

Predictive Analysis on Airline Passenger Satisfaction with Classification Methods



UIC Team

TAN JING
WEI DONGLIN
YANG ZIJING
CHEN XIAOFENG

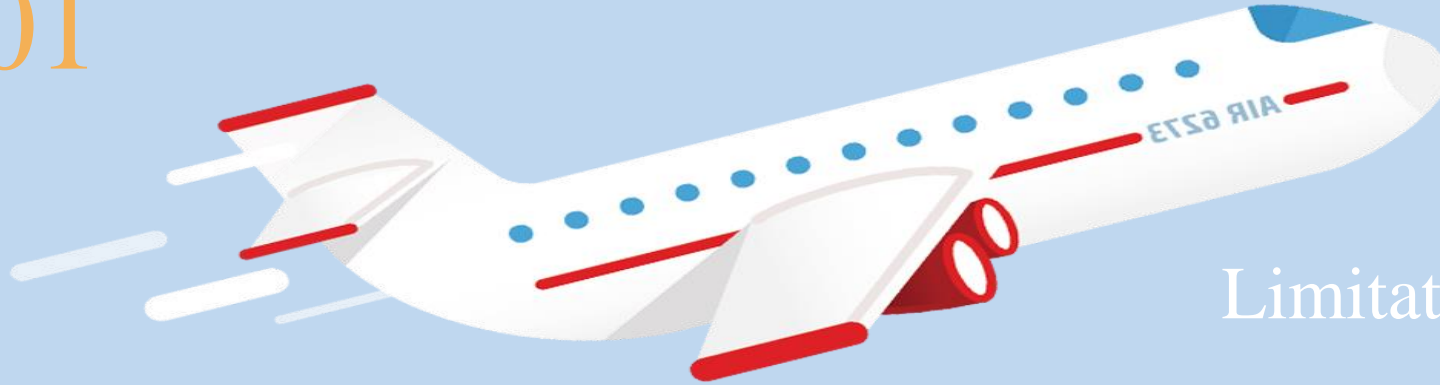
CONTENTS

Background & Purpose

01

Data pre-processing

02



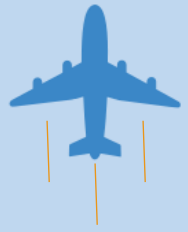
03
Data Description

04
Data processing

05
Limitation & Conclusion

01 Background & Purpose





Background



Demand for high quality services of air travel **has increased**.

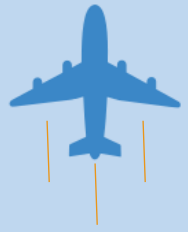


Airlines should focus on **improving the quality of services**.



Good services can help airlines win customer's heart and stand out in the market.





Background



Customer satisfaction plays a crucial role in **consuming behavior** and **the brand image of airlines**



Maintaining customer satisfaction at a good level can help the airline **occupy the aviation market** and get plenty of **repeat customers**.





Background



Nowadays, the aviation industry was hit hard by the Covid-19.




In order to profit and reduce the loss, some airlines introduce “flights to nowhere”


People can enjoy the beautiful views and good services during the flight.





Missing Flying Experience? Singapore Airlines Introduces ‘No-Destination’ Flights



 Image Source - Singapore Airlines/Twitter

QANTAS 7-HOUR FLIGHT TO NOWHERE SELLS OUT IN 10 MINUTES



Monday, September 21, 2020



"I cannot wait to go fly." Press play to hear from one of the pilots about what to expect on a flight with no destination.

Pre-pandemic, many of us viewed flying simply as a method of getting from one destination to another as quickly as possible.



Purpose of Research

The purpose of this research is to find out
what factors lead to customer satisfaction for an Airline.



02

Data pre-processing





Data collection

<https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction>

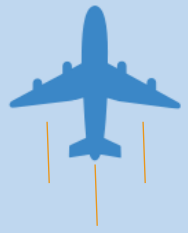
Kaggle Website



Dataset

training data: 103,904 records
testing data: 25,976 records

Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4 ...	5	4	3	4	4
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2 ...	1	1	5	3	1
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2 ...	5	4	3	4	4
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5 ...	2	2	5	3	1
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3 ...	3	3	4	4	3



Data cleaning

Process

- Drop outliers
- Drop missing value
- transfer all the categorical values into numerical value



Categorical value

Gender Type

1: Female 0: Male

Customer Type

1: Loyal customer 0: Disloyal customer

Travel Type

1: Business Travel 0: Personal Travel

Seat Class Type

2: Business Class 1: Economic Plus Class

0: Economic Class

Satisfaction Attitude Type

1: Satisfied Attitude

0: Neutral or Dissatisfied Attitude

03 Data Description





Data Description

General Trend of Dataset

Description of Target Feature

Count Plot Description

Numerical and
Categorical Features

Correlation



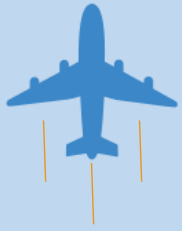


Numerical and Categorical Features

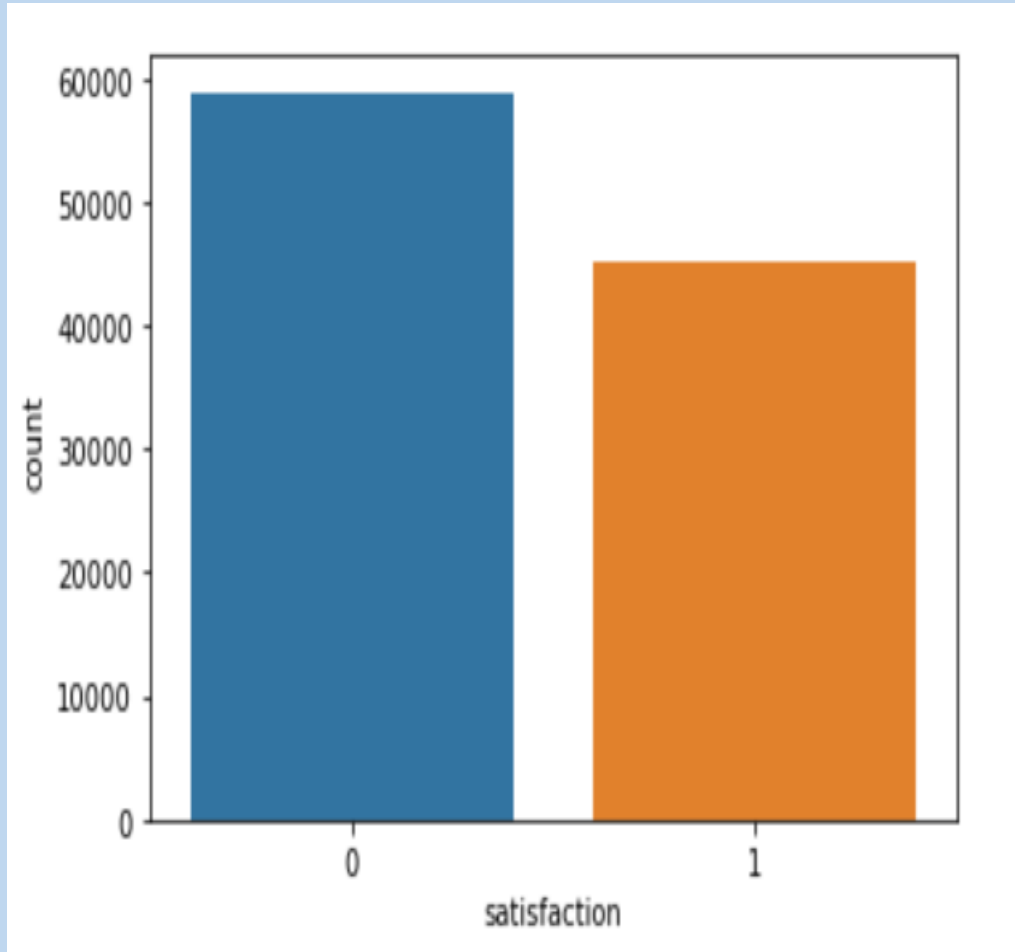
	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance
count	103594.000000	103594.000000	103594.000000	103594.000000	103594.000000	103594.000000
mean	0.507520	0.817248	39.380466	0.689857	1.028380	1189.325202
std	0.499946	0.386465	15.113125	0.462554	0.962868	997.297235
min	0.000000	0.000000	7.000000	0.000000	0.000000	31.000000
25%	0.000000	1.000000	27.000000	0.000000	0.000000	414.000000
50%	1.000000	1.000000	40.000000	1.000000	1.000000	842.000000
75%	1.000000	1.000000	51.000000	1.000000	2.000000	1743.000000
max	1.000000	1.000000	85.000000	1.000000	2.000000	4983.000000

