

สรุปบทเรียนการพัฒนาความรู้

หลักสูตร สำนักงานคณะกรรมการข้าราชการพลเรือน (สำนักงาน ก.พ.)

ชื่อหลักสูตร Data Visualization

ชื่อ-สกุล (ภาษาไทย) นางสาวณัฐพร ประคองเก็บ ตำแหน่ง นักวิทยาศาสตร์ชำนาญการพิเศษ

สังกัด กลุ่มวิจัยแร่และจุลสัณฐานดิน สำนักวิทยาศาสตร์เพื่อการพัฒนาที่ดิน

วันที่อบรม 19 ธันวาคม 2566

วัตถุประสงค์ของหลักสูตร

1. เข้าใจสาเหตุของการทำ Data Visualization ได้
2. สามารถนำเสนอ Visualization เพื่อการสื่อสารได้
3. ประยุกต์ใช้การเตรียมข้อมูล การทำความสะอาด และการจัดทำ Data Visualization ได้

สรุปบทเรียน

Data Visualization คือ การนำข้อมูลที่ได้จากแหล่งข้อมูลต่างๆ มาวิเคราะห์ประมวลผล แล้วนำเสนอออกมาในรูปแบบที่มองเห็นและทำความเข้าใจได้ง่าย จุดประสงค์สำคัญของการทำ Data Visualization คือ การนำเสนอข้อมูลให้ผู้รับสารสามารถเข้าใจได้ทันที ประโยชน์ของการทำ Data Visualization ก็สรุปได้ดังนี้

1. ช่วยให้เข้าใจข้อมูลได้ง่ายขึ้น เพราะเป็นข้อมูลที่ย่อยและจัดรูปแบบให้เข้าใจได้ทันทีด้วยภาพ
2. ช่วยให้เห็น Insight ได้ชัดเจน เห็นข้อเปรียบเทียบ เห็นแนวโน้มของข้อมูล ช่วยให้อนุมานความสัมพันธ์ของข้อมูลได้ง่ายขึ้น
3. ช่วยประหยัดเวลาในการตีความข้อมูลและตัดสินใจ ลดภาระการค้นหาและเปรียบเทียบข้อมูล
4. ช่วยให้สามารถมองเห็นจุดที่น่าสนใจของชุดข้อมูลได้ แม้ยังไม่ได้ตั้งข้อสันนิษฐานเกี่ยวกับข้อมูล ก็สามารถมองเห็นข้อมูลที่มีความโดดเด่นบางอย่างขึ้นมาได้จากการทำความเข้าใจภาพ
5. ช่วยให้ข้อมูลมีความน่าสนใจมากขึ้น จากการนำเสนอที่มีเรื่องราวหรือใช้สีที่ดึงดูดสวยงาม

องค์ประกอบหลักของ Data Visualization ได้แก่

1. ข้อมูล (Information): ข้อมูลถือเป็นองค์ประกอบตั้งต้นของการทำ Data Visualization
2. เรื่องราว (Story): เรื่องราวในที่นี้หมายถึงเรียงร้อยข้อมูลให้ออกมาเป็นลำดับขั้นตอนในการนำเสนอ อย่างการนำเสนอข้อมูลที่เป็นภาพรวมก่อนจะย่อยลงไปข้อมูลย่อยในส่วนต่างๆ และมุมต่างๆ เพื่อให้ผู้อ่านข้อมูลเข้าใจได้ง่าย
3. เป้าหมาย (Goal): จุดประสงค์คือการตั้งคำถามว่า การทำ Data Visualization ครั้งนี้ทำเพื่ออะไร ต้องการหาคำตอบเกี่ยวกับอะไร หรือต้องการนำเสนอข้อมูลในมุมไหน ซึ่งองค์ประกอบนี้ของ Data Visualization ต้องชัดเจน เป็นเหมือนแกนกลางยึดอีกสามองค์ประกอบเอาไว้

4. รูปแบบการนำเสนอ (Visual Form): รูปแบบการนำเสนอ เป็นอีกหนึ่งส่วนประกอบของการทำ Data Visualization เช่น การเลือกนำเสนอในรูปแบบ Infographic ขึ้นเดียว หรือนำเสนอผ่านแดชบอร์ดหลายหน้า และข้อมูลแต่ละส่วนจะนำเสนอรูปแบบใด เช่น กราฟแท่ง กราฟวงกลม ตารางตัวเลขจริง เป็นต้น

ประเภทของ Data Visualization ได้แก่

1. ตาราง (Table): เป็นรูปแบบการนำเสนอข้อมูลที่ประกอบด้วย Row หรือแถวแนวนอน และ Column หรือแถวแนวตั้ง เป็นการนำเสนอข้อมูลในการเปรียบเทียบตัวแปรหลากหลายตัว ใช้นำเสนอข้อมูลจำนวนมากแบบมีโครงสร้างได้

