CSV คืออะไร

CSV ย่อมาจาก Comma Separated Value เป็นไฟล์ข้อความประเภทหนึ่งที่ใช้สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาค หรือคอมม่า (,) ในการแบ่งแต่ละคอลัมน์ โดยปกติเราสามารถบันทึกไฟล์จาก Microsoft Excel ออกมาเป็น CSV ไฟล์ได้โดยตรง หรือ อาจได้ไฟล์ CSV จากการ export ไฟล์จากระบบฐานข้อมูลอื่นๆ

Pandas คืออะไร

Pandas มาจากคำว่า PANel DAta เป็น Library ใน Python ที่ทำให้เราเล่นกับข้อมูลได้ง่ายขึ้น เหมาะมากสำหรับทำ Data Cleaning / Wrangling ครับผม วิธีการใช้งาน Pandas คือ โหลดไฟล์ข้อมูล เช่น CSV เข้าไป แล้วเราจะได้ข้อมูลในรูปแบบตาราง (DataFrame) ที่แบ่งข้อมูลตามแถวและคอลัมน์ หรือเหมือน Excel ที่เราใช้กันนั่นเอง

โดยขั้นตอนการทำงานประกอบไปด้วย

1. อ่านข้อมูลจากไฟล์ CSV
2. ทำการจัดเรียงข้อมูล
3. ทำการ transform ข้อมูลในแต่ละ column
4. ทำการจัดการเรื่อง missing value
5. การ plotting หรือแสดงข้อมูลในรูปแบบ graph

Matplotlib คืออะไร

matplotlib เป็นมอดูลสำหรับวาดกราฟที่ใช้งานได้หลากหลายยืดหยุ่น กราฟที่วาดได้ใน matplotlib นั้นมีหลายชนิดด้วยกัน เช่นกราฟเส้นธรรมดา, แผนภูมิแท่ง, แผนภูมิวงกลม, การกระจาย, ฯลฯ ฟังก์ชันที่ใช้วาดกราฟเส้นธรรมดาก็คือฟังก์ชันที่ชื่อว่า plot อาร์กิวเมนต์ที่ต้องใส่ในฟังก์ชันนี้คือพิกัดในแนวแกน x และ y ตามลำดับ โดยข้อมูลที่ใส่ต้องเป็นออบเจ็กต์ชนิดลำดับ เช่น ลิสต์, ทูเพิล, เรนต์ และสามารถใช้อาเรย์ของ numpy ได้ด้วยการวาดกราฟด้วย matplotlib นั้นเราสามารถปรับแต่งอะไรได้มากมายหลากหลายตามต้องการ เช่นสีเส้น, รูปแบบเส้น, ขอบเขตกราฟ, ชื่อแกน, คำอธิบายเพิ่มเติม, ฯลฯ

สรุป

จากกราฟในแต่ละเดือนสรุปได้ว่าแบรนด์Toyotaนั้นเป็นแบรนด์ที่มียอดขายมากที่สุดและสาเหตุที่Toyotaมียอดขายมากที่สุดก็คือการที่Toyotaนั้นมีรถหลากหลายราคาทำให้คนทุกฐานะสามารถเข้าถึงได้ มีหลากหลายรุ่นเหมาะสมแก่ทุกการใช้งาน และอีกหนึ่งปัจจัยสำคัญที่ทำให้ได้ยอดขายมากที่สุดคือการที่Toyotaนั้นมีศูนย์บริการที่เยอะกระจายทั่วทุกพื้นที่เมื่อเทียบกับแบรนด์อื่นนั้น แบรนด์อื่นอาจจะมีรถที่ไม่หลากหลายรุ่นทำให้ต้องเลือกเจาะจงว่าจะนำไปใช้เพื่อประโยชน์อะไรหรือบางแบรนด์อาจจะมีราคาที่สูงเกินไป หรือมีศูนย์บริการที่ครอบคลุมไม่ทั่วถึงทุกพื้นที่หากรถมีปัญหาจะให้เกิดความลำบากได้