	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
count	103594.000000	103594.000000	103594.000000	103594.000000	103594.000000	103594.000000
mean	3.304323	3.640761	3.286397	14.747939	15.178678	0.433394
std	1.265396	1.175603	1.312194	38.116737	38.698682	0.495546
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3.000000	3.000000	2.000000	0.000000	0.000000	0.000000
50%	3.000000	4.000000	3.000000	0.000000	0.000000	0.000000
75%	4.000000	5.000000	4.000000	12.000000	13.000000	1.000000
max	5.000000	5.000000	5.000000	1592.000000	1584.000000	1.000000

- Categorical Features
- Numerical Features



Description of Target Feature

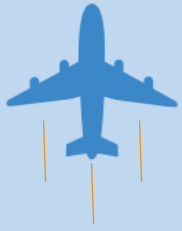


1: Satisfied attitude

0: Dissatisfied attitude

30% More

Fig. Satisfaction Level Description



General Trend of Dataset

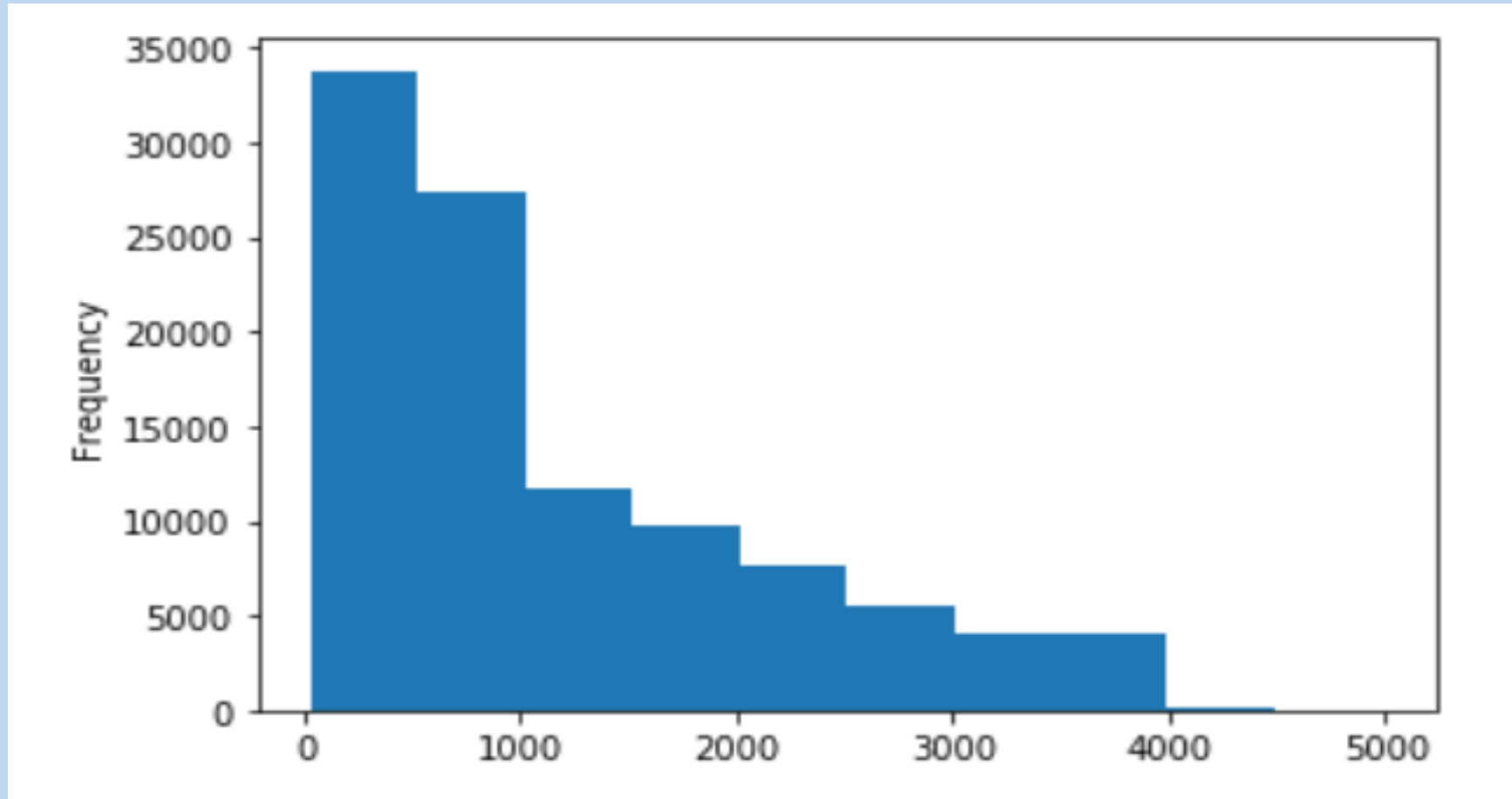
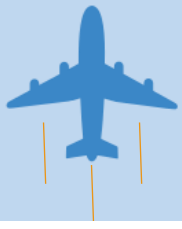


