航空業旅客滿意度調查

109-2 程式設計概論 期末專題報告

第13組

斯語四 羅珮瑄

資料名稱：Passenger Satisfaction -US Airline Passenger Satisfaction

資料來源： Kaggle（https://www.kaggle.com/tejmahal20/airline-passenger-satisfaction）

資料簡介:

- Train data (103904筆) + Test data (25976 筆) = 總計 129880 筆
- 總計有 23 個 columns
- 三種變數類型: 類別變數 (5) 數值變數 (4) 次序變數 (14)

df.head()

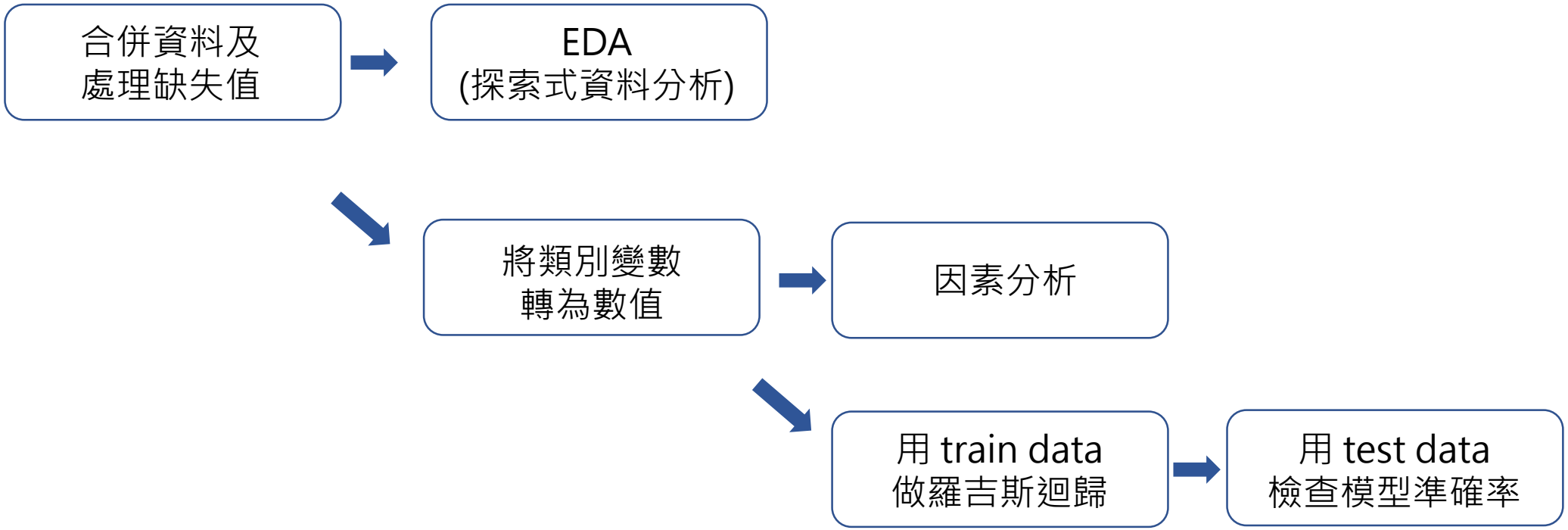
Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	3	1	...	5	4	3	4	4	5	5	25	18	neutral or dissatisfied
Male	disloyal Customer	25	Business travel	Business	235	3	2	3	3	...	1	1	5	3	1	4	1	1	6	neutral or dissatisfied
Female	Loyal Customer	26	Business travel	Business	1142	2	2	2	2	...	5	4	3	4	4	4	5	0	0	satisfied
Female	Loyal Customer	25	Business travel	Business	562	2	5	5	5	...	2	2	5	3	1	4	2	11	9	neutral or dissatisfied
Male	Loyal Customer	61	Business travel	Business	214	3	3	3	3	...	3	3	4	4	3	3	3	0	0	satisfied

專題題目：影響航空業旅客滿意度之重要因素及滿意度預測

- 整合問卷內容找出重要因素

- 建立能預測滿意度的模型

流程圖簡介


合併資料及
處理缺失值
EDA
(探索式資料分析)
將類別變數
轉為數值
因素分析
用 train data
做羅吉斯迴歸
用 test data
檢查模型準確率

