

## Data Analytics Project

จากการศึกษาว่าพฤติกรรมต่างๆของมนุษย์ ขึ้นอยู่กับเพศหรือไม่ เราจึงได้ทำการเก็บข้อมูลจาก Google Forms เพื่อทำการสำรวจข้อสงสัยดังกล่าว โดยจะแบ่ง Categories ดังนี้

1. พฤติกรรมการเลือกรับประทานอาหาร
2. กิจกรรมที่สนใจ
3. Platforms ที่ใช้
4. Content ที่สนใจ
5. วันเกิด

และได้ทำการวิเคราะห์แบบ Decision Tree เนื่องจากเราต้องการหาความสัมพันธ์ระหว่างตัวแปรต่างๆใน categories ว่ามีความสัมพันธ์กับตัวแปรเพศหรือไม่

จากนั้นได้ทำการแบ่งข้อมูลออกเป็นในส่วนของ Test Data และ Train Data เพื่อทดสอบว่า Model ของเรานั้นมีประสิทธิภาพหรือไม่ โดยการแบ่ง ได้แบ่งเป็น 2 ขั้นตอน ดังนี้

**1.จากแบ่งในการหาสถิติเบื้องต้น** เหตุเพราะ ข้อมูลของเพศชายและหญิงมีไม่เท่ากัน เราจึงต้องทำการสุ่ม sample ออกมาเพศละ 20 คน เพื่อที่จะนำมาหาสถิติเบื้องต้น และหา Mean Score ว่าตัวแปรที่มีค่ามากมีค่าน้อย และนำมาทำการ Bar Plot เพื่อให้เห็นภาพง่ายขึ้น

**2.และแบ่งในขั้นตอนการเข้า Model** เป็นการแบ่งข้อมูลปกติเพื่อให้มีข้อมูลในส่วนของ Test และ Train เนื่องจากป้องกันการที่คอมพิวเตอร์อ่านข้อมูลเดิมแล้วเกิดการจดจำข้อมูล โดยเลือกแบ่งที่

Train Data = 70%

Test Data = 30%

โดยในข้อมูลข้อที่ 1 ถึง 4 เราได้ทำการจัดระดับของความชอบเป็น 6 ระดับ เพื่อให้สะดวกต่อการวิเคราะห์ข้อมูล ได้แก่

-3 = ไม่สนใจอย่างมากที่สุด

-2 = ไม่สนใจอย่างมาก

-1 = ไม่สนใจ

0 = เฉยๆ

+1 = สนใจอย่างมากที่สุด

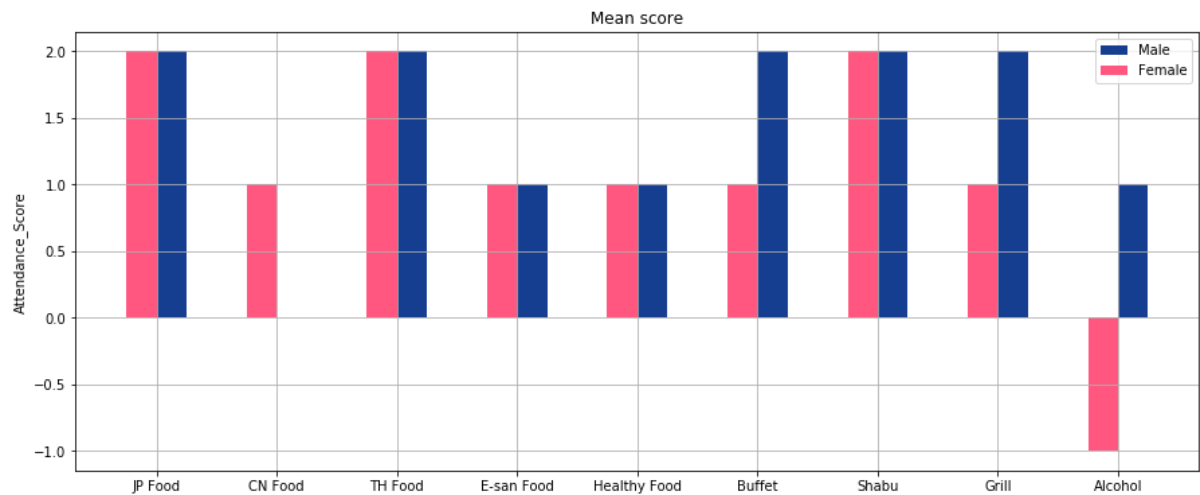
+2 = สนใจอย่างมากที่สุด

+3 = สนใจอย่างมากที่สุด

\*\*\*ส่วนใน categories ที่ 4 เราได้ทำการจำแนกเป็นข้อมูล เป็น วัน เดือน ปี แล้วทำการวิเคราะห์