Fig. Flight Distance Frequency Distribution



General Trend of Dataset

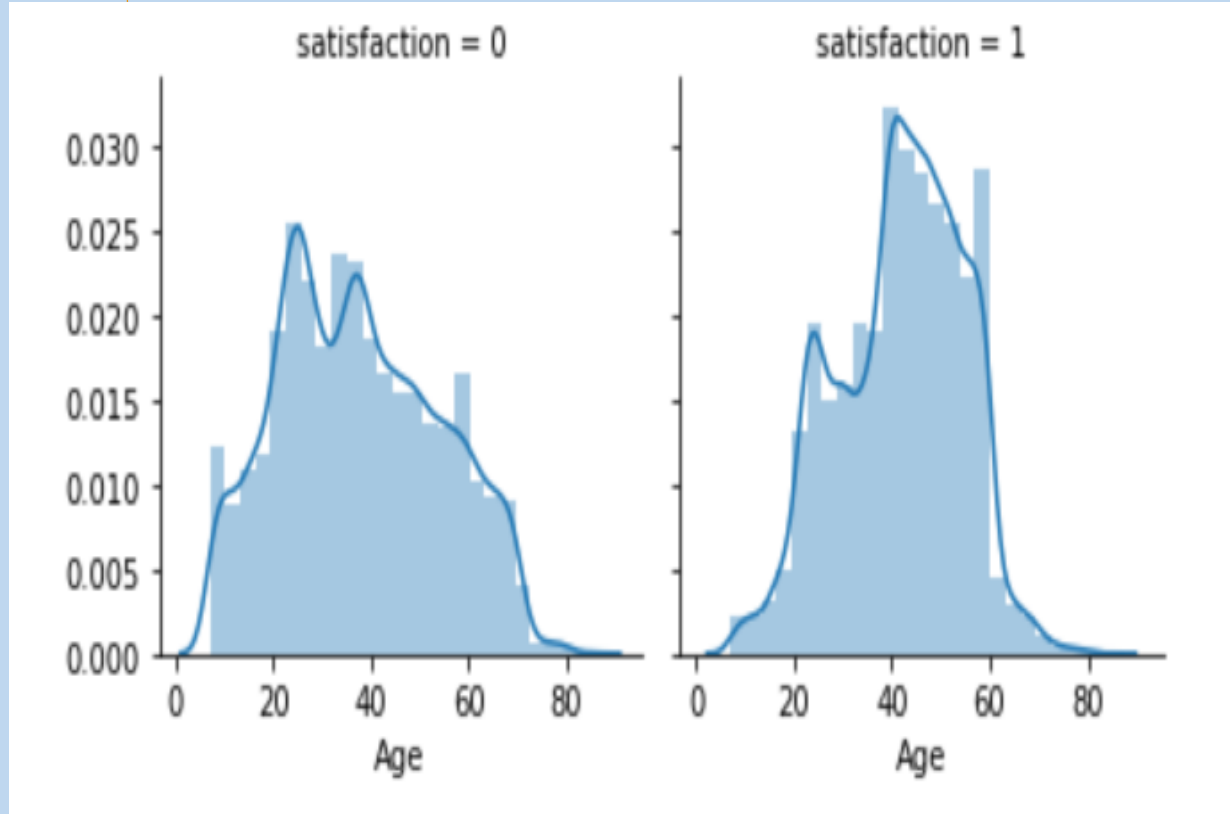


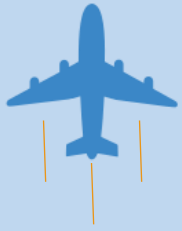
Fig. Age & Satisfaction Relation Tendency

```
# Age vs satisfaction  
train[['Age', 'satisfaction']].groupby(['Age'], as_index=False).mean().sort_values(by='satisfaction', ascending=False)
```

	Age	satisfaction
34	41	0.607755
39	46	0.599273
44	51	0.598287
50	57	0.590394
45	52	0.586089
...
2	9	0.137026
63	70	0.136808
3	10	0.130690
1	8	0.108320
0	7	0.100000

75 rows × 2 columns

Fig. Age & Satisfaction Correlation Graph



Count Plot Description

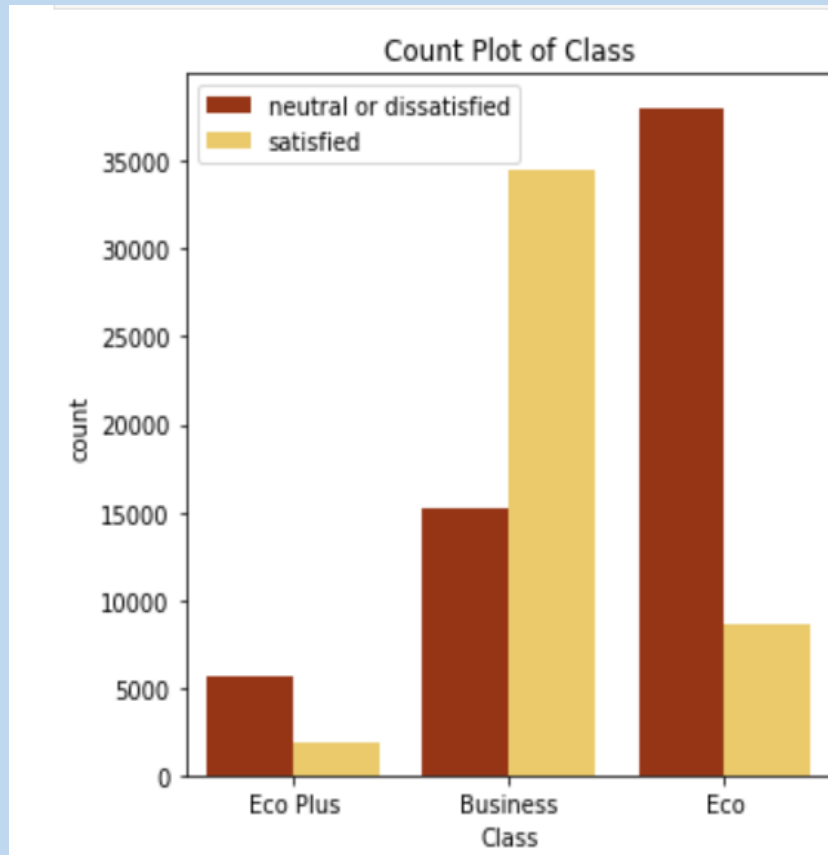


Fig. Count Plot of Flight Seat Class

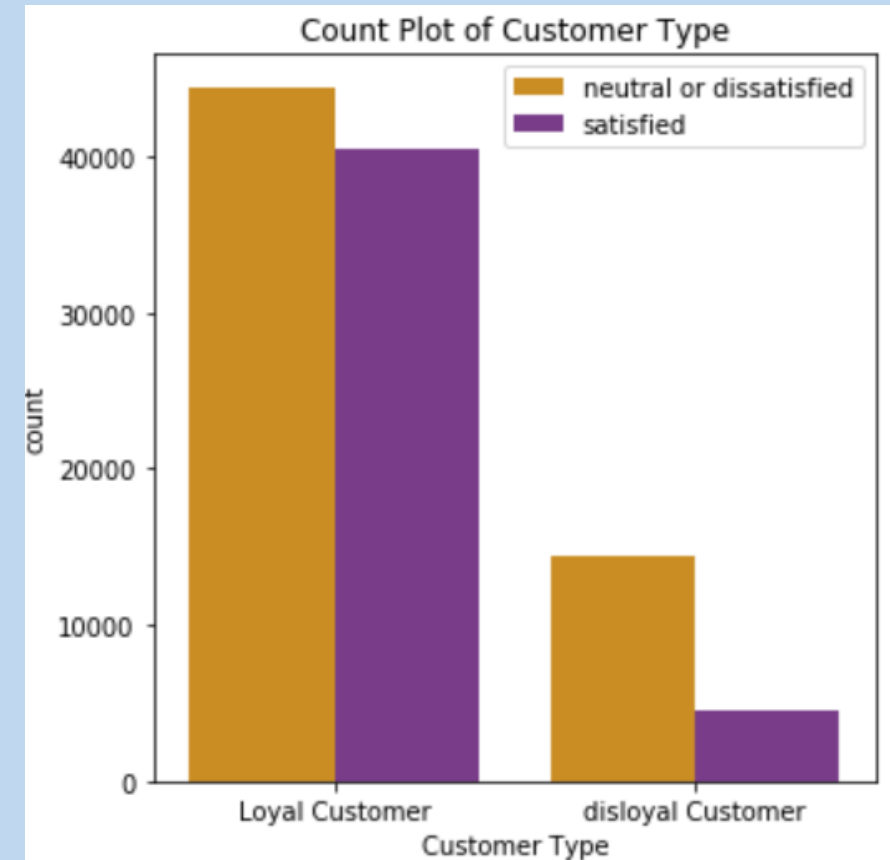
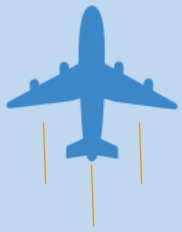


