพียงผู้เดียว

ประโยชน์ของการใช้ระบบนี้คือมีการนำเสนอข้อมูลทั้งในเชิงพรรณนาสถานการณ์ของตลาดหุ้นที่ซื้อขายในตลาดตามความเป็นจริงและข้อมูลเชิงวิเคราะห์ (Diagnosis) โดยการพยากรณ์ราคาหุ้นนี้ได้ตอบสนองผู้ลงทุนเพื่อนำไปใช้ประกอบการพิจารณา

วิไลลักษณ์ วงศ์ชัย (2560) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ ศึกษาต้นทุนและผลตอบแทนจากการปลูกข้าวโพดเลี้ยงสัตว์ของเกษตรกรในตำบลสถาน อำเภอเชียงของ จังหวัดเชียงราย การวิจัยใช้วิธีเก็บรวบรวมข้อมูลโดยใช้เครื่องมือ คือ แบบสอบถาม สอบถามเกษตรกรผู้ปลูกข้าวโพดเลี้ยงสัตว์ในตำบลสถาน อำเภอเชียงของ จังหวัดเชียงราย ที่สามารถเก็บเกี่ยวผลผลิตได้ในปี พ.ศ. 2559 จำนวน 131 ราย โดยการเลือกแบบเจาะจง ผลการวิจัย พบว่า มีต้นทุนรวมเฉลี่ย 4,229.48 บาทต่อไร่ มีกำไรสุทธิเฉลี่ย 5,311.29 บาทต่อไร่ ต้นทุนการปลูกข้าวโพดเลี้ยงสัตว์ประกอบด้วย 3 ส่วนที่สำคัญ ได้แก่ ต้นทุนค่าวัสดุคิด 13% ต้นทุนค่าแรงงาน 32% และต้นทุนค่าใช้จ่ายการผลิต 55% อัตรากำไรสุทธิต่อยอดขาย 20.36%

