

Final Project



Airline Passenger Satisfaction

Siwat Photinam

Agenda

- **Overview and Objective**
- Data Exploration
- Data Cleansing and Preparation
- Data Analysis, Model and Visualization
- Conclusion

Overview and Objective

แม้ว่าปัจจุบัน ธุรกิจการบินได้รับผลกระทบจากไวรัส Covid-19 อย่างมาก แต่การวิเคราะห์ข้อมูล โดยการใช้ผลสำรวจความพึงพอใจของผู้โดยสาร เป็นสิ่งที่ต้องกระทำอย่างสม่ำเสมอ เพื่อปรับปรุงผลิตภัณฑ์/การบริการ เพื่อรักษา Market Share และรักษา Customer Engagement ต่อไป

คำถามสำคัญในการจัดทำ ได้แก่

- ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยแต่ละผลิตภัณฑ์/บริการอย่างไร? ทุกช่วงอายุมีความเชื่อมโยงในความต้องการด้านผลิตภัณฑ์/บริการที่สอดคล้องกันด้านใด?
- ข้อบ่งชี้สำคัญที่มีผลต่อความ Satisfied/Dissatisfied ที่เกิดจากผลิตภัณฑ์/การบริการใด

Agenda

- Overview and Objective
- **Data Exploration**
- Data Cleansing
- Data Analysis, Model and Visualization
- Conclusion

Data Exploration

เนื่องจากไม่สามารถเข้าถึงข้อมูลความพึงพอใจของผู้โดยสารจริงได้ จึงใช้ข้อมูลจาก Kaggle.com เป็น Structure Data โดยมีจำนวน 103,904 แถว 24 คอลัมน์ และสามารถจัดกลุ่มข้อมูลได้จำนวน 4 ประเภทข้อมูล ดังนี้

- ประเภทกลุ่มข้อมูลผู้โดยสาร ได้แก่ ID, Age, Class, Gender, Type of Travel
- ประเภทกลุ่มข้อมูลด้านความรู้สึก ได้แก่ Customer Type, Satisfaction
- ประเภทกลุ่มข้อมูลผลิตภัณฑ์/การให้บริการ ได้แก่ Baggage handling, Check-in Service, Cleanliness, Departure/Arrival time convenient, Ease of Online booking, Food and drink, Gate location, Inflight entertainment Inflight service, Inflight wifi service, Leg room service, On-board service, Online boarding, Seat comfort -- คะแนน '0' คือ Not Applicable --
- ประเภทกลุ่มข้อมูลทางการบิน ได้แก่ Arrival Delay in Minutes, Departure Delay in Minutes, Flight Distance

Agenda

- Overview and Objective
- Data Exploration
- **Data Cleansing and Preparation**
- Data Analysis, Model and Visualization
- Conclusion

Data Cleansing and Preparation (1)

Before cleansing & Preparation

```
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 24 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   id                                    103904 non-null int64
 1   Gender                                103904 non-null object
 2   Customer Type                         103904 non-null object
 3   Age                                    103904 non-null int64
 4   Type of Travel                        103904 non-null object
 5   Class                                 103904 non-null object
 6   Flight Distance                       103904 non-null int64
 7   Inflight wifi service                 103904 non-null int64
 8   Departure/Arrival time convenient    103904 non-null int64
 9   Ease of Online booking                103904 non-null int64
10   Gate location                         103904 non-null int64
11   Food and drink                       103904 non-null int64
12   Online boarding                       103904 non-null int64
13   Seat comfort                         103904 non-null int64
14   Inflight entertainment                103904 non-null int64
15   On-board service                     103904 non-null int64
16   Leg room service                     103904 non-null int64
17   Baggage handling                     103904 non-null int64
18   Checkin service                      103904 non-null int64
19   Inflight service                      103904 non-null int64
20   Cleanliness                          103904 non-null int64
21   Departure Delay in Minutes            103904 non-null int64
22   Arrival Delay in Minutes              103904 non-null int64
23   satisfaction                          103904 non-null object
dtypes: int64(19), object(5)
```

How are cleansing and preparation?