Fig. Count Plot of Customer Type



Count Plot Description

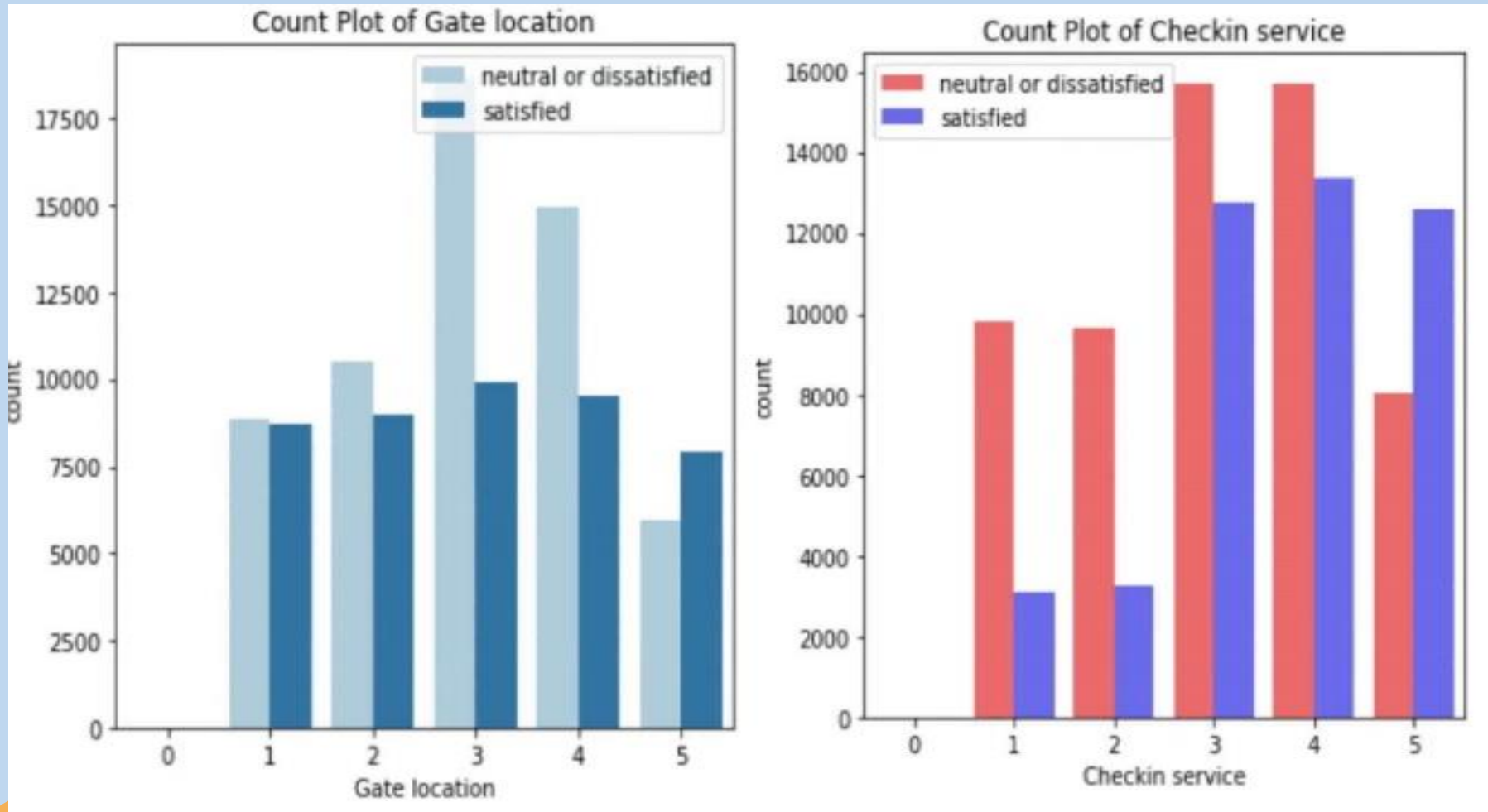
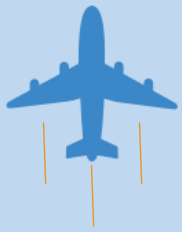


Fig. Count Plot of Gate location & Check in service



Correlation

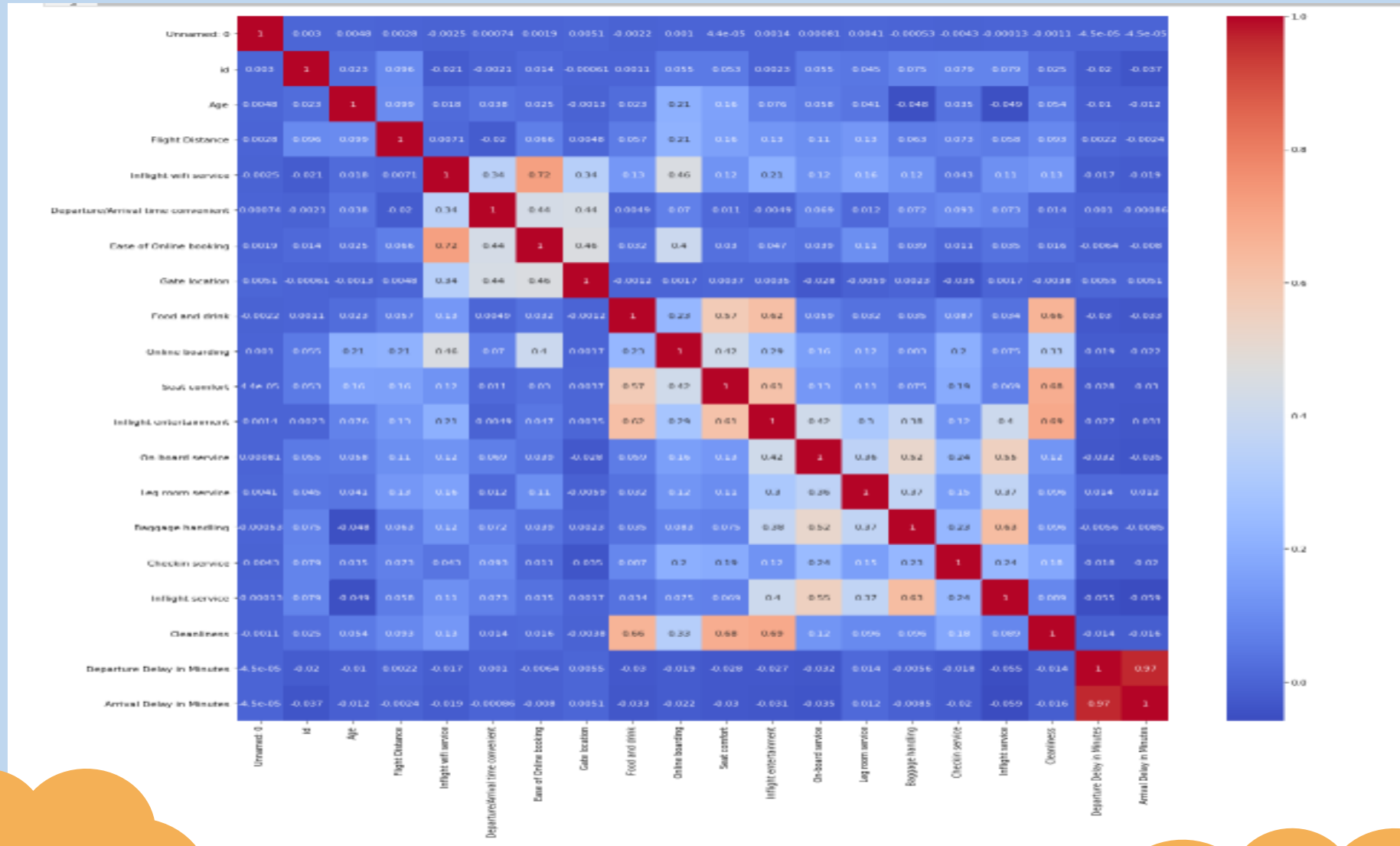
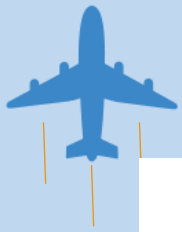


