

ตอนที่ ๓ การสร้าง Visualization

หลังจากที่ทราบถึงสาเหตุของการสร้าง Visualization และทราบว่า Visualization ที่ดีเป็นอย่างไรแล้วในลำดับต่อไปคือการทราบถึงวิธีการสร้าง Visualization

วิธีการสร้าง Visualization

- การวาดด้วยมือ (Draw by hand) เป็นวิธีที่ถูกต้องที่สุด
- ถ่ายภาพ (Take a photo) ตามตัวอย่างที่นำเสนอไปก่อนหน้า
- ใช้โปรแกรมคอมพิวเตอร์ (Use a computer software) เหมือนกับการวาดและการถ่ายภาพ แต่ทำด้วยเครื่องคอมพิวเตอร์
- การเขียน Code (Code it yourself) เพื่อแปลงข้อมูลออกมาเป็น Visualization

ตัวอย่างการสร้าง Visualization ด้วยมือ

พบว่าภาพที่แสดงผลไม่เหมือนกัน แม้ว่าอิมของข้อมูลเหมือนกัน ซึ่งจะเห็นว่าแต่ละคนสามารถนำเสนอแตกต่างกันได้ และ visualization ไม่จำเป็นต้องทำกับข้อมูลขนาดใหญ่หรือโกลด์ เราสามารถนำมาใช้กับข้อมูลในชีวิตประจำวันของเราได้

ตัวอย่างของการสร้าง Visualization ด้วยภาพถ่าย

สร้างโดยการจัดเรียงสิ่งที่เราสนใจและถ่ายรูปออกมา จากตัวอย่าง คือ เทปที่จัดแบ่งตามแนวเพลงที่นิยมฟังในอดีต ซึ่งจากภาพถ่ายจะเห็นว่าแนวเพลงไหนเป็นแนวเพลงที่เรานิยมมากที่สุด

ตัวอย่างของการสร้าง Visualization ด้วยภาพถ่าย อีกตัวอย่าง คือ เป็นการใช้เทคโนโลยีเข้ามาประกอบ เช่น แฟลช และ ชัตเตอร์สปีด ทำให้เห็นความเร็ว จำนวน และ เห็นสิ่งที่เราไม่สามารถเห็นได้ด้วยตาเปล่า ความน่าสนใจของภาพทางขวามือ คือ มีการบ่งปริมาณความเร็วของไม้กอล์ฟด้วย

ตัวอย่างของการวิเคราะห์จากภาพถ่าย

ซึ่งรูปภาพจะแสดงให้เห็นได้ชัดเจนกว่า VDO เพราะสิ่งที่เกิดขึ้นเร็วกว่าที่จะมองแบบภาพเคลื่อนไหว และจากรูปคือใช้ภาพในการวิเคราะห์สรีระในการแข่งขันของนักกีฬาซึ่งเคยได้รับบาดเจ็บ Visualization สามารถนำมาใช้วิเคราะห์ถึงการปรับรูปแบบของการจัดวางสรีระในการแข่งขันได้

การใช้ software และการเขียนโค้ด อาจจะมีห่างไกลจากผู้ที่เพิ่งจะเริ่มต้นวิชา Visualization ทั้งนี้ผู้ที่สนใจสามารถใช้ R, Tableau, d3.js เป็นต้น

Visualization คือการสื่อสารและแสดงข้อมูลที่มีขั้นตอนในการจัดทำอย่างชัดเจน

- การหาและรวบรวมข้อมูล (Data acquisition) ซึ่งเป็นขั้นตอนที่ยากที่สุดเพราะหากไม่มีข้อมูลก็ไม่สามารถทำขั้นตอนอื่นๆ ได้
- การเตรียมข้อมูล เช่น การทำความสะอาดข้อมูล เปลี่ยนรูปแบบของข้อมูล (Data preparation; cleaning, restructuring) เพื่อให้พร้อมต่อการใช้งาน
- การทำ Visualization ซึ่งจะต้องเลือกว่าจะสร้างแบบ Visual mapping หรือ Visual rendering