處理缺失值: 直接移除該行資料

將類別變數轉為數值

// by LebalEncoder

gender >>> Male:1 /Female:0

customer type >>> Loyal Customer:0 / disloyal Customer: 1

type of travel >>> Personal Travel:0 / Business travel:1

class >>> Business:0 / Eco: 1 / Eco Plus: 2

satisfaction >>> neutral or dissatisfied:0 / satisfied:1

Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
1	0	13	1	2	460	3	4	3	1	...	5	4	3	4	4	5	5	25	18	0
1	1	25	0	0	235	3	2	3	3	...	1	1	5	3	1	4	1	1	6	0
0	0	26	0	0	1142	2	2	2	2	...	5	4	3	4	4	4	5	0	0	1
0	0	25	0	0	562	2	5	5	5	...	2	2	5	3	1	4	2	11	9	0
1	0	61	0	0	214	3	3	3	3	...	3	3	4	4	3	3	3	0	0	1

EDA:變數名稱簡介

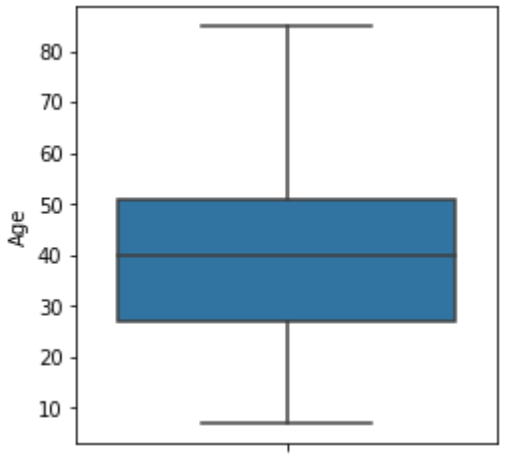
乘客個人基本資訊(3)
Gender
Customer Type
Age

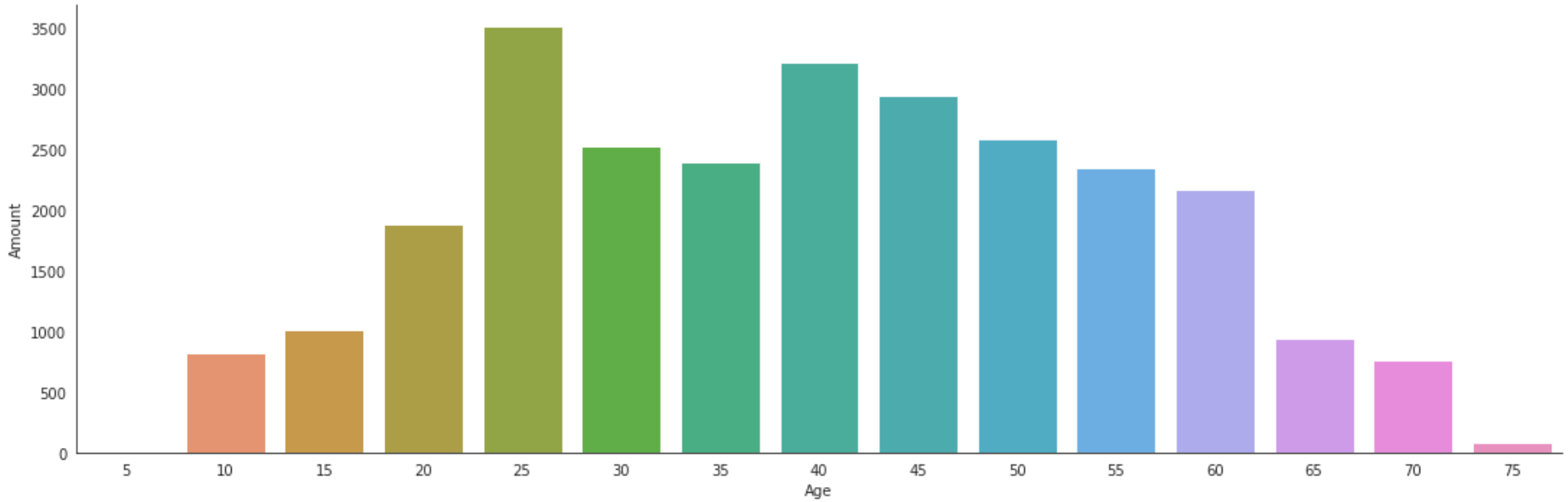
乘客搭乘資訊(5)
Type of Travel
Class
Flight Distance
Departure Delay in Minutes
Arrival Delay in Minutes