Fig. General Correlation Table



Correlation

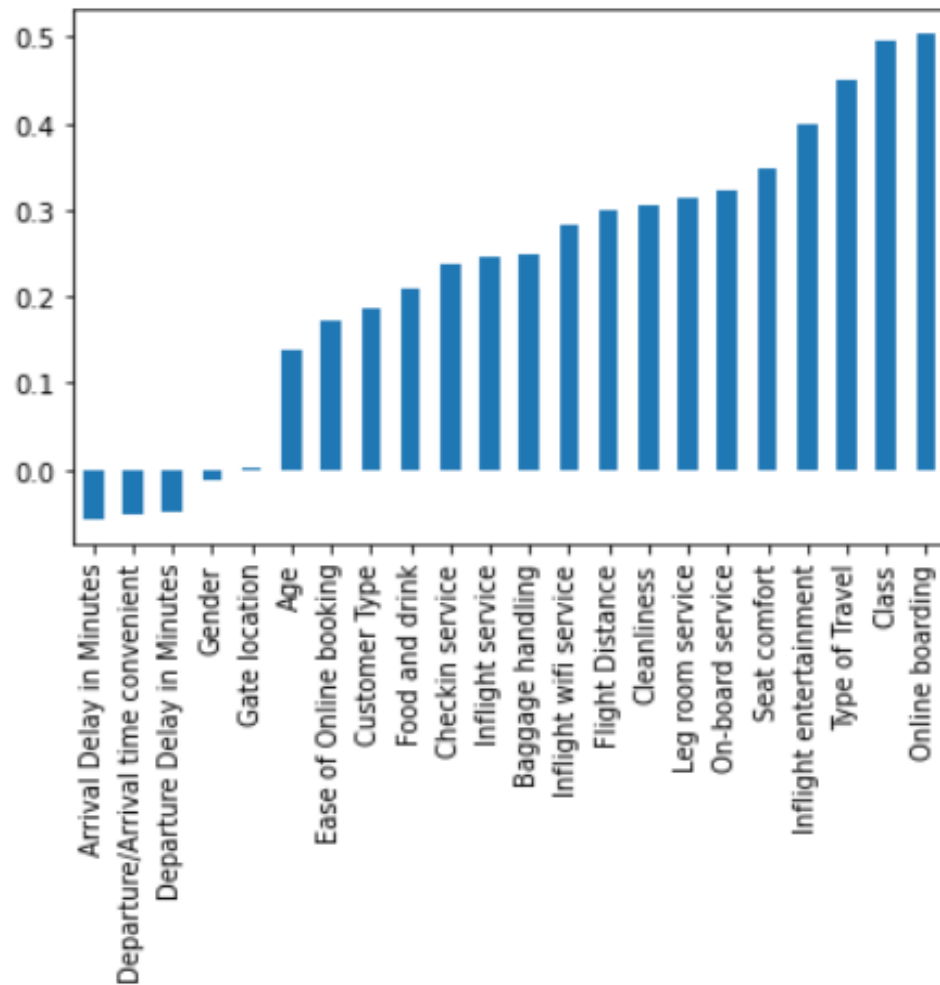


Fig. The Correlation Table between Satisfied Level
& Other Features

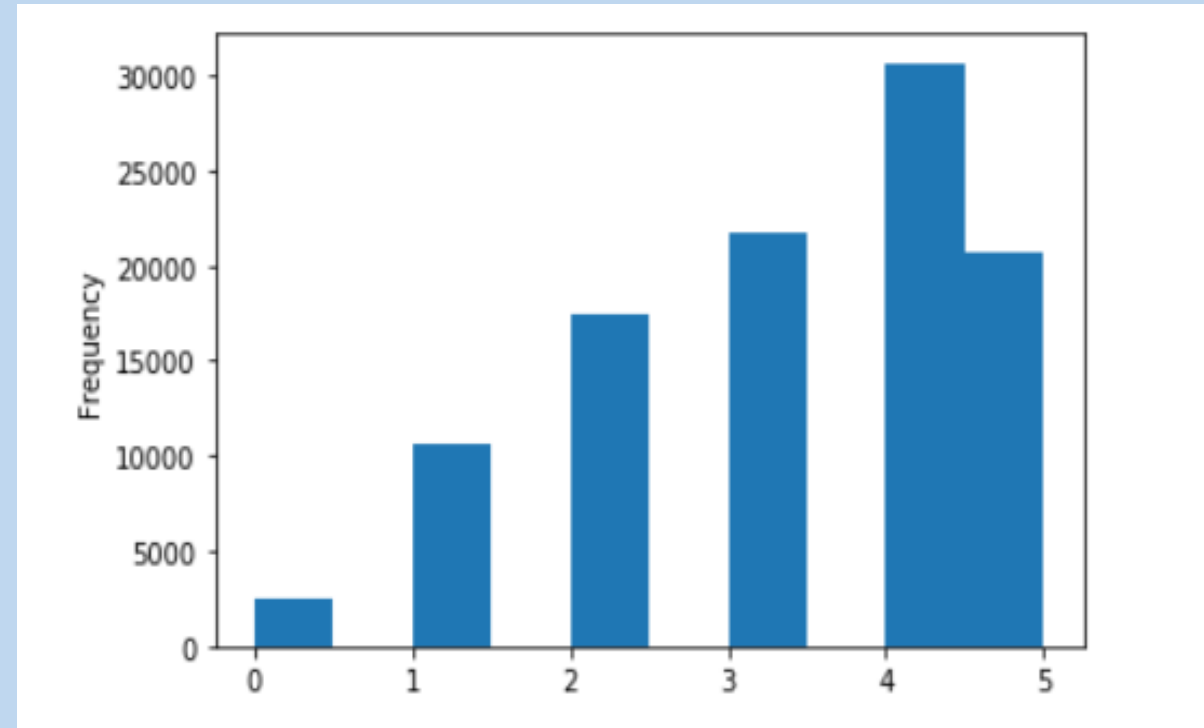
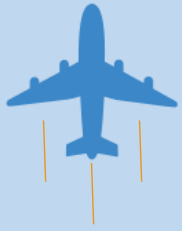