2. กราฟวงกลม (Pie chart) หรือกราฟแท่งแบบต่อกัน (Stacked bar chart): รูปแบบข้อมูลนี้เป็นการแบ่งข้อมูลออกเป็นส่วนต่างๆ โดยเทียบเป็นสัดส่วนของข้อมูลทั้งหมด เป็นประเภทของ Data Visualization ที่ใช้งานง่ายเพื่อแสดงองค์ประกอบต่างๆ ในเซตข้อมูล

3. กราฟเส้น (Line chart) และแผนภูมิพื้นที่แบบชั้น (Area chart): เป็นการนำเสนอสัดส่วนข้อมูลในแต่ละชุดในแต่ละช่วงเวลา ใช้สำหรับดูความเปลี่ยนแปลงของข้อมูลเปรียบเทียบกัน

4. ฮิสโตแกรม (Histogram): เป็นกราฟที่ใช้แผนภูมิแท่งเพื่อนำเสนอข้อมูลแต่ไม่มีพื้นที่ว่างระหว่างแท่ง ใช้นำเสนอความถี่ตามปริมาณของข้อมูล เพื่อดูการกระจายตัวแบบมีความต่อเนื่องกัน

5. แผนภาพการกระจาย (Scatter Plot): เป็นแผนภาพที่ใช้นำเสนอความสัมพันธ์ของข้อมูล 2 ตัวแปร โดยเป็นจุดที่กระจายอยู่ในแผนภาพ จุดต่างๆ อาจมีความสัมพันธ์หรือไม่มีความสัมพันธ์กันก็ได้

6. ฮีตแมพ (Heat maps): เป็นรูปแบบของการทำ Data Visualization ที่นำเสนอข้อมูลตามเขตพื้นที่ หรือเขตบริเวณ มักเห็นในรูปแบบแผนที่เมืองหรือประเทศ และแผนภาพฮีตแมพของหน้าเว็บไซต์

7. แผนภาพต้นไม้ (Tree maps): เป็นรูปแบบการแสดงผลเป็นรูปทรงสี่เหลี่ยมซ้อนกัน มักใช้ในการนำเสนอสัดส่วนของข้อมูลกลุ่มต่างๆ โดยแบ่งออกตามพื้นที่ของรูปทรง

การทำความสะอาดข้อมูล หรือ “Data Cleansing” หมายถึง กระบวนการตรวจสอบ สะสาง แก้ไข หรือจัดรูปแบบข้อมูลให้อยู่ในสภาพที่พร้อมใช้งานที่สุด รวมไปถึงคัดกรองข้อมูลที่ไม่ถูกต้องหรือไม่จำเป็นออกไปจากชุดข้อมูลที่จะใช้วิเคราะห์หรือประมวลผล เพื่อให้ชุดข้อมูลที่จะใช้มีความสมบูรณ์ มีคุณภาพ พร้อมนำไปวิเคราะห์และใช้ประโยชน์ ซึ่งอาจเรียกอีกอย่างว่า เป็นการทำให้ “re-organize” ข้อมูลใหม่ก็ได้ สาเหตุของการที่เราต้องทำความสะอาดข้อมูลก่อนใช้ นั่นก็เพื่อหลีกเลี่ยงผลลัพธ์ที่ไม่ถูกต้องแม่นยำจากปัจจัยหลายอย่าง ไม่ว่าจะเป็นความผิดพลาดของการบันทึกข้อมูล การพิมพ์ผิด รูปแบบข้อมูลที่แตกต่างกัน ชุดข้อมูลไม่สอดคล้องกับคำถาม ข้อมูลที่ไม่เป็นความจริง ข้อมูลที่ไม่สามารถอ้างอิงในระบบได้ เป็นต้น

ลักษณะข้อมูลที่ต้องทำ Data Cleansing ได้แก่ 1. ชุดข้อมูลที่ไม่ได้อยู่ในรูปแบบหรือไฟล์ประเภทเดียวกัน 2. ข้อมูลไม่ได้ถูกจัดเก็บในรูปแบบที่ต้องการ และ 3. ข้อมูลที่ไม่ถูกต้องหรือมีข้อผิดพลาดในการเก็บข้อมูล

กระบวนการทำ Data Cleansing ได้แก่

1. กำจัดข้อมูลที่ซ้ำซ้อนและข้อมูลที่ไม่เกี่ยวข้องออก คือ การกำจัดข้อมูลที่ซ้ำซ้อนและข้อมูลที่ไม่เกี่ยวข้องเพื่อลดภาระในการประมวลผล