- **Drop Column 'ID'**
-- เป็น Auto Run ID ไม่ซ้ำกัน --
- **Convert & Mapping Column**
'Gender', 'Customer Type', 'Type of Travel', 'Class', 'satisfaction'
from object to int64
- **Convert & Grouping Column**
'Age' from object to int64
-- เพื่อจัดทำ Data Analysis & Model --
- **Drop Column**
'Flight Distance', 'Departure Delay in Minutes', 'Arrival Delay in Minutes'
-- ไม่ได้นำข้อมูลชุดนี้ไปจัดทำ Data Model --
- **Rename & Sort Index Column**
-- เพื่อจัดลำดับข้อมูล

Data Cleansing and Preparation (2)

After cleansing & Preparation

```
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   A_ID                                  103904 non-null int64
1   B_Age_Code                           103904 non-null int64
2   C_Gender_Code                        103904 non-null int64
3   D_Satisfied_Code                     103904 non-null int64
4   E_Pax_Code                           103904 non-null int64
5   F_Travel_Code                        103904 non-null int64
6   G_Class_Code                         103904 non-null int64
7   H_Baggage_handling                  103904 non-null int64
8   I_Checkin_service                   103904 non-null int64
9   J_Cleanliness                       103904 non-null int64
10  K_Departure_Arrival_Convenient      103904 non-null int64
11  L_Ease_Online_booking                103904 non-null int64
12  M_Food_drink                         103904 non-null int64
13  N_Gate_location                     103904 non-null int64
14  O_Inflight_entertainment             103904 non-null int64
15  P_Inflight_service                   103904 non-null int64
16  Q_Inflight_wifi_service               103904 non-null int64
17  R_Leg_room_service                  103904 non-null int64
18  S_Onboard_service                    103904 non-null int64
19  T_Online_boarding                    103904 non-null int64
20  U_Seat_comfort                       103904 non-null int64
dtypes: int64(21)
```

Example Data

➤ `pax01_df.iloc[0]`

```
A_ID                1
B_Age_Code          3
C_Gender_Code       1
D_Satisfied_Code    2
E_Pax_Code          2
F_Travel_Code       1
G_Class_Code        1
H_Baggage_handling  5
I_Checkin_service   4
J_Cleanliness       5
K_Departure_Arrival_Convenient  3
L_Ease_Online_booking  3
M_Food_drink        5
N_Gate_location     3
O_Inflight_entertainment  5
P_Inflight_service   5
Q_Inflight_wifi_service  3
R_Leg_room_service   2
S_Onboard_service    3
T_Online_boarding    3
U_Seat_comfort       5
Name: 0, dtype: int64
```

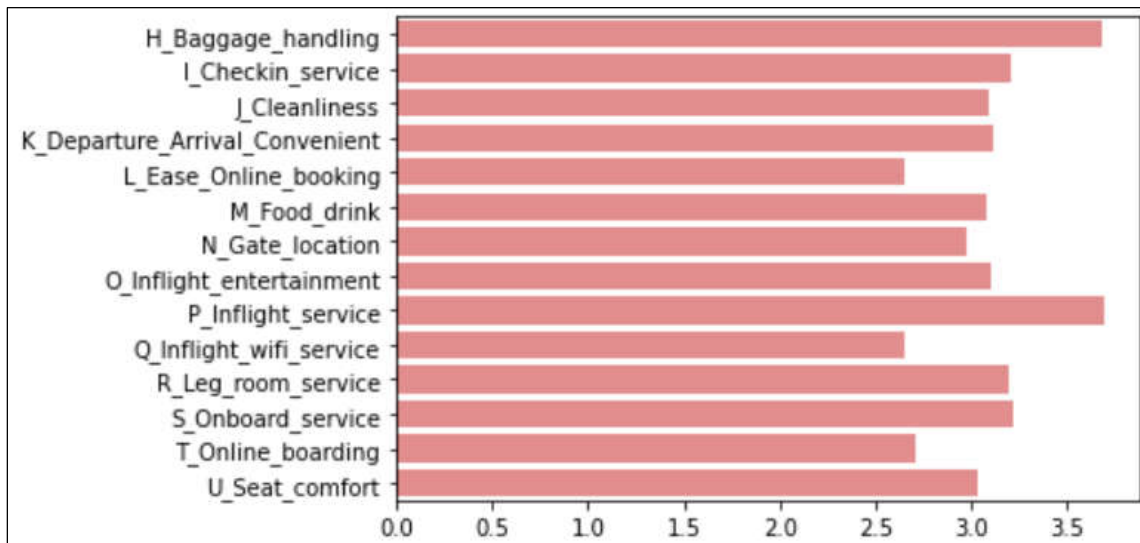

Agenda

- Overview and Objective
- Data Exploration
- Data Cleansing and Preparation
- **Data Analysis, Model and Visualization**
- Conclusion

Data Analysis, Model and Visualization (1)

➤ ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยในแต่ละผลิตภัณฑ์/บริการอย่างไร?