Fig. The Correlation Table between Online Boarding
& Other Variables



Correlation

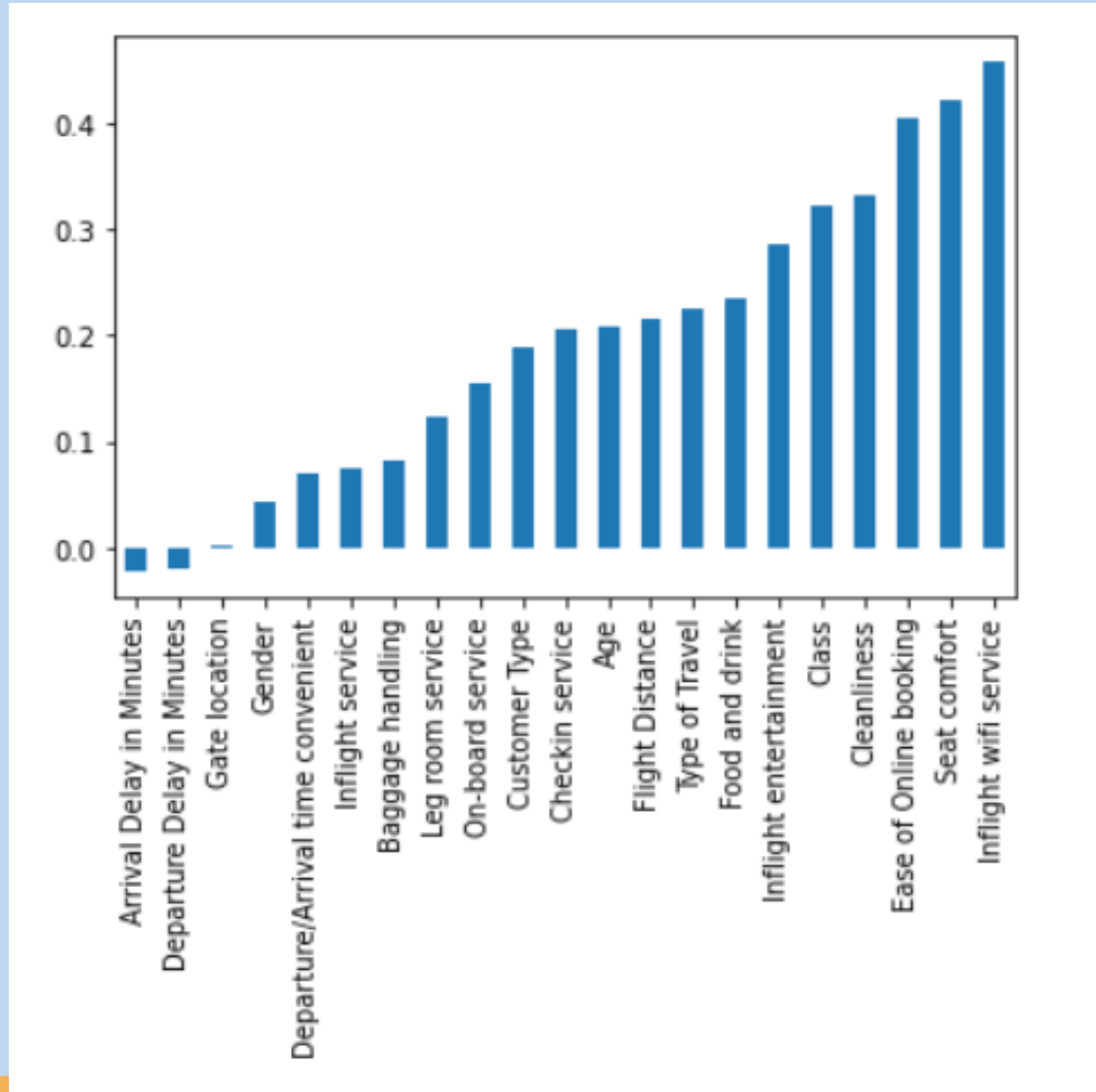


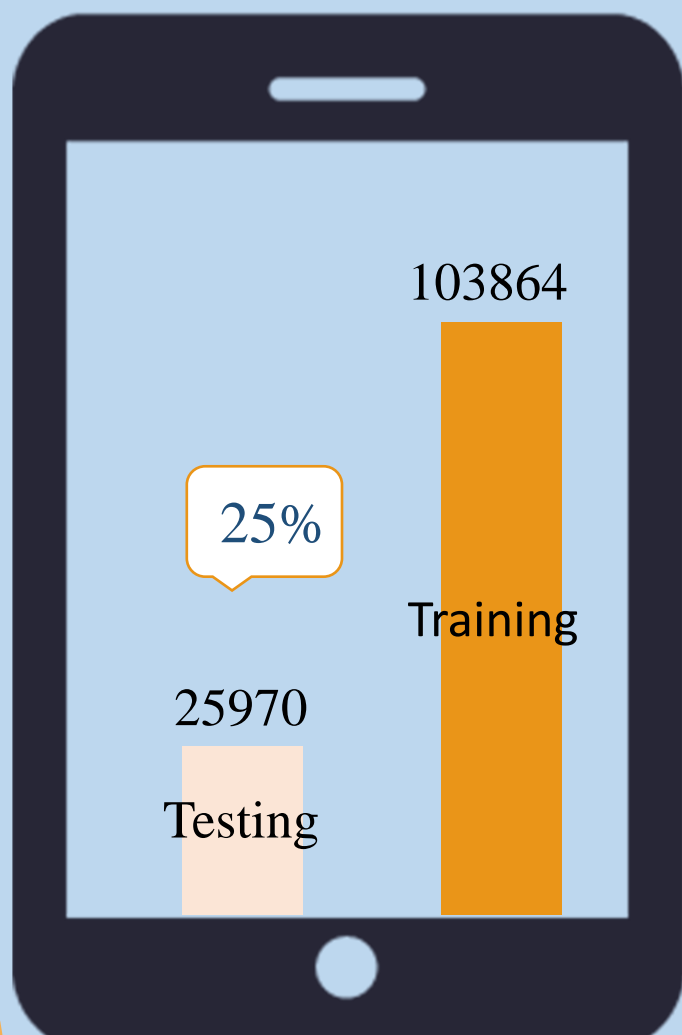
Fig. The Correlation Table between Food and Drink service & Passenger Satisfaction

04 Data processing





Train/Test Split



```
In [11]: train.head(3).append(train.tail(3))
```

Out[11]:

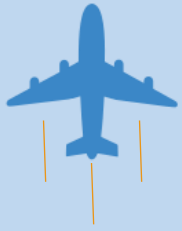
	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient
0	0	1	13	0	1	460	3	4
1	0	0	25	1	2	235	3	2
2	1	1	26	1	2	1142	2	2
103861	0	0	30	1	2	1995	1	1
103862	1	0	22	1	0	1000	1	1
103863	0	1	27	1	2	1723	1	3

6 rows x 23 columns

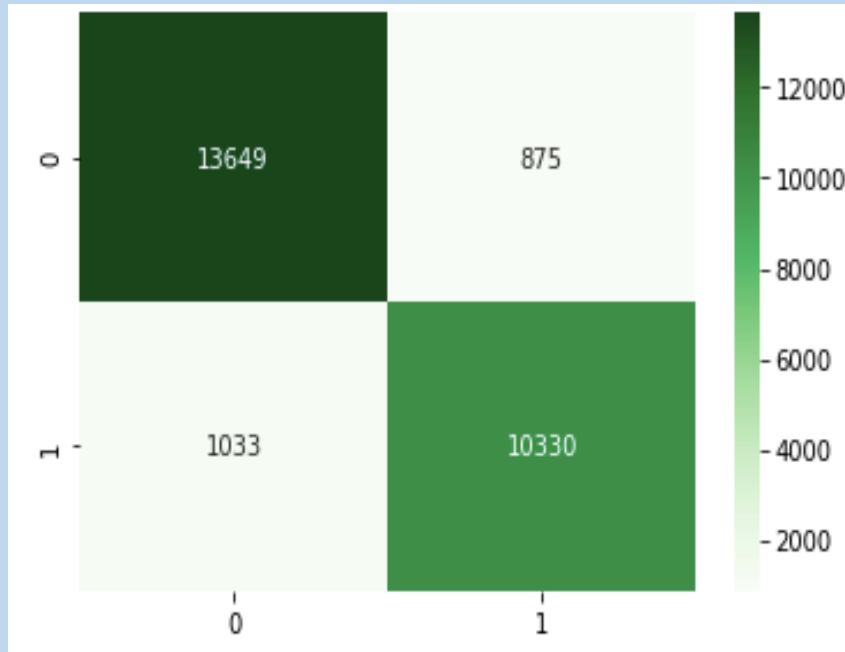
```
In [12]: test.head(3).append(test.tail(3))
```