ข้อมูลมีแหล่งที่มา เช่น

- เก็บข้อมูลด้วยตนเอง โดยการบันทึกทั้งในกระดาษ หรือบันทึกในแอปพลิเคชัน โดยหากเก็บทุกวันในช่วงระยะเวลาหนึ่งเราก็สามารถนำมาวิเคราะห์ข้อมูลได้
- การไหลตข้อมูล โดยมีการเปิดเผยให้ดาวน์โหลดทั้งไทยและต่างประเทศ
- การซื้อข้อมูล บริษัทเอกชนมีการขายข้อมูลมากขึ้น เช่น ข้อมูลเศรษฐกิจหรืออุตสาหกรรม เพื่อนำมาใช้ในการประกอบการตัดสินใจ
- การแฮ็คข้อมูล จากข้อมูลที่มีอยู่แล้วแต่ไม่พร้อมในการดาวน์โหลด โดยเราต้องเขียนสคริปเพื่อดึงข้อมูลปริมาณมากๆ ในเวลาเดียวกันเพื่อความรวดเร็วในการดึงข้อมูลมาใช้ในการทำ Visualization
- การเรียกร้องข้อมูลการขอข้อมูลที่สนใจจากองค์กรที่เป็นเจ้าของข้อมูล

การเตรียมข้อมูล

ข้อมูลที่ดี คือข้อมูลที่สามารถนำไปใช้ได้ง่ายโดยข้อมูลที่เหมาะสมในการใช้งาน คือ ข้อมูลที่เป็น file excel, csv, หรือ json ข้อมูลส่วนใหญ่มักอยู่ในรูป PDF นำไปใช้ต่อยากกว่าเพราะจะต้องเสียเวลาในการทำความสะอาดข้อมูล ตรวจสอบข้อมูลและปรับรูปแบบของข้อมูลก่อนนำไปใช้จริง

ข้อมูลที่ดี มักอยู่ในรูปแบบคอลัมน์ (columns)และเป็นแถว (Row) ที่ชัดเจน ซึ่งลักษณะจะมีลักษณะการจัดเรียงข้อมูลเป็นแบบเดียวกับ excel โดยแต่ละ sample จะแสดงผลในแต่ละแถว และลักษณะของ sample จะแสดงผลในแต่ละคอลัมน์ ซึ่งข้อมูลลักษณะนี้จะสามารถทำความเข้าใจได้ง่ายและนำมาใช้ต่อได้ง่ายกว่าแบบอื่น

การแสดงผลข้อมูลในรูปแบบ Columns สามารถจำแนกลักษณะของ sample ได้ง่าย และสามารถเขียนอธิบายข้อมูลได้

ในบางครั้งข้อมูลที่น่ามาการจัดเก็บไม่เหมือนกัน ทั้งในรูปแบบของ Format หรือ คำที่ใช้ เช่น จำนวนล้อ บาง sample ระบุเป็นตัวเลข บาง sample ระบุเป็นคำ ซึ่งจะต้องมีการปรับให้สอดคล้องกันก่อนนำไปวิเคราะห์โดยทุกคอลัมน์มีรูปแบบข้อมูลที่เหมือนกัน