搭乘體驗評價(14)	
Inflight wifi service	Inflight entertainment
Departure/Arrival time convenient	On-board service
Ease of Online booking	Leg room service
Gate location	Baggage handling
Food and drink	Checkin service
Online boarding	Inflight service
Seat comfort	Cleanliness

滿意度(1)
satisfaction

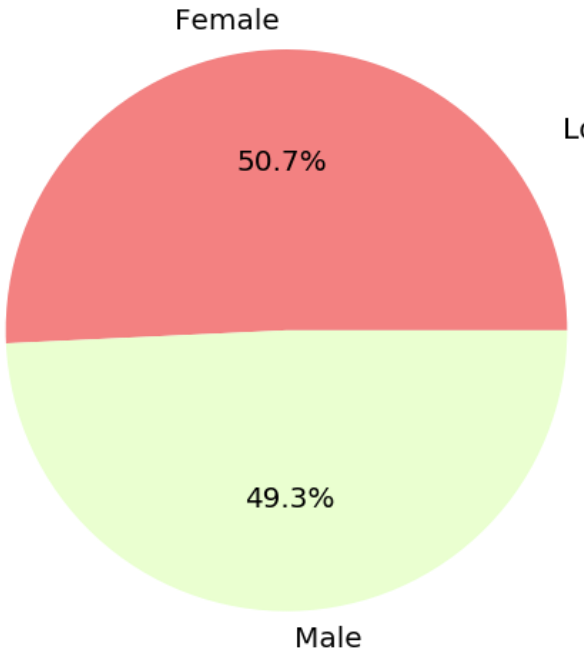
EDA: 乘客個人基本資訊



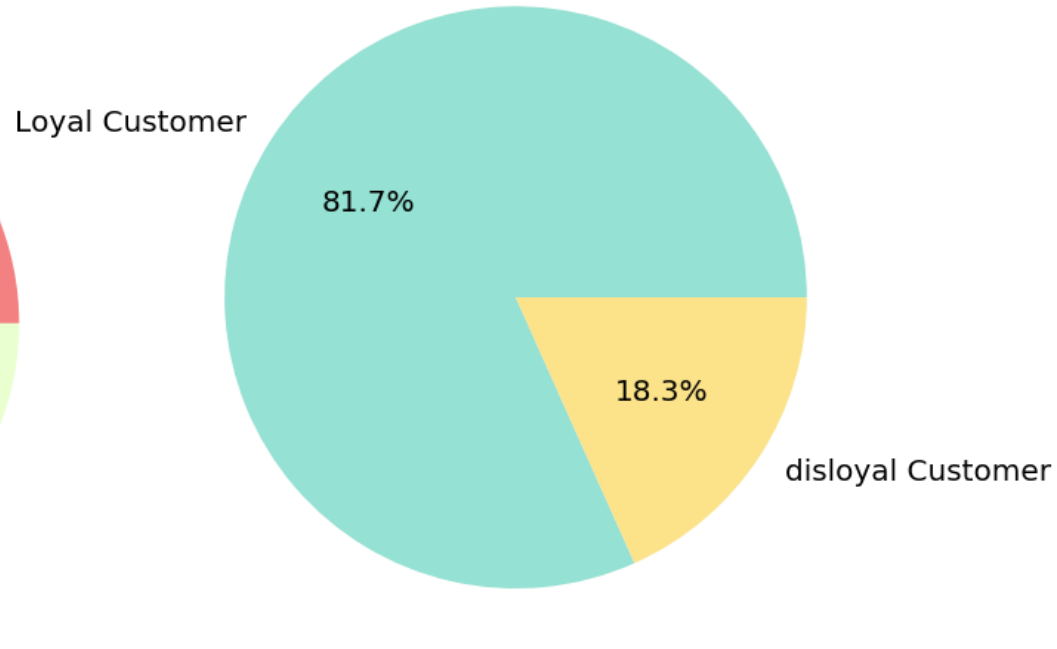


乘客年齡分布

EDA: 乘客個人基本資訊

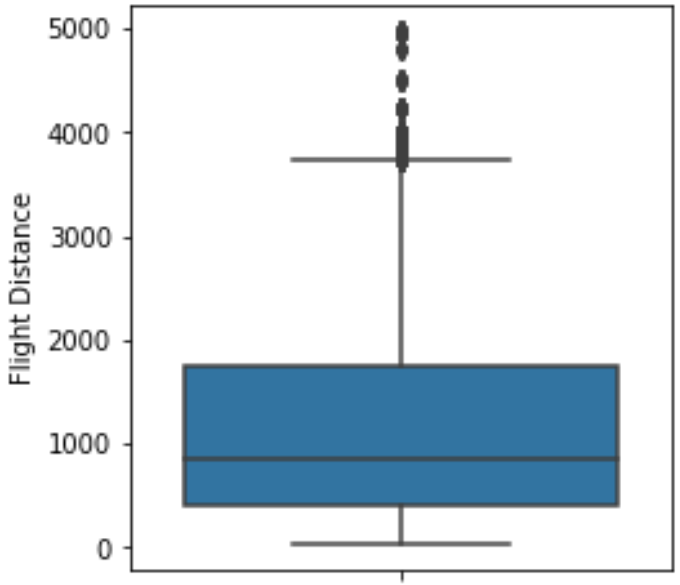


性別比例

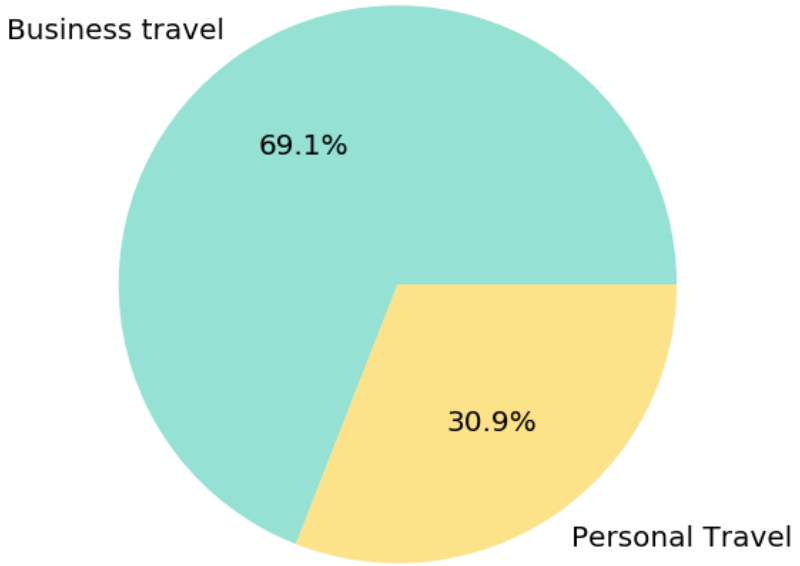


忠誠客戶比例

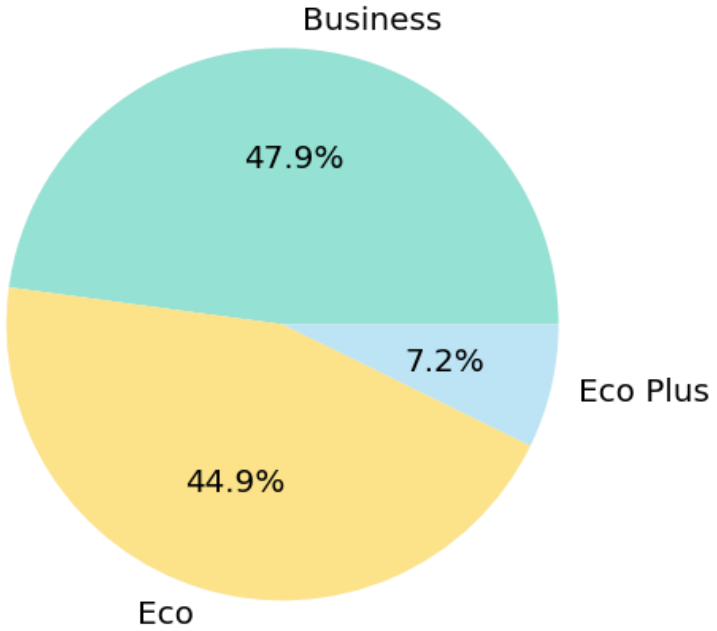
EDA: 乘客搭乘資訊