Out[12]:

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient
0	1	1	52	1	0	160	5	4
1	1	1	36	1	2	2863	1	1
2	0	0	20	1	0	192	2	0
25967	1	1	17	0	0	828	2	5
25968	0	1	14	1	2	1127	3	3
25969	1	1	42	0	0	264	2	5



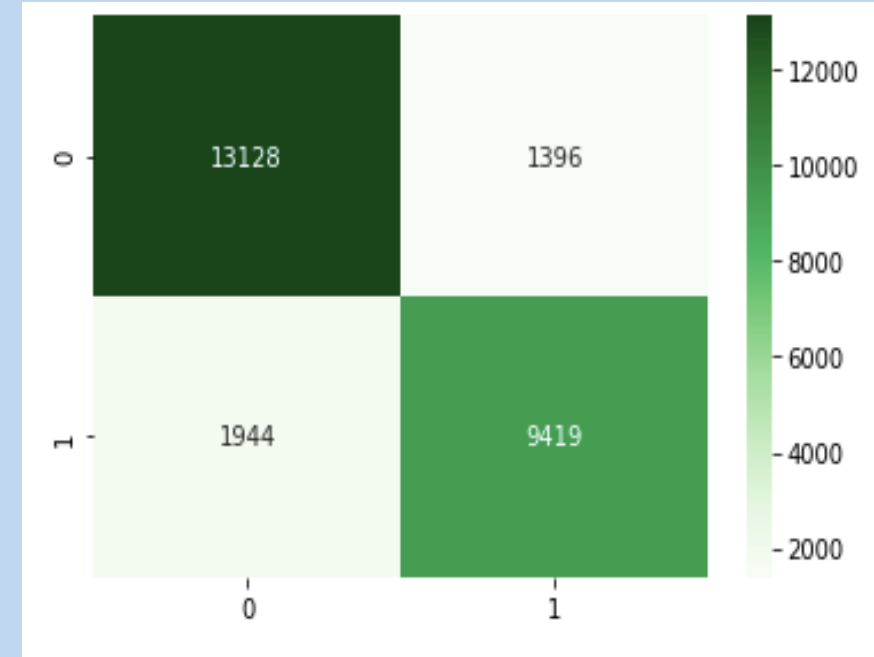
Model Evaluation



Confusion Matrix for Logistic Regression

Accuracy score : 0.8715571522385753
Precision score : 0.8702100221075902
Recall score : 0.8313825574232157
The Mean Square Error is 0.13

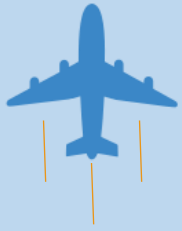
Evaluation Score for Logistic Regression



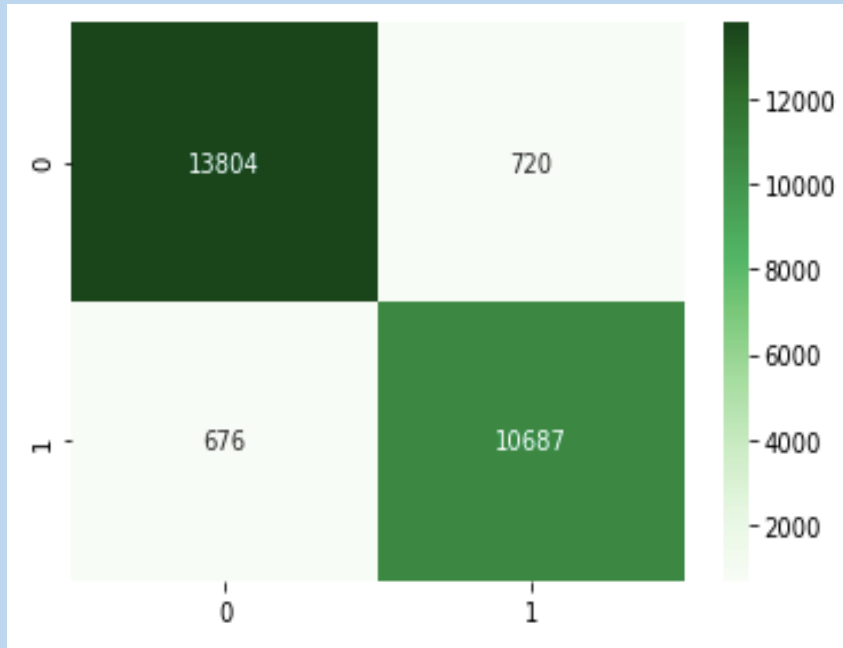
Confusion Matrix for Linear SVM

Accuracy score : 0.8709777108201028
Precision score : 0.8709200184928341
Recall score : 0.8289184194314881
The Mean Square Error is 0.13

Evaluation Score for Linear SVM



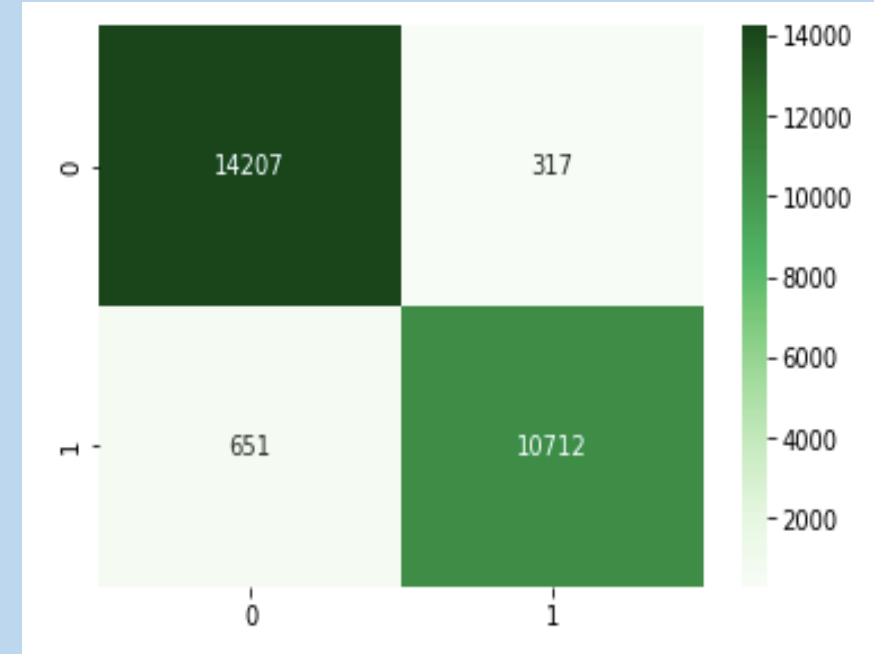
Model Evaluation



Confusion Matrix for Decision Tree

Accuracy score : 0.9460733186541508
Precision score : 0.9368808626282108
Recall score : 0.9405086684854352
The Mean Square Error is 0.05