GenZ (7-22 Years)

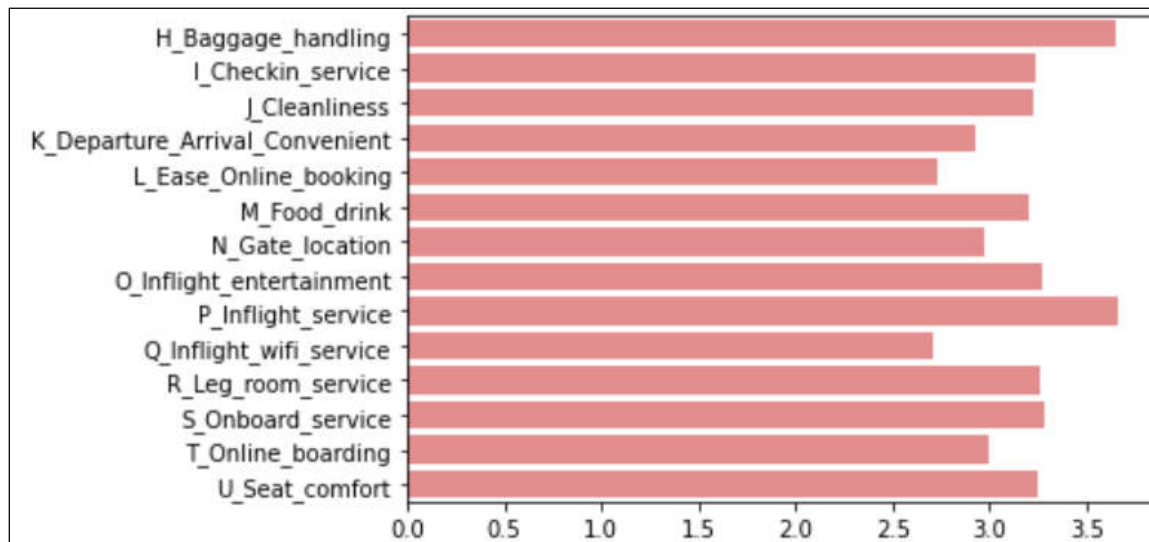


H_Baggage_handling	3.679152
I_Checkin_service	3.209861
J_Cleanliness	3.094859
K_Departure_Arrival_Convenient	3.116253
L_Ease_Online_booking	2.656507
M_Food_drink	3.078402
N_Gate_location	2.971233
O_Inflight_entertainment	3.099664
P_Inflight_service	3.690540
Q_Inflight_wifi_service	2.651965
R_Leg_room_service	3.196037
S_Onboard_service	3.221447
T_Online_boarding	2.714502
U_Seat_comfort	3.036535

Data Analysis, Model and Visualization (2)

➤ ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยในแต่ละผลิตภัณฑ์/บริการอย่างไร?

GenY (23-38 Years)

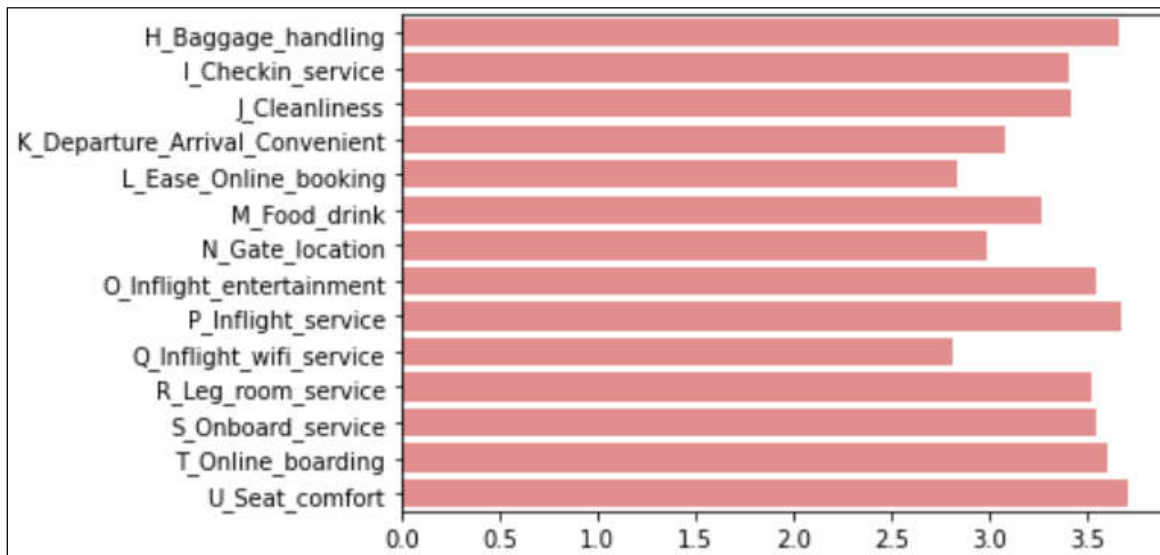


H_Baggage_handling	3.647625
I_Checkin_service	3.232619
J_Cleanliness	3.227202
K_Departure_Arrival_Convenient	2.931229
L_Ease_Online_booking	2.727141
M_Food_drink	3.201318
N_Gate_location	2.977638
O_Inflight_entertainment	3.268194
P_Inflight_service	3.658641
Q_Inflight_wifi_service	2.702853
R_Leg_room_service	3.260940
S_Onboard_service	3.281105
T_Online_boarding	2.997171
U_Seat_comfort	3.246795

Data Analysis, Model and Visualization (3)

➤ ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยในแต่ละผลิตภัณฑ์/บริการอย่างไร?