飛行里程數



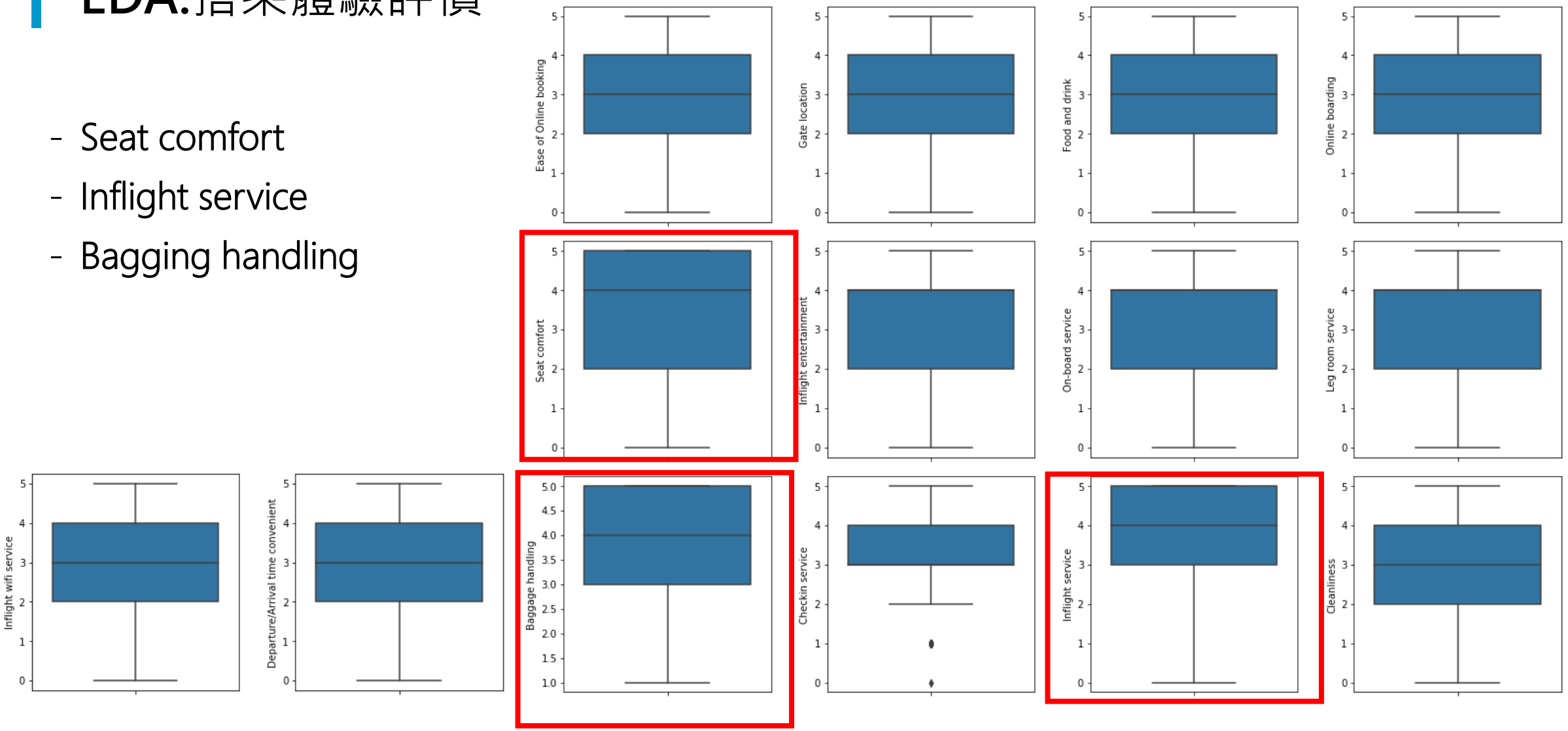
飛行目的



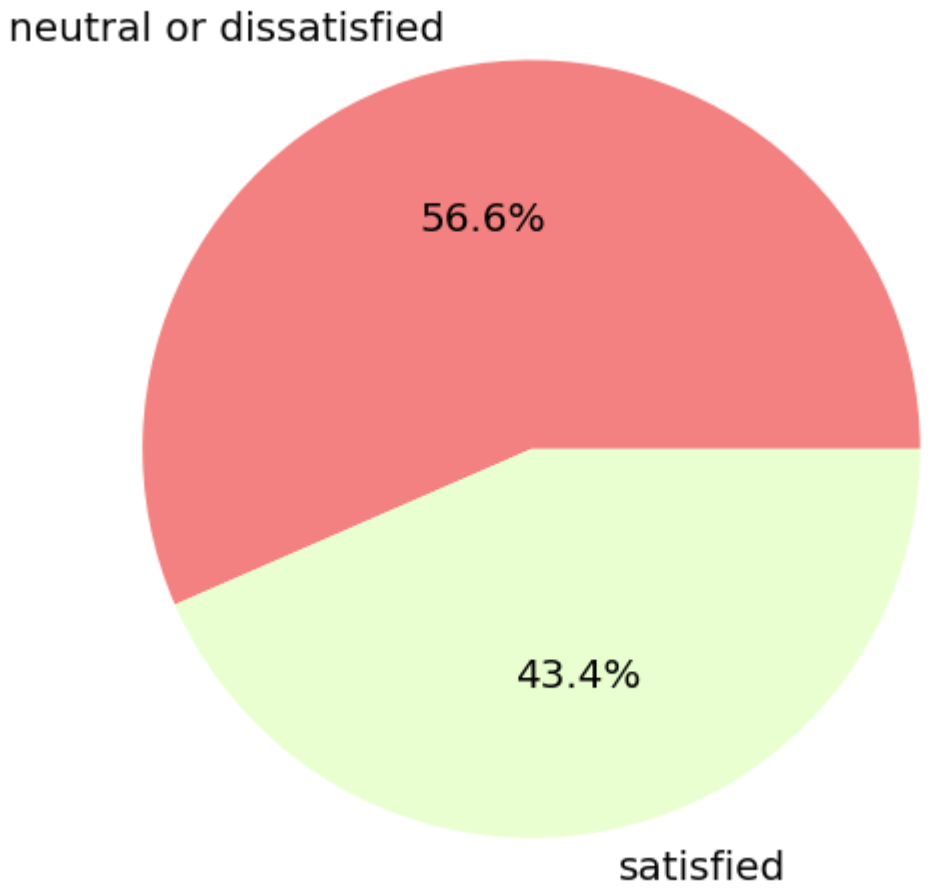
乘坐艙等

EDA:搭乘體驗評價

- Seat comfort
- Inflight service
- Bagging handling



EDA: 滿意度比例

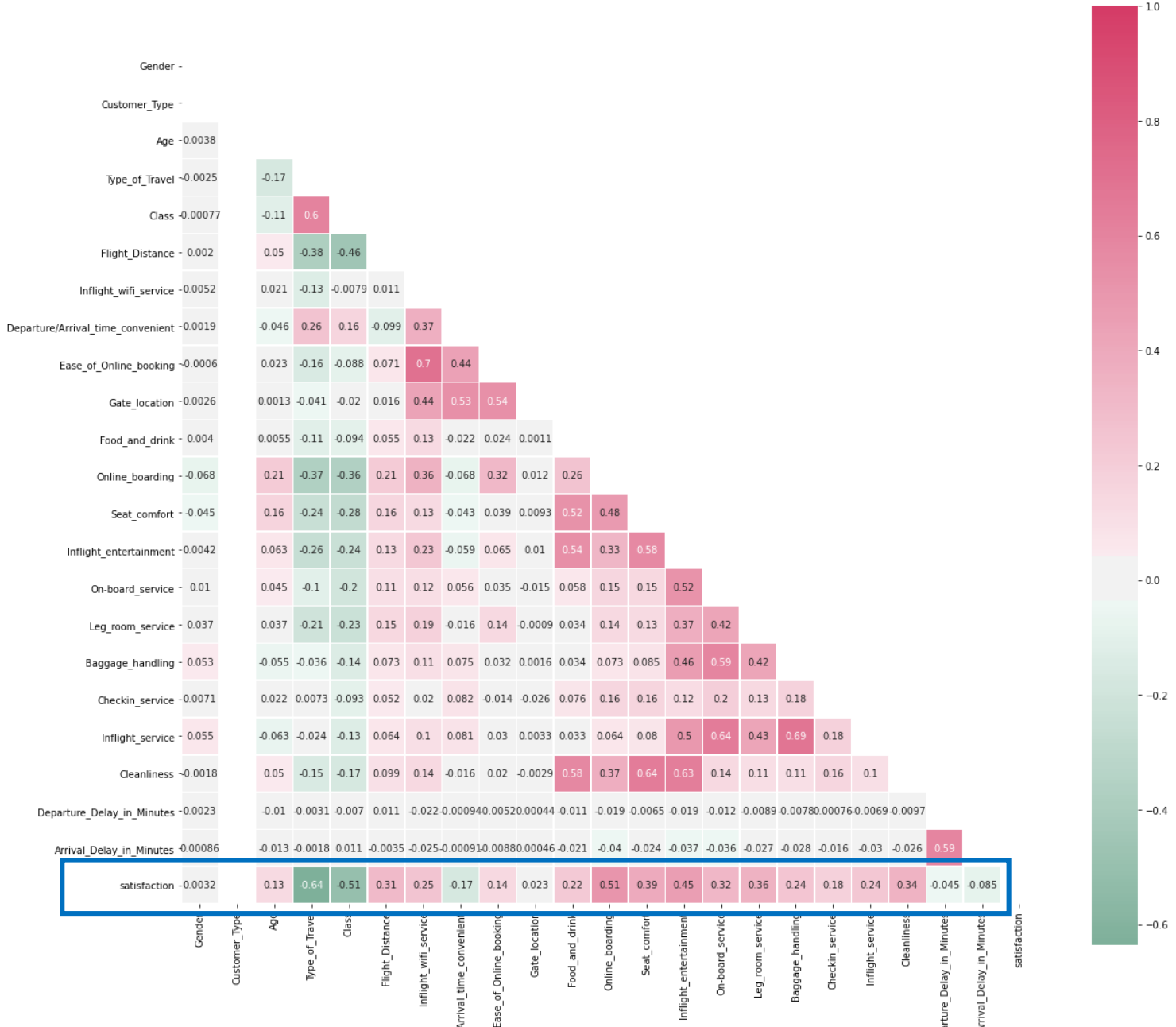


EDA: 相關係數

各變數與**滿意度**的相關程度

正相關:

- online broading(0.51)
- Inflight entertainment(0.45)



因素分析

```
fa = FactorAnalyzer()
df1 = df.drop(["satisfaction","Gender","Flight Distance","Class","Type of Travel","Age","Customer Type"], axis=1)
fa.fit(df1)
ev, v = fa.get_eigenvalues() #eigenvalues
print(ev)
```

Factor 1:機上飲食及設備

(Cleanliness / Seat comfort / Inflight entertainment / Food and drink)

Factor 2:服務

(Seat comfort / Inflight service / Baggage handling / On-board service / Inflight wifi service)

Factor 3:便利性

(Inflight wifi service / Ease of Online booking / Gate location / Departure/Arrival time convenient):