Evaluation Score for Decision Tree



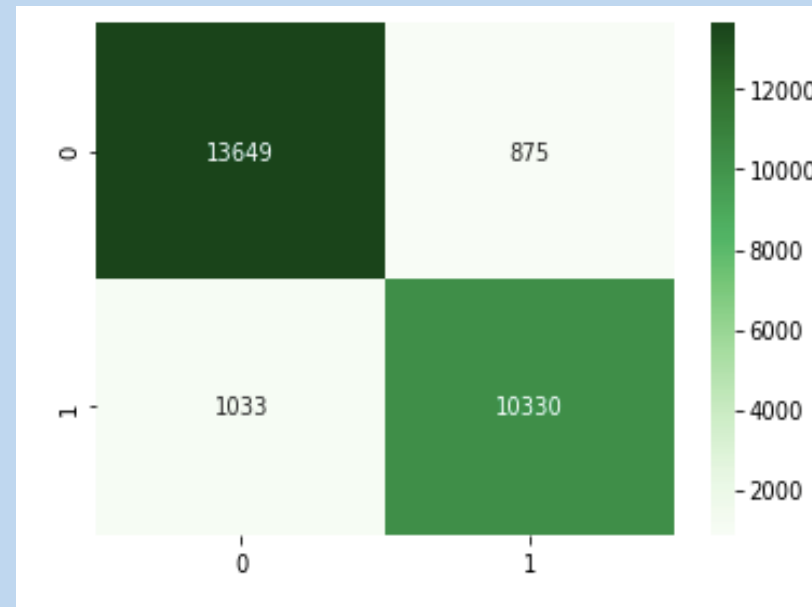
Confusion Matrix for Random Forest

Accuracy score : 0.9626067137945687
Precision score : 0.9712575936168284
Recall score : 0.9427087916923348
The Mean Square Error is 0.04

Evaluation Score for Random Forest



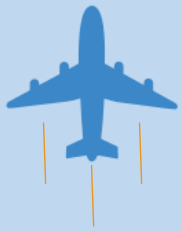
Model Evaluation



Confusion Matrix for Ada-Boost

Accuracy score : 0.9262950515702862
Precision score : 0.9219098616688978
Recall score : 0.9090909090909091
The Mean Square Error is 0.07

Evaluation Score for Ada-Boost



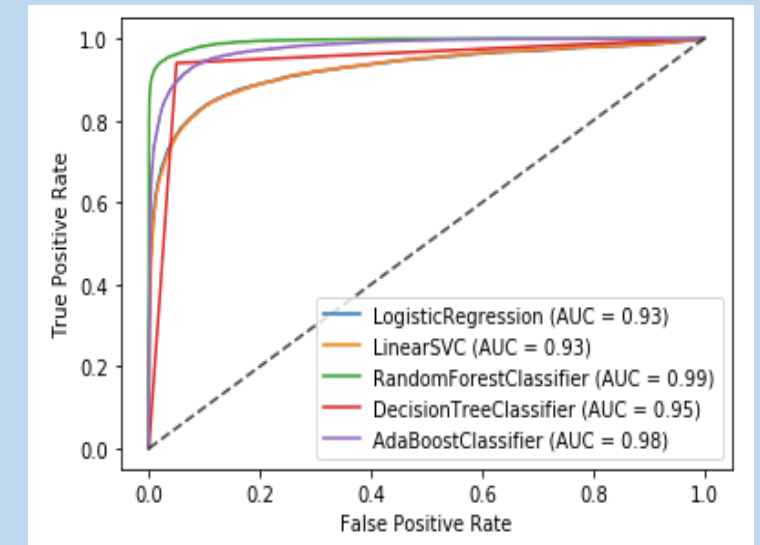
Comparison & Model Choosing

	AUC
Logistic Regression	0.871557
Linear SVM	0.870978
RandomForest	0.962607
Decision Tree	0.946073
AdaBoost	0.926295

Accuracy for five classifiers

	CPU times	Wall time
Logistic Regression	482ms	275ms
Linear SVM	31.8s	31.8s
Decision Tree	732ms	733ms
Random Forest	14.2s	14.4s
Ada-Boost	5.63s	5.79s

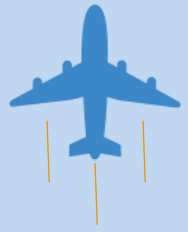
Running time for five classifier



ROC-Curve for five classifiers

05 Limitation & Conclusion



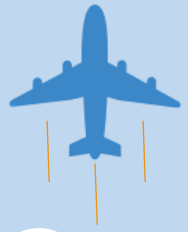


Limitation



Limitation

There are too many variables so that some classifier models such as logistic model and decision trees, cannot visualize the relationships of variables.

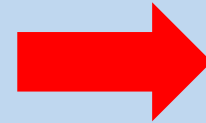


Conclusion

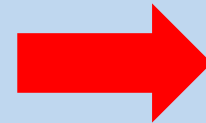


Conclusion

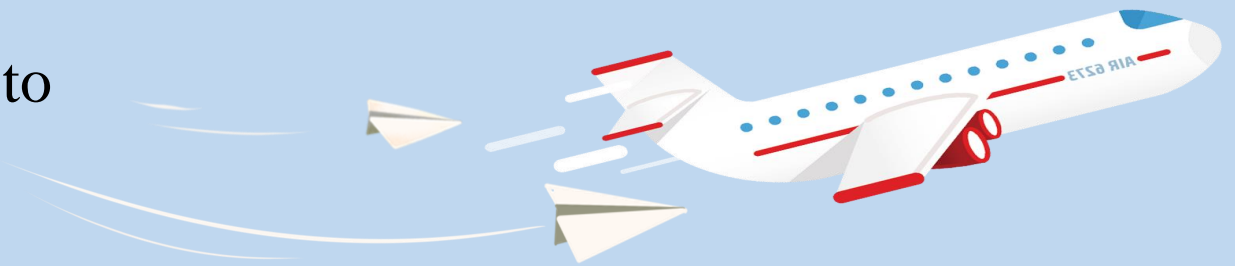
- Online booking service, seat class, and type of travel have a strong positive correlation with customers' satisfaction attitude
- Gate location, gender, departure or arrival time convenience, and arrival and departure delay have a negative correlation with customers' satisfaction attitude
- Random forest is the appropriate classifier to analyze flight passengers' satisfaction.

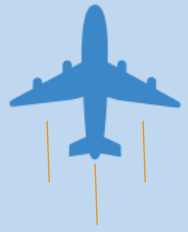


Airlines can improve customers' satisfaction by improving online booking service.



Airlines should try their best to reduce the situation of arrival and departure delay caused by human factors.





Reference

Asyiqin, S., 2020. Singapore Airlines Introduces No-Destination Flights For Those Who Miss Flying.

[online] WORLD OF BUZZ. Available at: <<https://worldofbuzz.com/singapore-airlines-introduces-no-destination-flights-for-those-who-miss-flying/>> [Accessed 12 November 2020].

Qantas 7-hour flight to nowhere sells out in 10 minutes. (2020). Retrieved 12 November 2020, from <https://abc30.com/flight-to-nowhere-qantas-airlines-flights-with-no-destination-flying-in-covid-19-pandemic/6454022/>

THANK YOU