GenX (39-54 Years)

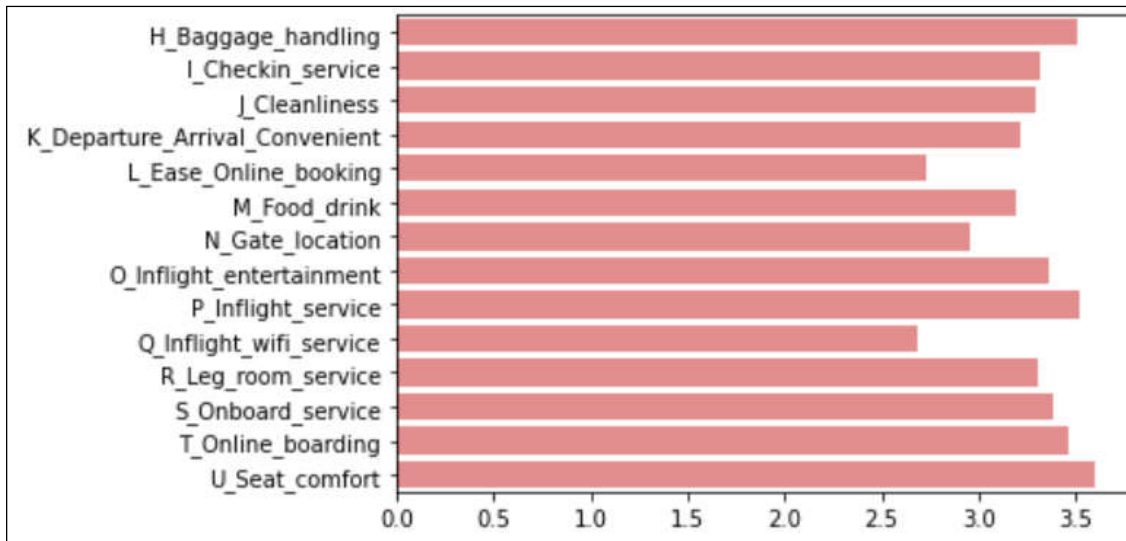


H_Baggage_handling	3.663177
I_Checkin_service	3.406626
J_Cleanliness	3.418078
K_Departure_Arrival_Convenient	3.077682
L_Ease_Online_booking	2.836101
M_Food_drink	3.262904
N_Gate_location	2.990266
O_Inflight_entertainment	3.548439
P_Inflight_service	3.669639
Q_Inflight_wifi_service	2.807962
R_Leg_room_service	3.525999
S_Onboard_service	3.543995
T_Online_boarding	3.599046
U_Seat_comfort	3.704894

Data Analysis, Model and Visualization (4)

➤ ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยในแต่ละผลิตภัณฑ์/บริการอย่างไร?

Boomer (55-73 Years)