## 1. พฤติกรรมการเลือกรับประทานอาหาร

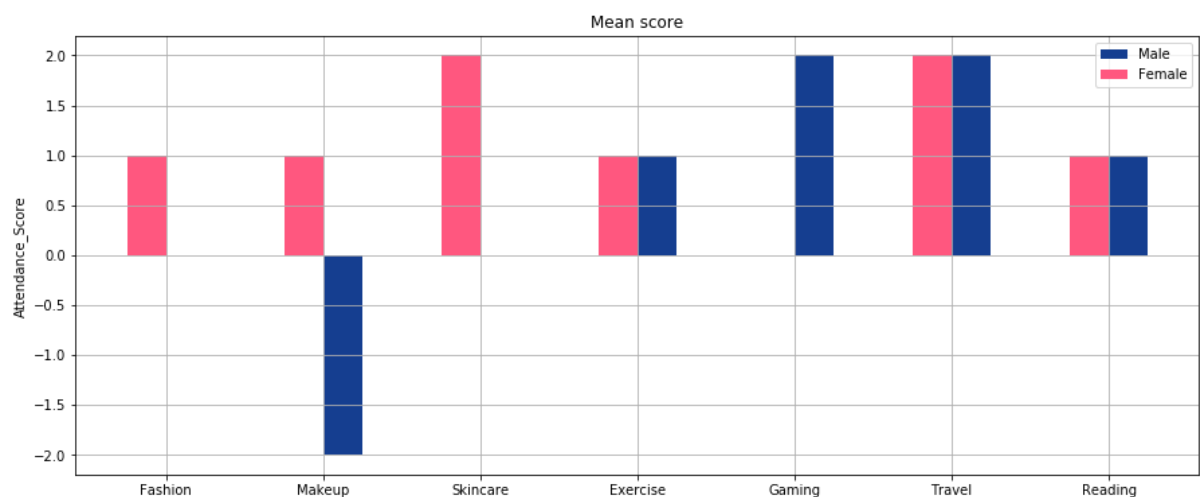
จากข้อมูลที่ได้สำรวจพบว่า สิ่งที่แตกต่างกันเด่นชัดที่สุดระหว่าง เพศชาย และ เพศหญิง คือ เพศชายเลือกกินเนื้อย่าง, รับประทาน Buffet และ ดื่ม Alcohol มากกว่าเพศหญิงดังกราฟ



จากการนำข้อมูลเข้ามาวิเคราะห์ต่อใน Model พบว่า มี R square = 0.7368421052631579

## 2. กิจกรรมที่สนใจ

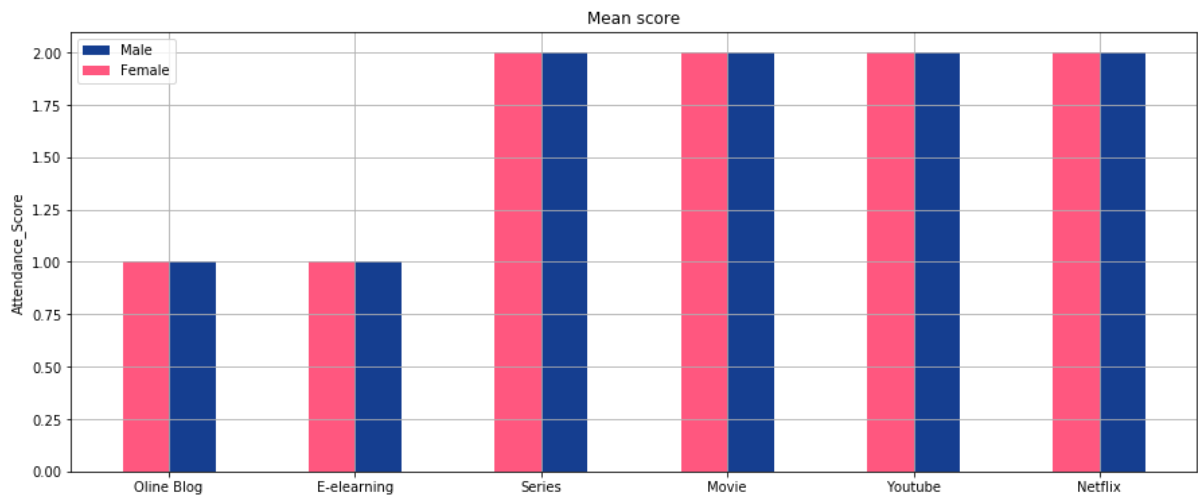
จากข้อมูลที่ได้ทำการสำรวจที่เห็นเด่นชัดว่า เพศชายจะไม่สนใจการแต่งหน้า แต่เพศหญิงไม่มีความสนใจในเกมเลย ดังกราฟ