<p>ในบางคอลัมน์ที่มีการระบุถึงลักษณะของ sample ไข่มากกว่าหนึ่งอย่าง ควรจับแยกออกมาเป็นคอลัมน์และนำเสนอข้อมูลแบบเดียว เช่น จากเดิมที่คอลัมน์สุดท้ายมีการระบุมีแอร์หรือไม่มีแอร์ มีมิเตอร์หรือไม่มีมิเตอร์ ก็ให้แยกเป็นคอลัมน์ใหม่ เพื่อให้เห็นข้อมูลที่ละเอียดขึ้น และอาจจะทำให้เราสามารถตัดข้อมูลบางข้อมูลที่ไม้อำคัญออกไปได้</p>
<p>หลังจากได้ข้อมูลที่มีโครงสร้างที่เหมาะสมแล้ว ลำดับต่อไปคือการทำความสะอาดข้อมูล โดยเน้นถึงการทำให้สิ่งเดียวกันให้เหมือนกัน ซึ่งเป็นขั้นตอนของการทำความสะอาดข้อมูล ในส่วนนี้จะใช้เวลาประมาณครึ่งหนึ่งของการทำ Visualization</p>
<p>การปรับโครงสร้างข้อมูล ควรปรับลักษณะการจัดข้อมูลให้สอดคล้องกับการนำไปใช้งาน โดยข้อมูลมีลักษณะเป็นอิสระต่อกันทำให้เราสามารถปรับรูปแบบได้เสมอ แยกตามดอกและตามเลข</p>
<p>หรือการจัดเรียงตามตัวเลข เอาตัวเลขเดียวกันอยู่ด้วยกันก็ได้ ขึ้นอยู่กับความเหมาะสมของการนำไปใช้</p>
<p>กิจกรรม ฝึกการจัดเรียงข้อมูลตามความเข้าใจ</p> <ul style="list-style-type: none"> - จัดเรียงตัวเลขสองตัว เลขคือ ๗๕ และ ๓๗ โดยให้คิดว่าจะแสดงเป็นภาพได้อย่างไรได้บ้าง - สามารถจัดเรียงได้กี่แบบ - วิธีไหนดีไม่ใช่ว่า และแต่ละวิธีเหมาะสมกับสถานการณ์ใดไม่เหมาะสมกับสถานการณ์ใด
<p>ตอนที่ ๔ มาตรฐาน Visualization และรูปแบบการใช้งาน</p> <p>ลักษณะทั่วไปของ Visualization ก่อนอื่นจะพูดถึงกิจกรรมที่แล้วว่าจากตัวเลขสองตัว คือ ๗๕ และ ๓๗ เราสามารถแสดงผลอย่างไรได้บ้าง</p>
<p>สามารถทำได้ในหลายลักษณะ เช่น</p> <ul style="list-style-type: none"> - ให้หนึ่งจุดแทนตัวเลขหนึ่งตัว ๗๕ มีจำนวนจุด ๗๕ จุดและ ๓๗ ก็มีจุด ๓๗ จุด - เป็น pie chart - เป็นความยาว โดยใช้สัดส่วนที่เหมาะสม - ใช้ความเข้มของสี เลขค่ามากสีเข้มกว่า ค่าน้อยเข้มน้อยกว่า เป็นต้น
<p>แสดงเป็นจำนวน ความยาว มุม พื้นที่ สี ซึ่งแต่ละวิธีต้องพิจารณาต่อการใช้ดีหรือไม่ใช่ว่าอะไรขึ้นอยู่กับการนำไปใช้งานหรือไม่</p>
<p>การแสดงผลด้วยภาพเป็น ๒ กลุ่มตามชนิดของข้อมูล</p> <ol style="list-style-type: none"> ๑) Ordinal data = ข้อมูลที่มีลำดับชัดเจน เช่น ระยะห่างจากจุด ความยาว มุม ขนาด และความเข้มของสี ซึ่งสามารถเรียงได้ ๒) Categorical data = ข้อมูลที่แบ่งเป็นหมวดหมู่ชัดเจน ที่ไม่มีการแบ่งลำดับอย่างชัดเจน เช่น ข้อมูลชื่อจังหวัดที่มีลำดับไม่ชัดเจน เหมาะกับการแสดงผลด้วยชนิดของสีและรูปร่างของ symbol ที่ไม่มีการจัดลำดับในตัวเอง โดยไม่จำเป็นต้องใช้ข้อมูลแบบเดียว สามารถนำมาประกอบกันได้เพื่อให้เห็นข้อมูลในหลายมิติ
<p>การทำชาร์ตแตกต่างกันออกไปตามลักษณะของข้อมูล ที่มีการนำมาใช้ประกอบกัน</p>