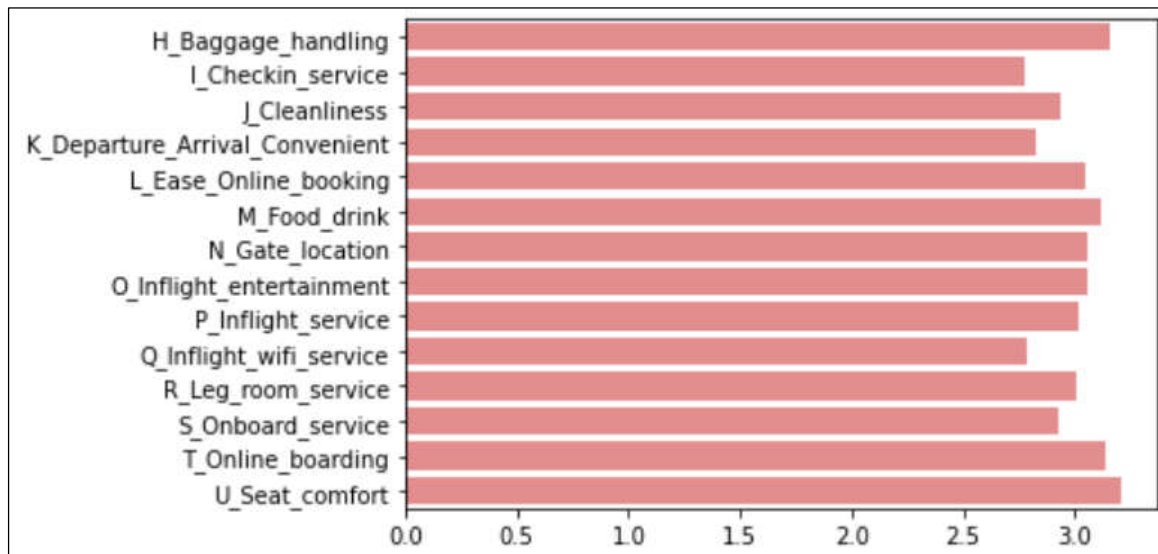


H_Baggage_handling	3.512281
I_Checkin_service	3.319585
J_Cleanliness	3.296598
K_Departure_Arrival_Convenient	3.217639
L_Ease_Online_booking	2.729269
M_Food_drink	3.186501
N_Gate_location	2.951853
O_Inflight_entertainment	3.361374
P_Inflight_service	3.521954
Q_Inflight_wifi_service	2.685089
R_Leg_room_service	3.300783
S_Onboard_service	3.386099
T_Online_boarding	3.457505
U_Seat_comfort	3.595805

Data Analysis, Model and Visualization (5)

➤ แต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยในแต่ละผลิตภัณฑ์/บริการอย่างไร?

Silent (From 74 Years)



H_Baggage_handling	3.160976
I_Checkin_service	2.770732
J_Cleanliness	2.931707
K_Departure_Arrival_Convenient	2.829268
L_Ease_Online_booking	3.043902
M_Food_drink	3.117073
N_Gate_location	3.051220
O_Inflight_entertainment	3.060976
P_Inflight_service	3.012195
Q_Inflight_wifi_service	2.782927
R_Leg_room_service	3.004878
S_Onboard_service	2.924390
T_Online_boarding	3.139024
U_Seat_comfort	3.204878

Data Analysis, Model and Visualization (6)

➤ ทุกช่วงอายุมีความเชื่อมโยงในความต้องการด้านผลิตภัณฑ์/บริการที่สอดคล้องกันด้านใด?

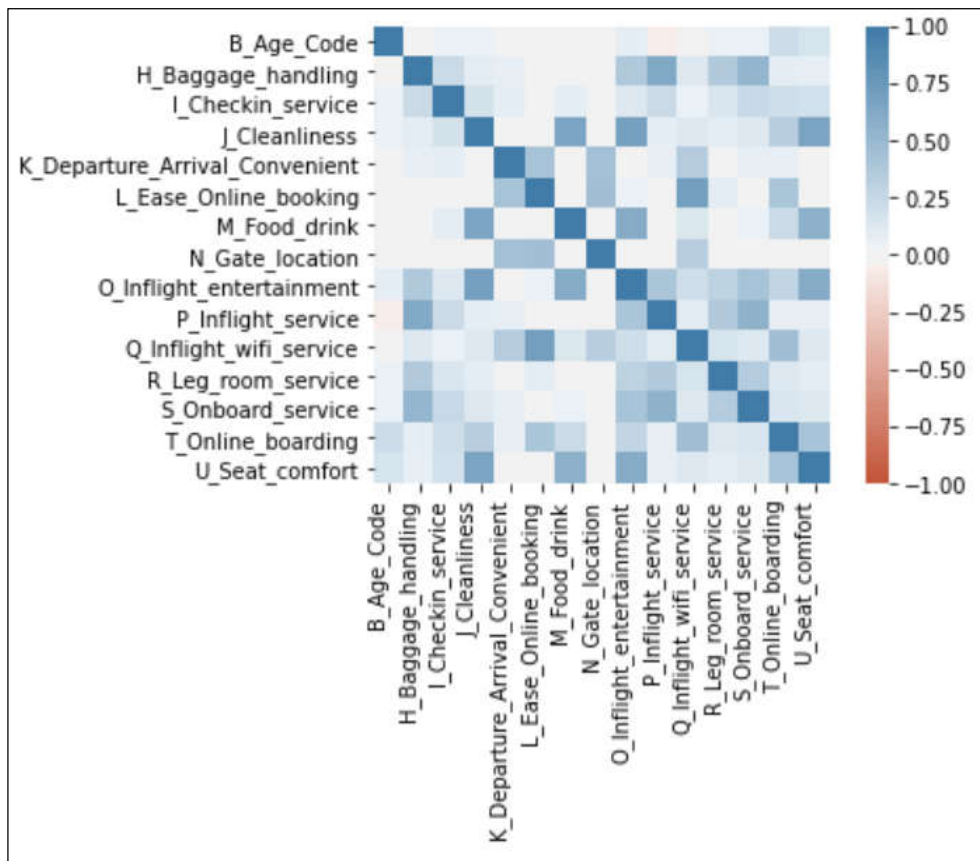
Correlation Matrix (With Correlation Information)