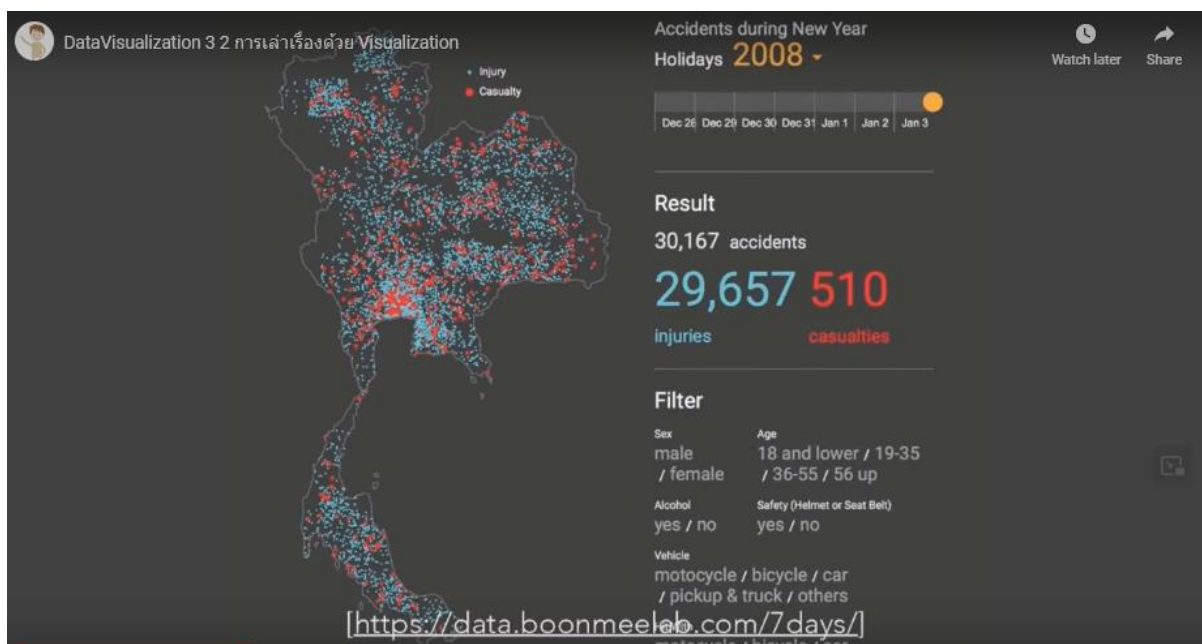
2. แก้ไขข้อผิดพลาดในเชิงโครงสร้างหรือรูปแบบ คือ การจัดการกับข้อมูลที่ผิดพลาดในเชิงโครงสร้างหรือรูปแบบ (structural errors)

3. กรองข้อมูลที่มีค่าผิดปกติออกจากชุดข้อมูล ตัวอย่างเช่น ค่าที่ไม่มีทางเป็นไปได้ ยกตัวอย่างเช่น จำนวนสมาชิกในครอบครัว 500 คน เป็นต้น จึงควรมีการกำจัดช่วงของค่าหรือตัวเลขที่ไม่มีทางเป็นจริงออกในคราวเดียวด้วยการกำหนด “Outlier” ขีดเส้นกันเอาเฉพาะช่วงข้อมูลที่ต้องการ

4. จัดการกับข้อมูลที่หายไปหรือไม่สมบูรณ์ ได้แก่ 1. ตัดข้อมูลที่หายไปออกจากการประมวลผล 2. ใส่หรือแทนค่าข้อมูลที่หายไป และ 3 เปลี่ยนวิธีประมวลผลหรือใช้ชุดข้อมูลอื่นๆ ที่สามารถให้ Insight ได้ใกล้เคียง

5. ตรวจสอบความถูกต้อง (QA) เป็น ขั้นตอนสุดท้ายของการทำ Data Cleansing คือ ขั้นตอนการตรวจสอบชุดข้อมูลว่าได้ชุดข้อมูลที่สมบูรณ์พร้อมสำหรับการวิเคราะห์เพื่อให้ได้คำตอบที่ต้องการแล้วหรือไม่

ภาพตัวอย่างการเล่าเรื่องด้วย Visualization ที่ดี ภาพแสดงจำนวนการเกิดอุบัติเหตุบนท้องถนนในช่วงวันปีใหม่ของปี 2008 จากภาพแสดงข้อมูลให้เข้าใจได้ง่าย และผู้รับสารได้รับข้อมูลครบถ้วน



ประโยชน์ที่ได้รับต่อตนเอง

1. ทำให้เข้าใจสาเหตุของการทำ Data Visualization ได้มากกว่าก่อนเรียน
2. สามารถนำเสนอ Visualization เพื่อการสื่อสารได้มากกว่าก่อนเรียน
3. ประยุกต์ใช้การเตรียมข้อมูล การทำความสะอาด และการจัดทำ Data Visualization ได้มากกว่าก่อนเรียน

ประโยชน์ที่ได้รับต่อหน่วยงาน

1. สามารถนำความรู้ได้จากการเรียน มาจัดการกับข้อมูลของหน่วยงานเพื่อนำเสนอผลงานหรือสื่อสารได้ดีขึ้น