<p>กรณีที่เป็น ข้อมูลที่เป็น Categorical data ทั้งคู่ ส่วนใหญ่จะแสดงผลในลักษณะของตาราง หรือคอลัมน์ ความน่าสนใจอยู่ที่ข้อมูลลักษณะนี้สามารถสลับข้อมูลได้เนื่องจากเป็นข้อมูล Categorical data ทั้งคู่ การสลับเพื่อให้เห็น pattern ในการวิเคราะห์ข้อมูล</p>
<p>เหมาะกับข้อมูลที่เป็น Categorical data และ Ordinal data ความสูงของแท่งและความยาวของแท่งเป็น Ordinal แต่ในแต่ละแท่งสามารถสลับกันได้เพราะไม่ได้มีการเรียงลำดับไว้</p>
<p>ข้อมูล Ordinal data เช่น ความสูง และน้ำหนัก โดยกำหนดแกน X และ แกน Y โดยแกน X อาจจะเป็นน้ำหนักและแกน Y คือส่วนสูง</p>
<p>คล้ายกับ Scatterplot มักจะใช้กับข้อมูลที่ยากดู pattern หรือ รูปแบบบางอย่าง โดย LINE CHART จะเหมาะกับข้อมูลบางประเภทเท่านั้น</p>
<p>ตัวอย่าง LINE CHART แสดงยอดขายแต่ละวัน โดยข้อมูลยอดขายเป็นข้อมูลที่รวมหรือสะสมมา แต่จะไม่ทราบว่ายอดขายระหว่างวัน</p>
<p>หากมีข้อมูลมากกว่า ๒ มิติหรือ ๒ หลัก โดยการใช้ excel ได้</p> <ul style="list-style-type: none"> - Barchart (localtask) ด้านบน เหมาะสมที่จะนำมาใช้ในการเปรียบเทียบข้อมูลในระยะสั้นๆ - Bar chart (Global Task) ด้านล่าง เหมาะกับการนำไปใช้ในการนำเสนอข้อมูลที่ต้องการให้เห็นภาพรวมทั้งหมด การนำข้อมูลสองชนิดรวมกันมาเปรียบเทียบกันในระยะยาว ทั้งนี้งานแต่ละประเภทเหมาะกับ chart คนละแบบขึ้นอยู่กับลักษณะการนำไปใช้งาน
<p>LINE chart บนและล่างแตกต่างกันงานนำเสนอที่เหมาะสม หากต้องการเปรียบเทียบข้อมูลหลายๆ ตัวให้นำมาซ้อนทับกัน ด้านบนเหมาะกับ local task ที่สามารถเปรียบเทียบข้อมูลทั้งสองข้อมูลได้ แต่หากต้องการดูเฉพาะข้อมูลหรือ Global Task กราฟล่างจะเหมาะสมกว่าเพราะเป็นข้อมูลในภาพรวม</p>
<p>แผนที่มีหลายลักษณะ ใช้ในการแสดงพื้นที่ ข้อมูลต่างๆ จัดเป็น Visualization ชนิดหนึ่ง ในบางครั้งใช้ความเข้มของสีเพื่อแสดงลักษณะของพื้นที่</p>
<p>แผนที่ที่ใช้ความเข้มของสีในการแสดงข้อมูล ทำให้เห็นข้อมูลที่จัดแบ่งประเภทไว้ได้ชัดขึ้น</p>

นอกจากนี้มีการวิเคราะห์รูปอื่นๆ ในด้านการตลาด เช่น บางบริษัทมีการวิเคราะห์ข้อมูลการใช้ emoji ใน Instagram ว่ามีการใช้งานอย่างไร การจับคู่การใช้ emoji เพื่อดูว่าปกติแล้ว emoji รูปไหนใช้กับสินค้าอะไรจะสามารถนำข้อมูลนี้ไปใช้ทำการตลาดของบริษัทได้

การวิเคราะห์สามารถแยกข้อมูลเพื่อวิเคราะห์ความสัมพันธ์ได้ เช่น การแยก emoji ที่ปกติใช้คู่กันออกจากกันแล้วลากเส้นความสัมพันธ์ การวิเคราะห์ในลักษณะนี้จะช่วยให้เราทราบถึงความเชื่อมโยงที่เกิดขึ้นได้ ซึ่งจะเห็นว่าการนำเสนอข้อมูลมีหลายแบบขึ้นอยู่กับลักษณะของข้อมูล