	B_Age_Code	H_Baggage_handling	I_Checkin_service	J_Cleanliness	K_Departure_Arrival_Convenient	L_Ease_Online_booking	M_Food_drink	N_Gate_location	O_Inflight_entertainment	P_Inflight_service	Q_Inflight_wifi_service	R_Leg_room_service	S_Onboard_service	T_Online_booking	U_Seat_comfort
B_Age_Code	1	-0.038753	0.040506	0.058892	0.034928	0.027085	0.025947	-0.001993	0.080448	-0.041276	0.018759	0.049672	0.062059	0.211846	0.162659
H_Baggage_handling	-0.038753	1	0.233122	0.095793	0.072126	0.038762	0.034746	0.002313	0.37821	0.628561	0.120923	0.369544	0.519134	0.08328	0.074542
I_Checkin_service	0.040506	0.233122	1	0.179583	0.093333	0.011081	0.087299	-0.035427	0.120867	0.237197	0.043193	0.153137	0.243914	0.204462	0.191854
J_Cleanliness	0.058892	0.095793	0.179583	1	0.014292	0.016179	0.65776	-0.00383	0.691815	0.088779	0.132698	0.09637	0.12322	0.331517	0.678534
K_Departure_Arrival_Convenient	0.034928	0.072126	0.093333	0.014292	1	0.436961	0.004906	0.444757	-0.004861	0.073318	0.343845	0.012441	0.068882	0.070119	0.011344
L_Ease_Online_booking	0.027085	0.038762	0.011081	0.016179	0.436961	1	0.031873	0.458655	0.047032	0.035272	0.715856	0.107601	0.038833	0.404074	0.030014
M_Food_drink	0.025947	0.034746	0.087299	0.65776	0.004906	0.031873	1	-0.001159	0.622512	0.033993	0.134718	0.032498	0.059073	0.234468	0.574556
N_Gate_location	-0.001993	0.002313	-0.035427	-0.00383	0.444757	0.458655	-0.001159	1	0.003517	0.001681	0.336248	-0.005873	-0.028373	0.001688	0.003669
O_Inflight_entertainment	0.080448	0.37821	0.120867	0.691815	-0.004861	0.047032	0.622512	0.003517	1	0.404855	0.209321	0.299692	0.420153	0.285066	0.61059
P_Inflight_service	-0.041276	0.628561	0.237197	0.088779	0.073318	0.035272	0.033993	0.001681	0.404855	1	0.110441	0.368656	0.550782	0.074573	0.069218
Q_Inflight_wifi_service	0.018759	0.120923	0.043193	0.132698	0.343845	0.715856	0.134718	0.336248	0.209321	0.110441	1	0.160473	0.1215	0.45697	0.122658
R_Leg_room_service	0.049672	0.369544	0.153137	0.09637	0.012441	0.107601	0.032498	-0.005873	0.299692	0.368656	0.160473	1	0.355495	0.12395	0.105559
S_Onboard_service	0.062059	0.519134	0.243914	0.12322	0.068882	0.038833	0.059073	-0.028373	0.420153	0.550782	0.1215	0.355495	1	0.155443	0.131971
T_Online_booking	0.211846	0.08328	0.204462	0.331517	0.070119	0.404074	0.234468	0.001688	0.285066	0.074573	0.45697	0.12395	0.155443	1	0.420211
U_Seat_comfort	0.162659	0.074542	0.191854	0.678534	0.011344	0.030014	0.574556	0.003669	0.61059	0.069218	0.122658	0.105559	0.131971	0.420211	1