จากการนำข้อมูลเข้ามาวิเคราะห์ต่อใน Model พบว่า มี R square = 0.8947368421052632

### 3. Platforms ที่ใช้งาน

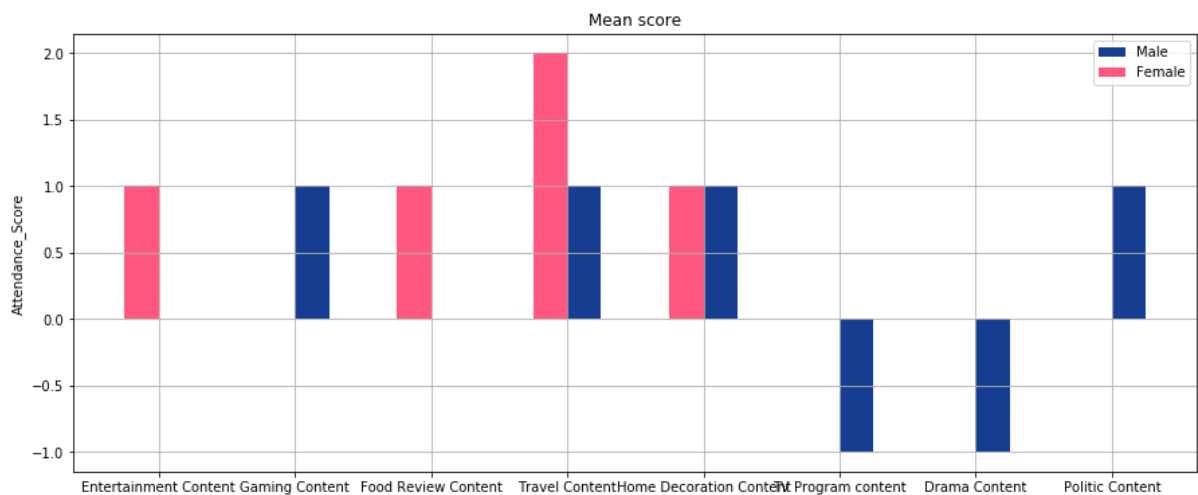
จากข้อมูลที่ได้ทำการสำรวจเป็น ดังกราฟ



จากการนำข้อมูลเข้ามาวิเคราะห์ห้ต่อใน Model พบว่า มี R square = 0.47368421052631576

### 4. Content ที่สนใจ

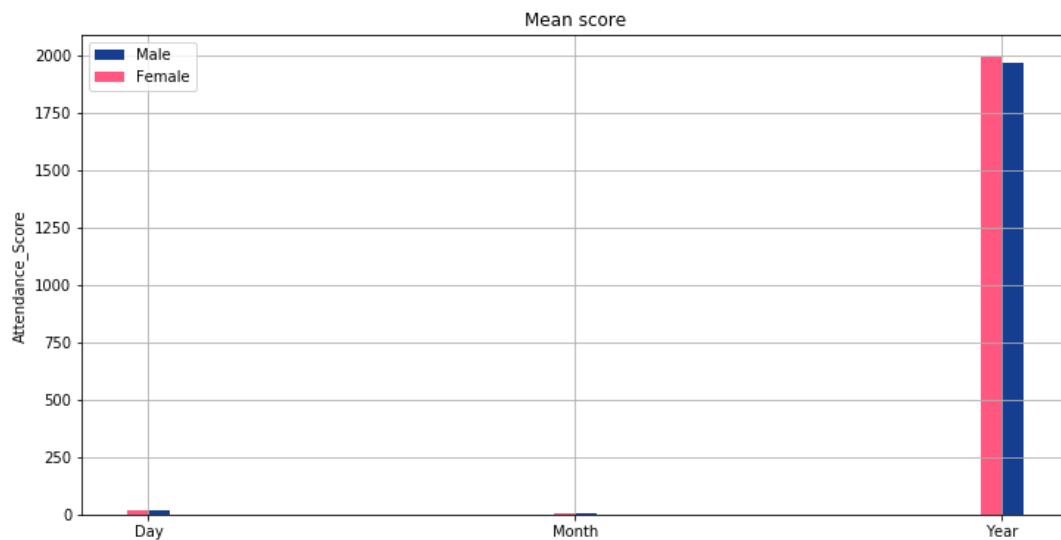
จากข้อมูลที่ได้ทำการสำรวจที่เห็นเด่นชัดว่า เพศชายไม่นิยม Drama Content และ รายการ TV อย่างชัดเจน แต่เพศหญิงนิยม Entertainment และ Travel อย่างชัดเจน ดังกราฟ



จากการนำข้อมูลเข้ามาวิเคราะห์ห้ต่อใน Model พบว่า มี R square = 0.631578947368421

## 5. วันเกิด

จากข้อมูลที่ได้ทำการสำรวจ ทำให้พบว่าวันที่เกิดนั้นไม่มีผลต่อเพศเท่าที่ควร



จากการนำข้อมูลเข้ามาวิเคราะห์ต่อใน Model พบว่า มี R square = 0.42105263157894735

## Summary

Categories	R square
พฤติกรรมการเลือกรับประทานอาหาร	0.737
กิจกรรมที่สนใจ	0.895
Platform ที่ใช้งาน	0.474
Content ที่สนใจ	0.632
วันเกิด	0.421

จากข้อมูลที่ได้ทำการวิเคราะห์โดยเลือกใช้ Model Decision Tree พบว่า กิจกรรมที่สนใจ และพฤติกรรมการเลือกรับประทานอาหารมีผลต่อเพศสภาพมากที่สุด เพราะมีค่า R square ที่สูงที่สุด