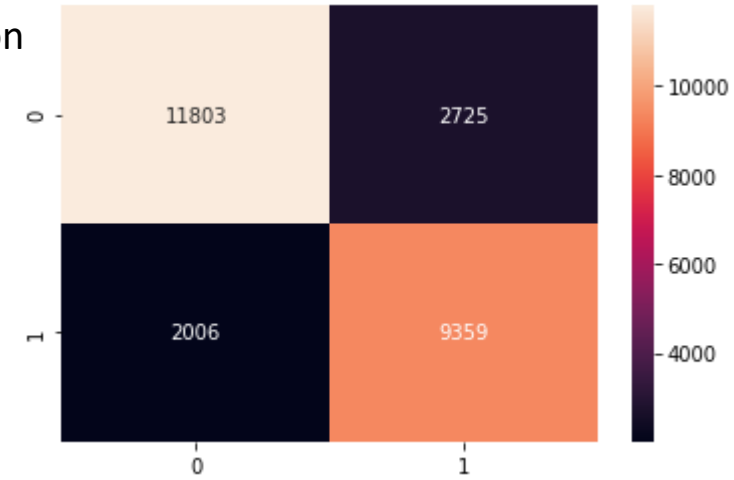
Factor 4:班機延誤情形

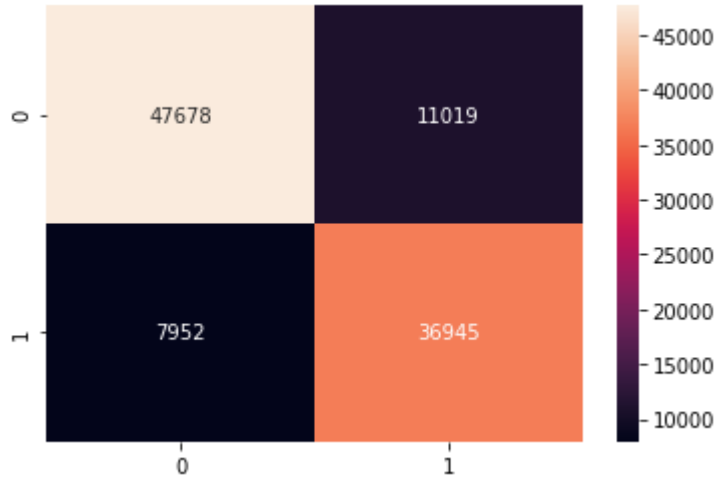
(Departure Delay in Minutes / Arrival Delay in Minutes)

	Factor 1	Factor 2	Factor 3	Factor 4
Inflight wifi service	0.167059	0.129103	0.758434	-0.010062
Departure/Arrival time convenient	-0.032744	0.058272	0.503835	0.001029
Ease of Online booking	0.030221	0.023083	0.930572	-0.00326
Gate location	-0.034743	-0.032968	0.507054	0.006394
Food and drink	0.753137	0.012164	0.004244	-0.016821
Online boarding	0.393821	0.114393	0.360317	-0.012785
Seat comfort	0.791138	0.081543	0.026203	-0.013805
Inflight entertainment	0.745879	0.468087	0.010898	-0.006612
On-board service	0.090762	0.701895	0.027572	-0.018924
Leg room service	0.073675	0.484083	0.080573	0.023345
Baggage handling	0.023219	0.766125	0.026458	0.007819
Checkin service	0.140255	0.285337	0.026927	-0.01347
Inflight service	0.017757	0.799698	0.01666	-0.043263
Cleanliness	0.857953	0.090292	-0.003341	0.00154
Departure Delay in Minutes	-0.016937	-0.015028	-0.001216	0.967434
Arrival Delay in Minutes	-0.018951	-0.020189	-0.002827	0.997112

羅吉斯迴歸

```
# logistic Regression
from sklearn.linear_model import LogisticRegression
X_train = train.drop("satisfaction", axis=1)
y_train = train["satisfaction"]
X_test = test.drop("satisfaction", axis=1)
y_test = test["satisfaction"]
print("X_train",len(X_train))
print("X_test",len(X_test))
print("y_train",len(y_train))
print("y_test",len(y_test))
print("test",len(test))
```






```
logreg=LogisticRegression()
logreg.fit(X_train,y_train)
acc_log_train=round(logreg.score(X_train,y_train)*100,2)
acc_log_test=round(logreg.score(X_test,y_test)*100,2)
print("Training Accuracy: % {}".format(acc_log_train))
print("Test Accuracy: % {}".format(acc_log_test))
```

Training Accuracy	81.69 %
Test Accuracy	81.73 %

結論

航空公司可以根據分析結果中針對旅客較在乎的服務
或設備優先進行改善

e.g. 針對飲食和設備進行改善調整

延伸的可能分析

- 將資料分群 (e.g K-means) >>> 再針對用戶特性規劃行銷策略
- 多試不一樣的監督式學習，再進一步比較預測最精準的模型
  (e.g. 隨機森林、SVM)

後續改善： 因應疫情新增衛生消毒保護措施相關的問卷內容

e.g. 安全旅行晴雨表 (Safe Travel Barometer)

	Name	Safe Travel Score *	Disinfection Frequency	Thermal Screening	Face Masks	Health Declaration Form	Staff Face Masks
	Qatar Airways	4.5	After Every Flight	✓	Provided	✓	✓
	Emirates	4.5	After Every Flight	✓	Provided	✓	✓
	Etihad Airways	4.4	After Every Flight	✓	Provided	✓	✓
	China Southern Airlines	4.1	After Every Flight	✓	Provided	✓	✓

Reference:

https://dashboard.safetravelbarometer.com/airlines?bucket=COVID-19%20Safety%20Protocols&page=1&perPage=10&tag=Full%20Service%20Carrier

Reference:

- https://www.kaggle.com/mig555/mushroom-classification
- https://www.kaggle.com/teejmahal20/classification-predicting-customer-satisfaction