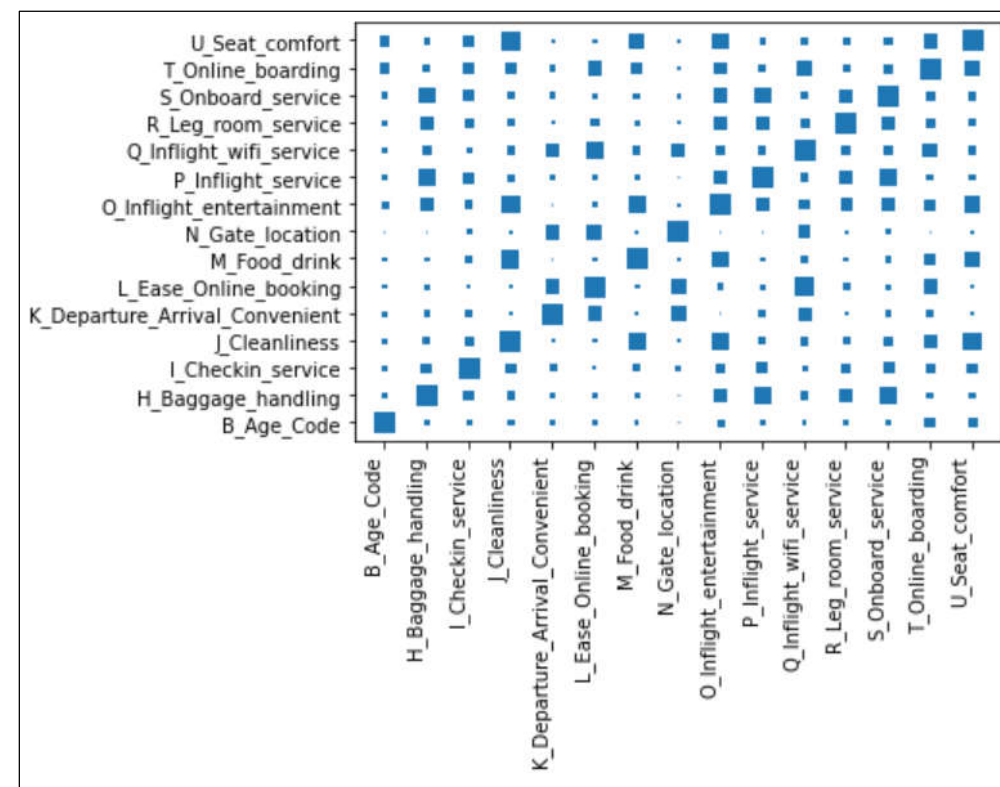
Data Analysis, Model and Visualization (7)

➤ ทุกช่วงอายุมีความเชื่อมโยงในความต้องการด้านผลิตภัณฑ์/บริการที่สอดคล้องกันด้านใด?

Correlation Matrix (Heat Map)



Correlation Matrix (Scatter Plot)



Data Analysis, Model and Visualization (8)

- ข้อบ่งชี้สำคัญที่มีผลต่อความ Satisfied/Dissatisfied เกิดจากผลิตภัณฑ์/การบริการด้านใด
กำหนดให้ **X** คือ ผลิตภัณฑ์/การบริการของสายการบิน และ **y** คือ ความพึงพอใจ/ไม่พึงพอใจ

Decision Tree

X_train

T_Online_boarding	0.352604
Q_Inflight_wifi_service	0.214191
R_Leg_room_service	0.078315
K_Departure_Arrival_Convenient	0.043456
M_Food_drink	0.043453
O_Inflight_entertainment	0.040566
N_Gate_location	0.037859
J_Cleanliness	0.032730
I_Checkin_service	0.028353
P_Inflight_service	0.027063
H_Baggage_handling	0.026948
U_Seat_comfort	0.026575
S_Onboard_service	0.024266
L_Ease_Online_booking	0.023621

y_train

2	39434
1	30181
Name: D_Satisfied_Code, dtype: int64	

Performance

0.922336609408265					
	precision	recall	f1-score	support	
1	0.91	0.91	0.91	14844	
2	0.93	0.93	0.93	19445	
accuracy			0.92	34289	
macro avg	0.92	0.92	0.92	34289	
weighted avg	0.92	0.92	0.92	34289	

Prediction

col_0	1	2
D_Satisfied_Code		
1	13578	1266
2	1397	18048

Agenda

- Overview and Objective
- Data Exploration
- Data Cleansing and Preparation
- Data Analysis, Model and Visualization
- Conclusion

Conclusion (1)

①ผู้โดยสารในแต่ละช่วงอายุ (Generation) มีความพึงพอใจเฉลี่ยแต่ละผลิตภัณฑ์/บริการอย่างไร?

➢เนื่องจากอายุในข้อมูลมีความหลากหลาย จึงทำการกำหนดเป็น Generation พบว่าใน Generation Z และ Y ให้ความพึงพอใจเฉลี่ยกับ Inflight Service มากที่สุด แต่เมื่อช่วงอายุเพิ่มขึ้นเป็น Generation X, Boomer, Silent ให้ความพึงพอใจเฉลี่ยกับ Seat Comfortable มากที่สุด แต่พบว่า Generation Z, Y, X, Boomer ให้ความพึงพอใจเฉลี่ยกับ Inflight Wi-Fi Service น้อยที่สุด ยกเว้น Silent ให้ความพึงพอใจเฉลี่ยกับ Check-in Service น้อยที่สุด

②ทุกช่วงอายุมีความเชื่อมโยงในความต้องการด้านผลิตภัณฑ์/บริการที่สอดคล้องกันด้านใด?

