**01076253 PROBABILITY AND STATISTICS HW 1**

**ชื่อชุดข้อมูล** Travel Review Rating Dataset ( From the Machine Learning Repository of University of California )

**ชื่อคอลัมน์**

Numeric

* ประเภทการรีวิวแยก 5 ประเภท

Category

* หมายเลข User ผู้รีวิวสถานที่เหล่านั้น

**ทำไมถึงสนใจ**

ผมเป็นคนที่รัก และชอบในการเดินทางท่องเที่ยวไปยังสถานที่ต่าง ๆ แต่บางครั้งผมก็มีปัญหาอยู่ในใจมากมาย เช่น ผมควรจะไปเยี่ยมชมสถานที่ไหนดี ? มีสถานที่ที่น่าสนใจที่ตรงกับไลฟ์สไตล์ของผม หรือไม่ ในสถานที่ที่ ๆ ใกล้บ้าน ? บ่อยครั้งที่ผมใช้เวลาหลายชั่วโมงเพื่อค้นหาสถานที่ที่น่าสนใจที่จะออกไปข้างนอก ผมว่ามันเสียเวลาเปล่า โดยใช่เหตุ

จะเกิดอะไรขึ้นละครับ ถ้าเราสามารถสร้างระบบแนะนํา หรือการรีวิว ซึ่งสามารถแนะนําสถานที่ที่น่าสนใจหลายแห่งตามความต้องการของแต่ละคนได้ ด้วยข้อมูลจากการตรวจสอบของ Google ซึ่งผมจะพยายามแบ่งผู้ใช้รีวิว Google ออกเป็นกลุ่มที่สนใจคล้ายกัน ดังตัวอย่างข้อมูลที่นำมานี้ เพื่อที่จะได้เป็นทางเลือกในการตัดสินใจต่อใครหลาย ๆ คนที่ประสบพบเจอปัญหาแบบเดียวกับผม

**แหล่งที่มาของชุดข้อมูล** [Travel Review Rating Dataset | Kaggle](https://www.kaggle.com/wirachleelakiatiwong/travel-review-rating-dataset)

**คำอธิบายชื่อคอลัมน์ข้อมูล และวิธีการรวบรวมข้อมูล**

**คำอธิบายชื่อคอลัมน์ข้อมูล**

Numeric (เลือกมาจาก 5 ประเภท ใน 24 ประเภทของข้อมูลจริง)

Attribute 1 : Unique user id

Attribute 2 : Average ratings on beaches

Attribute 3 : Average ratings on parks

Attribute 4 : Average ratings on cafes

Attribute 5 : Average ratings on malls

Attribute 6 : Average ratings on beauty & spas

Category

จำนวนผู้ใช้งานที่เข้ามาให้การแนะนำ หรือรีวิวในประเภทต่าง ๆ

**วิธีการรวบรวมข้อมูล**

ชุดข้อมูลนี้มาจากที่เก็บแมชชีนเลิร์นนิงของมหาวิทยาลัยแคลิฟอร์เนีย, เออร์ไวน์ (UC Irvine) : ข้อมูลการจัดอันดับรีวิวการเดินทาง ชุดข้อมูลนี้จะถูกเติมโดยการจับคะแนนของผู้ใช้จากรีวิวของ Google รีวิวเกี่ยวกับสถานที่ท่องเที่ยวจาก 24 หมวดหมู่ทั่วยุโรปได้รับการพิจารณา คะแนนผู้ใช้ Google มีตั้งแต่ 1 ถึง 5 และมีการคํานวณคะแนนผู้ใช้เฉลี่ยต่อหมวดหมู่