ภัทราวดี ธิเลอร์ (2563) ในยุคของข้อมูลข่าวสารที่มีอยู่อย่างล้นหลาม ผู้บริโภคสื่อร้องหาความถูกต้องและความน่าเชื่อถือจากข่าวสารมากขึ้น “นักข่าวเชิงข้อมูล” หรือ Data Journalist จึงเป็นคำตอบของการทำงานสื่อสารมวลชนในยุคปัจจุบัน เนื่องจากต้องมีความรู้ความสามารถในการค้นหา จัดการ และประมวลผลข้อมูลได้ หรือต้องทำงานร่วมกับหลายอาชีพมากขึ้น องค์กรกระจายเสียงและแพร่ภาพสาธารณะแห่งประเทศไทย (ส.ส.ท.) หรือสถานีโทรทัศน์ไทยพีบีเอสร่วมมือกับสมาคมนักข่าวนักหนังสือพิมพ์แห่งประเทศไทย (Thai Journalist Association) จัดอบรมภายในชื่อว่า หลักสูตรการพัฒนานักสื่อสารข้อมูลเชิงลึก (Thai PBS Data Journalism In -House Training) เพื่อเพิ่มทักษะการผลิตเนื้อหาให้มีคุณภาพ สร้างความแตกต่าง และสามารถบูรณาการข้อมูลเพื่อนำไปใช้งานกับงานด้านอื่น ๆ รวมถึงส่งเสริมให้พนักงาน ส.ส.ท. นำกระบวนการวารสารศาสตร์ข้อมูล (Data Journalism) มาใช้ในกระบวนการทำงานได้อย่างมีประสิทธิภาพ หลักสูตรการพัฒนานักสื่อสารข้อมูลเชิงลึก (Thai PBS Data Journalism In -House Training) ใช้แนวปฏิบัติที่ดีในการจัดอบรมเพื่อพัฒนาทักษะการรายงานข่าวเชิงข้อมูล หรือ KTCP Model ในการจัดการเนื้อหาในการอบรม ซึ่งแบ่งเป็น 4 ขั้นตอนดังต่อไปนี้ 1) Knowledge ให้ความรู้พื้นฐาน สร้างความเข้าใจ 2) Cheat Sheet เรียนรู้เทคนิคการทำงานของโปรแกรมการคำนวณแบบสเปรดชีต (Spreadsheet software) ในขั้นต้น ขั้นกลาง และขั้นสูง 3) Tip Visual เข้าใจหลักการออกแบบ Visual ในงานข่าวที่ถูกต้องและ 4) Partnership การทำงานร่วมกัน โดยนำข้อมูลจากทั้ง 3 ขั้นตอนข้างต้นมาเขียนข่าว เล่าเรื่องด้วยภาพ ด้วยหลักการเล่าเรื่องที่นำเสนอจับความใช้ข่าว เรื่องผลงานข่าว 365 วัน PM 2.5 เพื่อแสดงให้เห็นว่าการใช้ KTCP Model สามารถเสริมทักษะที่เป็นจุดอ่อนของนักข่าว ไม่ว่าจะเป็นขั้นตอนการกรองข้อมูล (Filter) ที่ต้องใช้โปรแกรมการคำนวณแบบสเปรดชีต ในการวิเคราะห์ข้อมูล หรือแม้แต่ขั้นตอนการแปลงข้อมูลให้เป็นภาพ (Visualization) ที่มีการสอนการออกแบบอย่างง่าย เครื่องมือออนไลน์ที่ประหยัดเวลา ภายใต้คอนเซ็ปต์การออกแบบที่ว่า “ไม่จำเป็นต้องสวยงาม แต่ต้องง่าย และไม่ซับซ้อนในการตีความ” ส่งผลให้ผู้เข้าอบรมสามารถผลิตผลงานเชิงข้อมูลได้จริง ตอบโจทย์ที่ว่า นักข่าวยุคใหม่ต้องมี Multi-Skills และสามารถเล่าเรื่องที่มาจากชุดข้อมูลมหาศาล (Data Set) ได้อย่างมีประสิทธิภาพนั่นเอง

สุทธิวัฒน์ แสงคล้าย (2558) วัตถุประสงค์ของงานวิจัยนี้คือการพิจารณาความแตกต่างของอัตราการมรณะในแต่ละรุ่นเกิด ขั้นตอนการศึกษาคือการเลือกตัวแบบอัตราการมรณะที่เหมาะสมที่สุดกับอัตราการมรณะไทยจาก 5 ตัวแบบ คือ ตัวแบบของลี คาร์เตอร์(Lee-Carter model : LC) ตัวแบบของเรนชอร์และฮาร์เบอร์แมน เรนชอ (Renshaw and Haberman model: RH) ตัวแบบอายุ-เวลา-รุ่นประชากร (Age-Period-Cohort model: APC) ตัวแบบของแคร์น เบรค ดอร์ (Cairns-Blake-Dowd model : CBD) และตัวแบบของแคร์น เบรค ดอร์ แบบพิจารณาผลกระทบรุ่น (Generalized Cairns-Blake-Dowd model: GCBD) ข้อมูลที่ใช้ศึกษาคือข้อมูลจำนวนประชากรและจำนวนการตาย แยกอายุ แยกเพศ ของปี พ.ศ. 2506-2557 เกณฑ์ที่ใช้ในการเลือกตัวแบบคือค่าเฉลี่ยร้อยละความคลาดเคลื่อนสัมบูรณ์(Mean Absolute percentage error :MAPE) และค่าเกณฑ์สารสนเทศของเบย์(Bayes Information Criterion :BIC) หลังจากเลือกตัวแบบอัตราการมรณะที่เหมาะสมแล้ว งานวิจัยนี้ได้หาตัวแบบออริมา(Autoregressive integrated moving average: ARIMA) ที่เหมาะสมสำหรับค่าประมาณพารามิเตอร์ของตัวแบบอัตราการมรณะที่เลือกได้ และพยากรณ์อัตราการมรณะไปข้างหน้า 70 ปี หลังจากนั้นค่าพยากรณ์ของอัตราการมรณะของรุ่นเกิดที่แตกต่างกันได้ถูกใช้ในการคำนวณมูลค่าเบี้ยประกันภัยของผลิตภัณฑ์ประกันชีวิตแบบเงินรายปีตัวอย่าง และเปรียบเทียบมูลค่าของเบี้ยประกันภัยเพื่อสรุปผลกระทบของการเปลี่ยนแปลงของอัตราการมรณะในแต่ละรุ่นเกิดที่มีต่อมูลค่าของเบี้ยประกันภัย ผลการศึกษาพบว่า ตัวแบบ RH เป็นตัวแบบที่ประมาณค่าอัตราการมรณะของไทยได้ดี ค่าอัตราการมรณะที่พยากรณ์จากตัวแบบ RH แสดงให้เห็นว่า อัตราการมรณะมีแนวโน้มลดลงในแต่ละรุ่นเกิด โดยอัตราการมรณะของเพศหญิงมีแนวโน้มลดลงเร็วกว่าของเพศชาย อัตราการมรณะตามรุ่นเกิดลดลงช้ากว่าอัตราการมรณะตามปีปฏิทิน และมูลค่าของเบี้ยประกันภัยที่คำนวณจากอัตราการมรณะของคนที่เกิดรุ่นหลังมีแนวโน้มเพิ่มสูงขึ้นเนื่องจากการลดลงของอัตราการมรณะในแต่ละรุ่น