➢เมื่อใช้ Correlation Matrix (Heat Map, Scatter Plot) นำทุก Generation เทียบกับผลิตภัณฑ์/บริการที่มี พบว่า Online-Boarding เป็นสิ่งสัมพันธ์มากที่สุด และ Inflight Service เป็นสิ่งที่มีสัมพันธ์น้อยที่สุด

Conclusion (2)

③ เชื่อมโยงคำถามที่ 1 และ 2

- เมื่อนำความพึงพอใจเฉลี่ยสูงสุดเทียบกับค่าความสัมพันธ์ พบว่า Seat Comfortable เป็นไปในทิศทางเดียวกัน จึงเป็นจุดเด่นของสายการบินที่ต้องรักษามาตรฐานสม่ำเสมอ แต่ Inflight Service ไม่เป็นไปในทิศทางเดียวกัน จึงเป็นจุดเด่นที่สายการบินต้องกลับไปพิจารณาเพื่อหาจุดสมดุลของการให้บริการรูปแบบนี้
- เมื่อนำความพึงพอใจเฉลี่ยต่ำสุดเทียบกับค่าความสัมพันธ์ พบว่า Inflight Wi-Fi Service และ Check-in Service เป็นไปในทิศทางเดียวกัน จึงเป็นจุดด้อยของสายการบินที่ต้องแก้ไขเร่งด่วน เพื่อให้เกิดความพึงพอใจเฉลี่ยที่เพิ่มขึ้น
- เมื่อนำค่าความสัมพันธ์สูงสุดเทียบกับความพึงพอใจเฉลี่ย พบว่า Online Boarding เป็นไปในทิศทางเดียวกัน และพบว่า Generation Z, Y ให้ความพึงพอใจเฉลี่ยต่อบริการนี้อยู่ในลำดับท้ายๆ ของการให้บริการ/ผลิตภัณฑ์ที่มีทั้งหมด

Conclusion (3)

④ ข้อบ่งชี้สำคัญที่มีผลต่อความ Satisfied/Dissatisfied เกิดจากผลิตภัณฑ์/การบริการใด

➢ จากการแบ่งประเภทกลุ่มข้อมูลและจากการวิเคราะห์โดยการเทียบค่าความพึงพอใจเฉลี่ยและค่าความสัมพันธ์ จึงได้ทำการเลือกกลุ่มข้อมูลผลิตภัณฑ์/การให้บริการ เปรียบเทียบกับกลุ่มข้อมูลด้านความรู้สึก โดยใช้วิธีการ Decision Tree เพื่อดูข้อบ่งชี้สำคัญที่มีผลต่อความ Satisfied/Dissatisfied ที่เกิดจากผลิตภัณฑ์/การบริการใด

➢ เมื่อทำการแบ่งกลุ่มข้อมูล test (33%) และที่เหลือเป็น train กำหนด random state เท่ากับ 42 และใช้ criterion เป็น Gini Index พบว่าประสิทธิภาพในการทำนายและผลลัพธ์ที่ได้จากการทำนายอยู่ในเกณฑ์ดี โดยสามารถนำไปใช้ในวิเคราะห์ได้

➢ อย่างไรก็ตาม ผู้จัดทำยังไม่สามารถ Export รูป Decision Tree ออกมาได้จึงยังไม่สามารถหาข้อบ่งชี้สำคัญที่มีผลต่อความ Satisfied/Dissatisfied เกิดจากผลิตภัณฑ์/การบริการใด เมื่อมีความพึงพอใจอยู่ในระดับใด

Conclusion (4)

⑤ สิ่งที่ต้องปรับ พัฒนาต่อไป

- การศึกษาและพัฒนาการเขียน Data Model รูปแบบอื่น เพื่อเปรียบเทียบผลลัพธ์และเลือกใช้ Data Model ที่มีประสิทธิภาพในการทำนายและค้นหา Insight ของข้อมูลที่เหมาะสมที่สุด
- การศึกษาและพัฒนาการเขียนเพื่อหำนำข้อมูลที่เป็น Un-Structure มาทำการวิเคราะห์ได้ เช่น ถอดข้อมูลที่เป็นข้อความมาหาความสัมพันธ์ในด้านความพึงพอใจหรือไม่พึงพอใจต่อบริการ/ผลิตภัณฑ์ต่างๆ ของบริษัท