ณัฐศิริ เชาว์ประสิทธิ์ และ สมชาย เล็กเจริญ (2560) งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบกระบวนการวิเคราะห์คำหายาภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิง โดยใช้โมเดลในการตรวจจับคำหายาภาษาไทยด้วย พจนานุกรมคำหายาที่ผ่านการปรับปรุงใช้เทคนิค TFICF (Term Frequency - Inverse Class Frequency) จากการวิจัยครั้งนี้พบว่า การศึกษาเปรียบเทียบกระบวนการวิเคราะห์คำหายาด้วยเทคนิคดาต้าไมน์นิง ได้แก่ เทคนิคต้นไม้ตัดสินใจมีค่าความถูกต้อง (Accuracy) เท่ากับ 0.96และค่าความคลาดเคลื่อนเฉลี่ย กำลังสอง (RMSE) เท่ากับ 0.19 รองลงมา เทคนิคนาอิวเบย์มีค่าความถูกต้อง เท่ากับ 0.96 และค่าความคลาดเคลื่อนเฉลี่ย กำลังสอง (RMSE) เท่ากับ 0.21 และ เทคนิคเคเนียร์เรสเนเบอร์ให้ค่าความถูกต้องที่น้อยกว่า โดยมีค่าความถูกต้อง (Accuracy) เท่ากับ 0.95 และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เท่ากับ 0.22 แม้ว่าเทคนิคต้นไม้ตัดสินใจและเทคนิคนาอิวเบย์จะให้ค่าความถูกต้องที่เท่ากัน แต่พบว่าการวิเคราะห์คำหายาด้วยเทคนิคต้นไม้ตัดสินใจมีค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) น้อยที่สุด และมีรูปแบบการวิเคราะห์ข้อความที่ง่ายต่อความเข้าใจมากกว่าเทคนิคอื่น ๆ

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้น ทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Regression เพื่อจัด กลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล คือ โปรแกรม SQL Server 2014 Management Studio & SQL Server Data Tools for Visual Studio 2013 เพื่อทำการ เปรียบเทียบ และเลือกผลลัพธ์ที่ดีที่สุดมานำเสนอ จากนั้นนำข้อมูลสารสนเทศมาทำการ แสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน web browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา