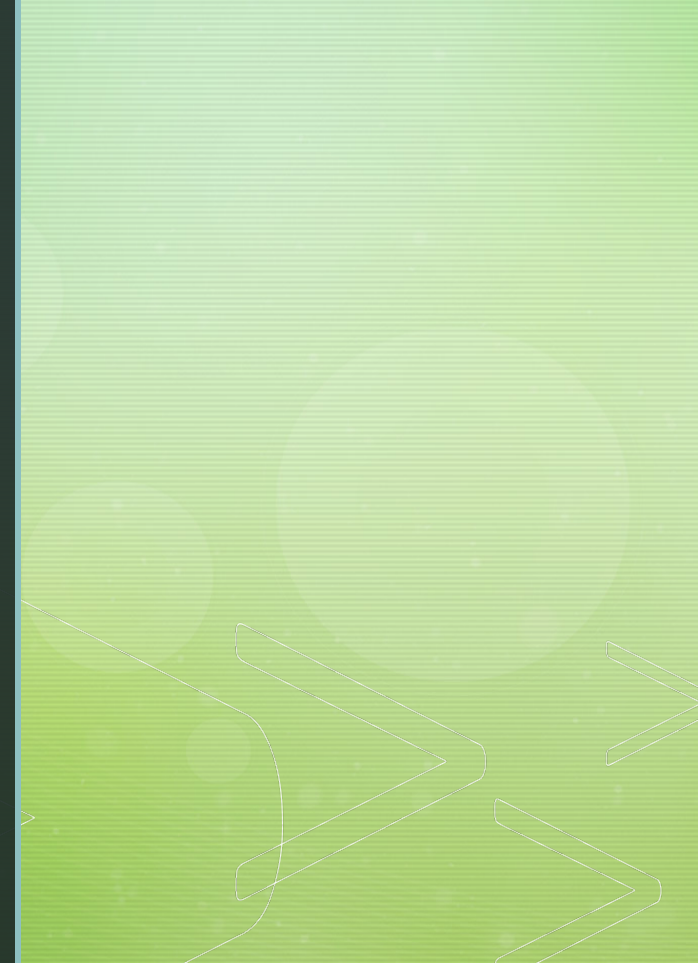


Airlines Customer Satisfaction and Business Planning



By Mujiati Nurokhmah



Database / Dataset

- <https://www.kaggle.com/datasets/sjleshrac/airlines-customer-satisfaction>

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129880 entries, 0 to 129879
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   satisfaction                             129880 non-null  object
1   Gender                                   129880 non-null  object
2   Customer Type                           129880 non-null  object
3   Age                                       129880 non-null  int64
4   Type of Travel                          129880 non-null  object
5   Class                                    129880 non-null  object
6   Flight Distance                         129880 non-null  int64
7   Seat comfort                            129880 non-null  int64
8   Departure/Arrival time convenient       129880 non-null  int64
9   Food and drink                          129880 non-null  int64
10  Gate location                           129880 non-null  int64
11  Inflight wifi service                   129880 non-null  int64
12  Inflight entertainment                  129880 non-null  int64
13  Online support                          129880 non-null  int64
14  Ease of Online booking                  129880 non-null  int64
15  On-board service                       129880 non-null  int64
16  Leg room service                       129880 non-null  int64
17  Baggage handling                       129880 non-null  int64
18  Checkin service                        129880 non-null  int64
19  Cleanliness                            129880 non-null  int64
20  Online boarding                        129880 non-null  int64
21  Departure Delay in Minutes              129880 non-null  int64
22  Arrival Delay in Minutes                129487 non-null  float64
dtypes: float64(1), int64(17), object(5)
memory usage: 22.8+ MB
```

Missing Value dan Duplicated Data

```
Gender 0
Customer Type 0
Age 0
Type of Travel 0
Class 0
Flight Distance 0
Seat comfort 0
Departure/Arrival time convenient 0
Food and drink 0
Gate location 0
Inflight wifi service 0
Inflight entertainment 0
Online support 0
Ease of Online booking 0
On-board service 0
Leg room service 0
Baggage handling 0
Checkin service 0
Cleanliness 0
Online boarding 0
Departure Delay in Minutes 0
Arrival Delay in Minutes 393
satisfaction 0
dtype: int64
```

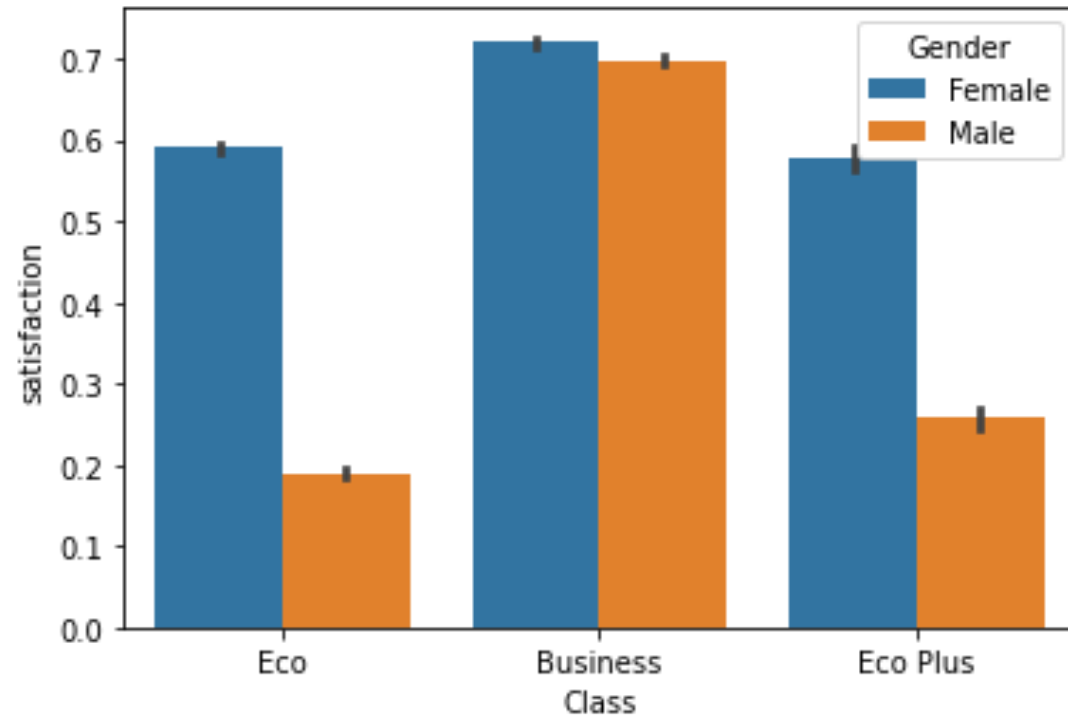
Duplicated Data

```
[17] # jumlah data duplicated
      df.duplicated().sum()
```

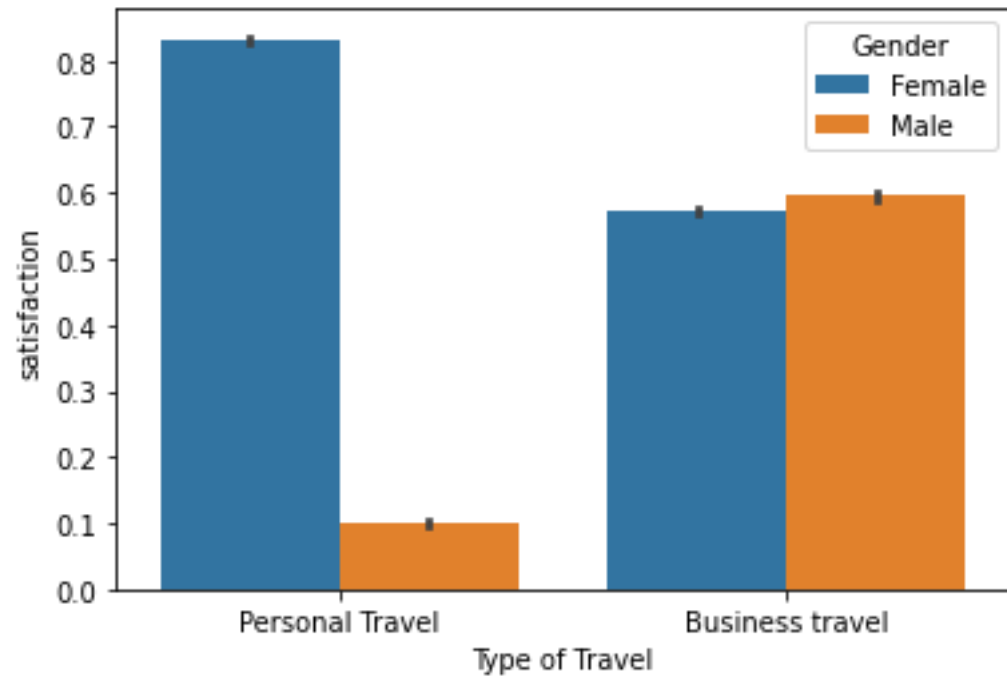
```
0
```

Visualisasi Data

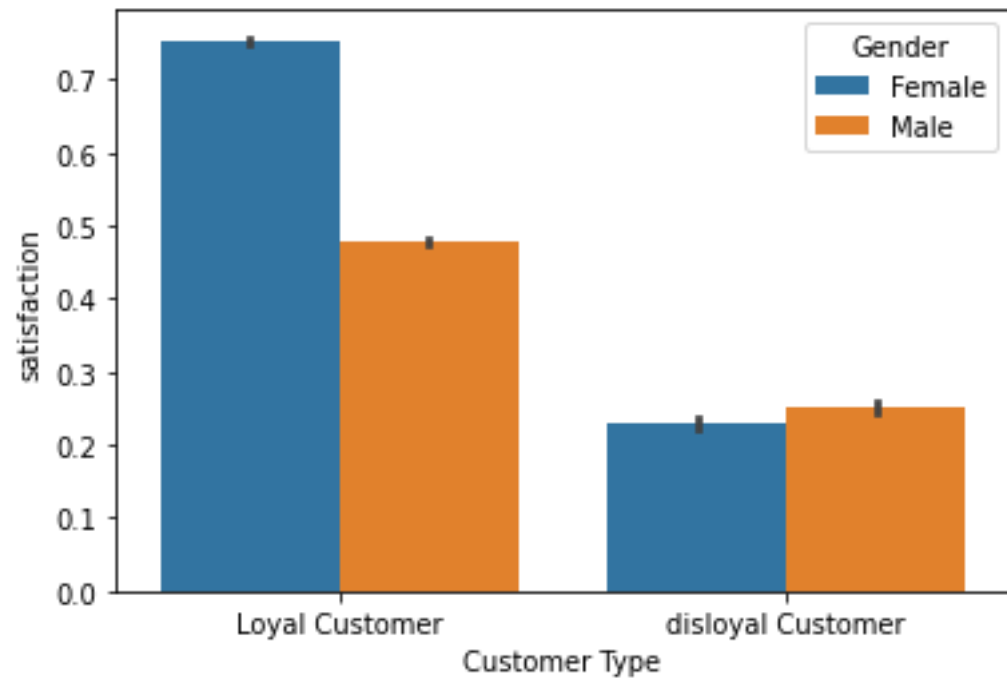
- Perbandingan satisfaction, class dan gender



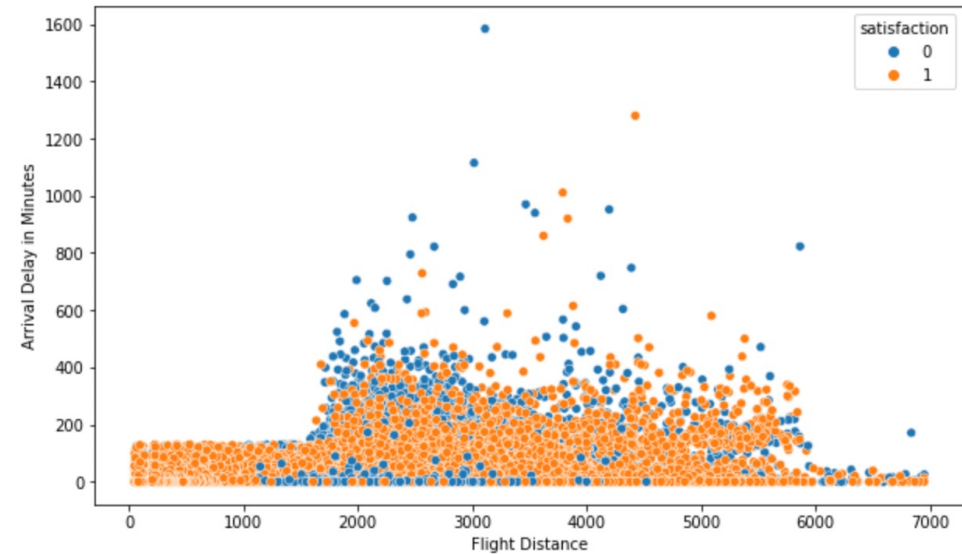
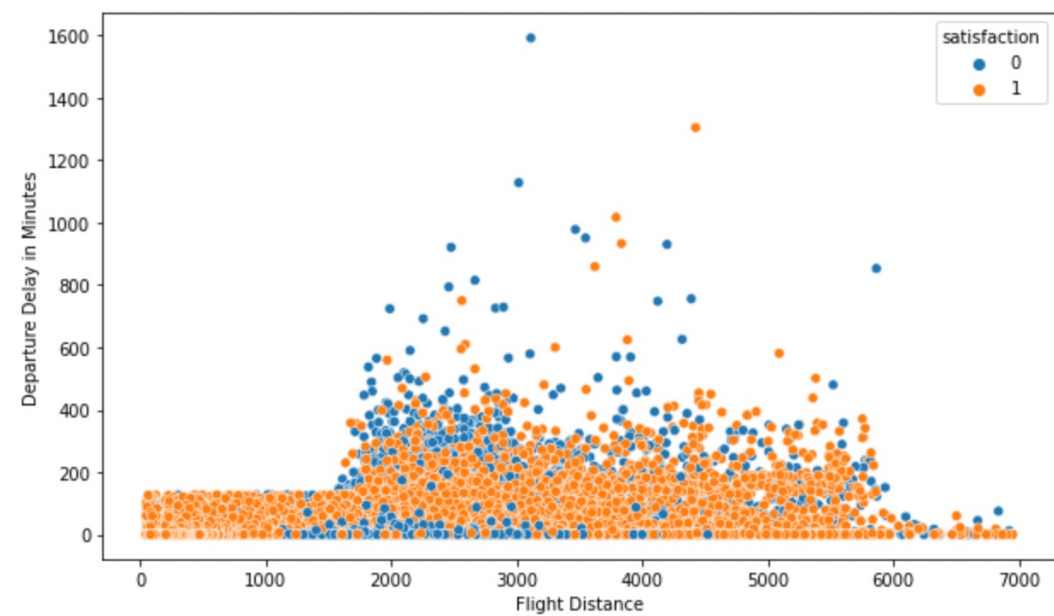
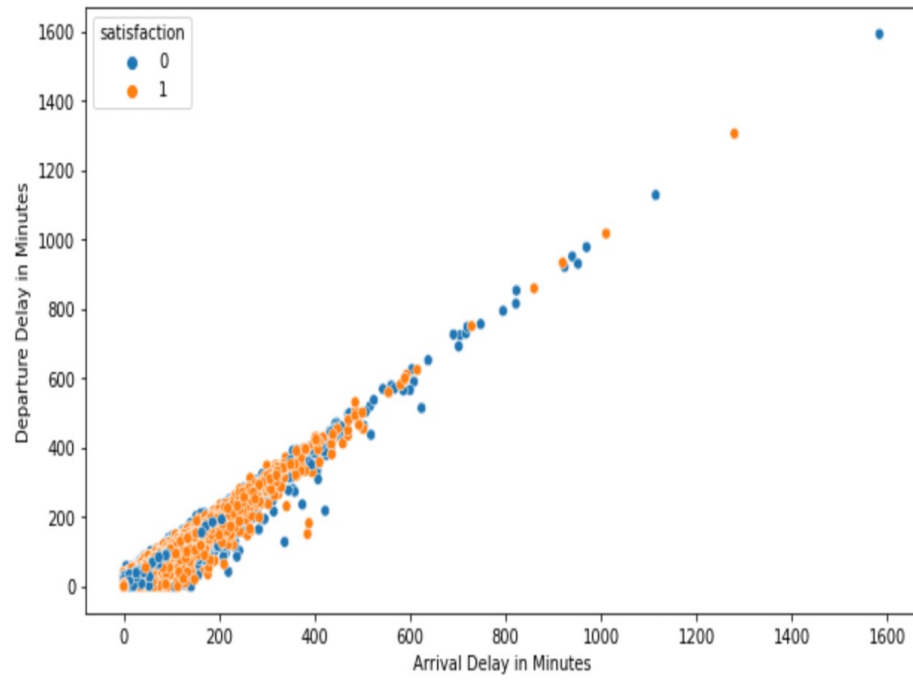
Perbandingan variabel satisfaction,
Type of Travel dan Gender



Perbandingan variabel satisfaction,
Customer Type dan Gender

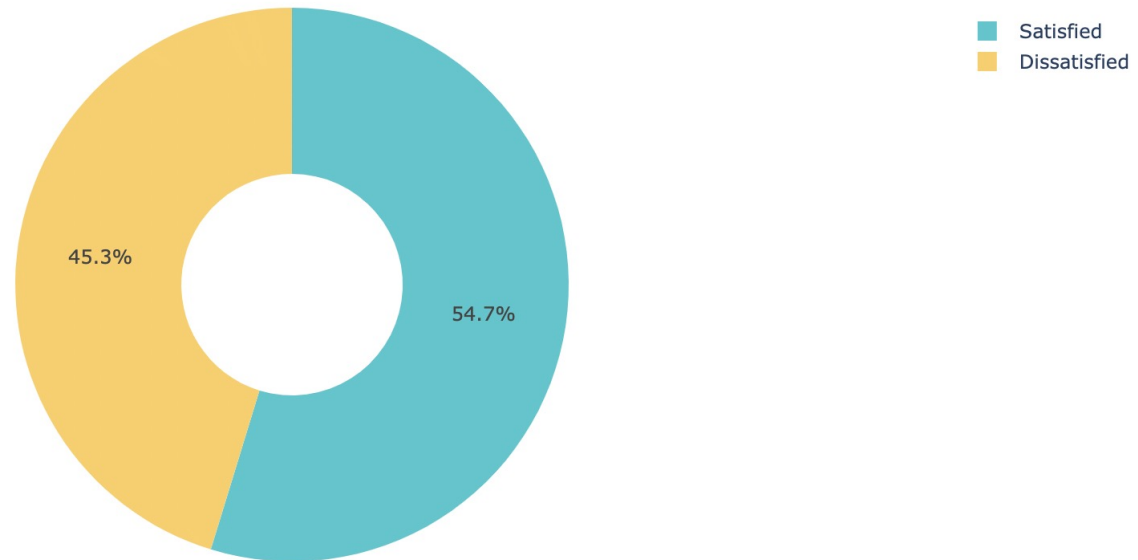


Perbandingan Variabel Delay Lengths (Arrival Dan Departure Delay in Minutes), Flight Distance

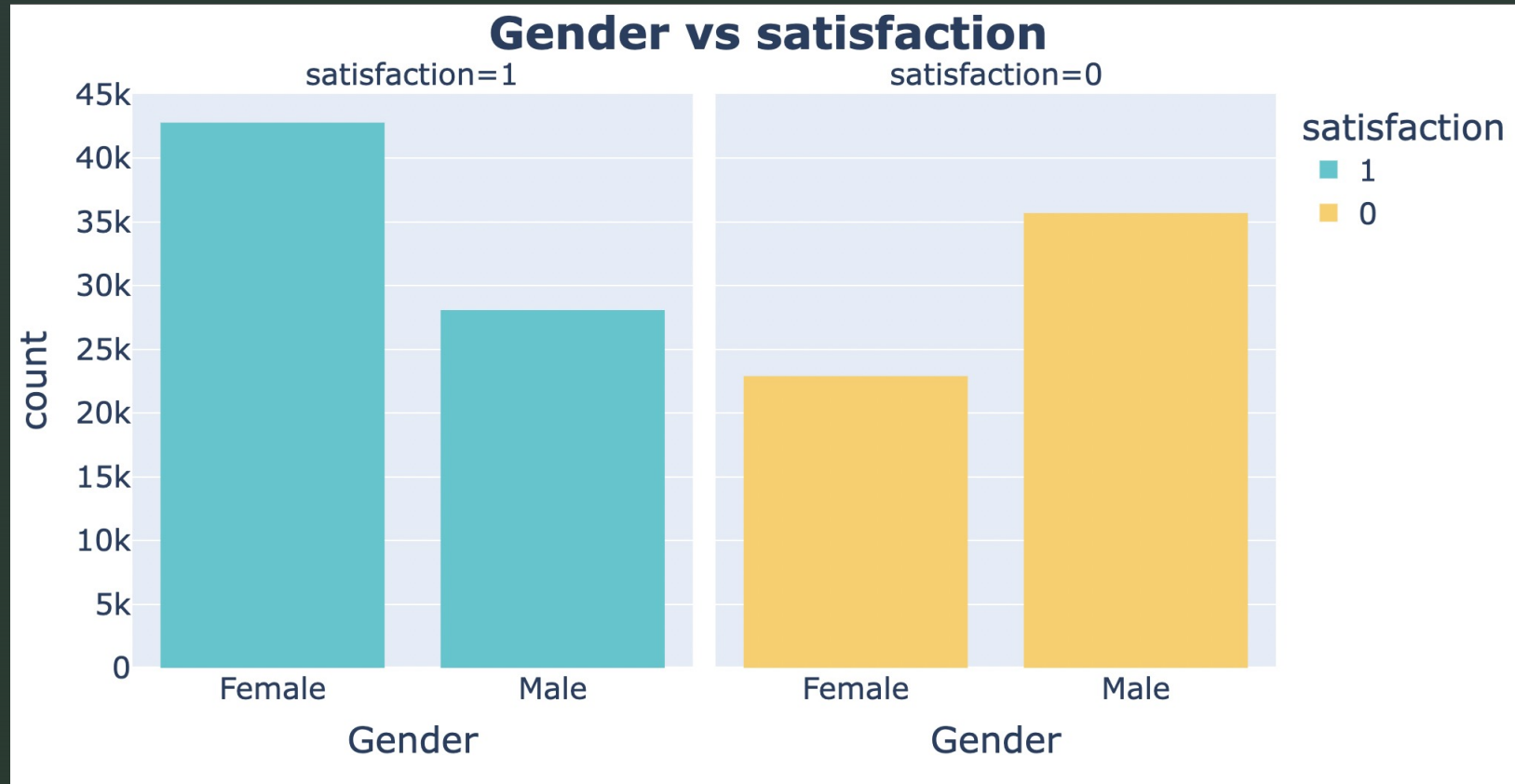


Ratio orang yg satisfied dan dissatisfied

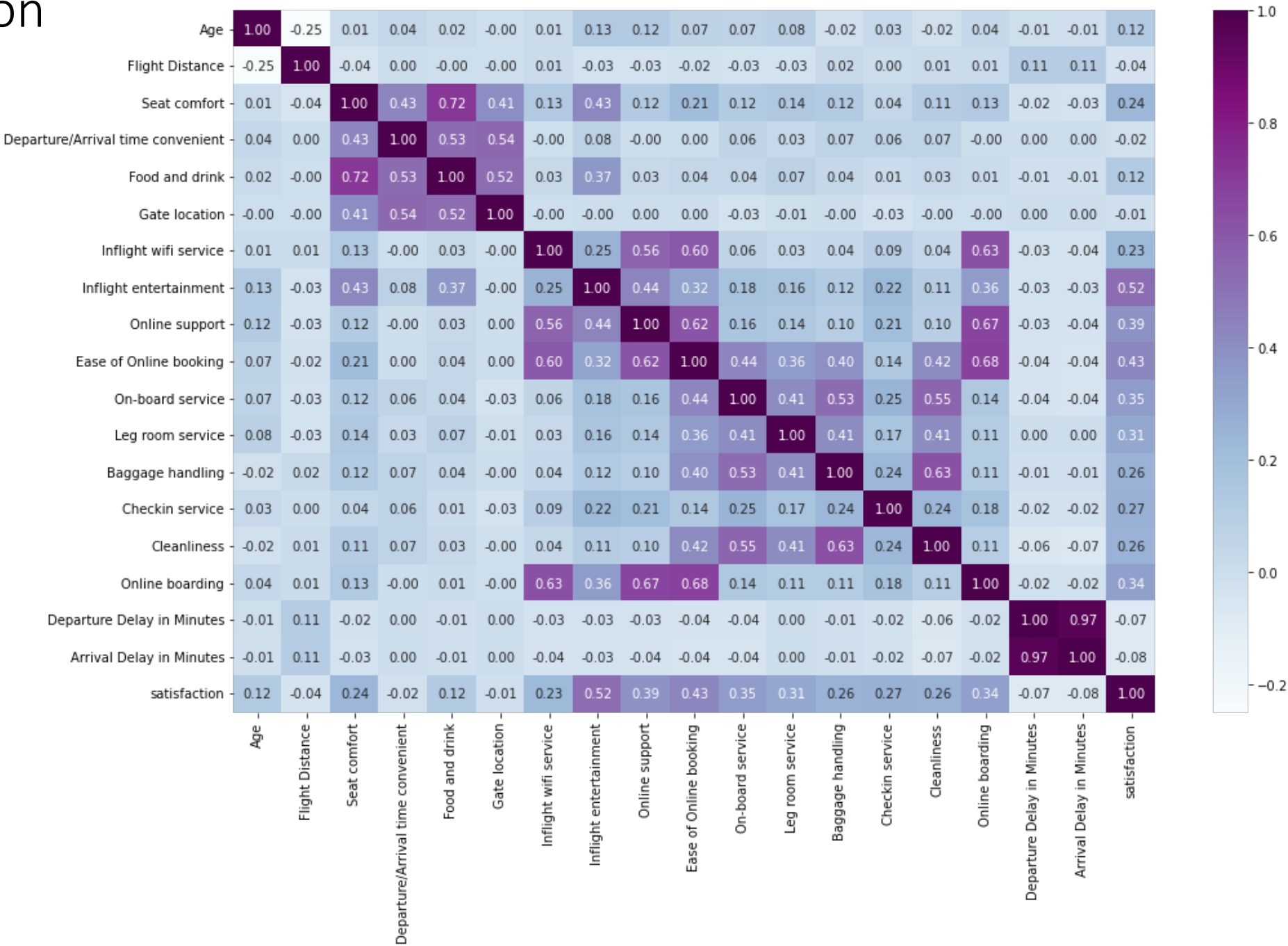
Ratio Satisfied dan Dissatisfied



Rasio Female dan Male yg satisfied dan dissatisfied



Heatmap Correlation



Modelling – Random Forest

Accuracy: 0.9601198566662548

Confusion Matrix:

	precision	recall	f1-score	support
0	0.95	0.97	0.96	14732
1	0.97	0.95	0.96	17640
accuracy			0.96	32372
macro avg	0.96	0.96	0.96	32372
weighted avg	0.96	0.96	0.96	32372

Assuming 1 as positive, we get:

True Positive: 16813

True Negative: 14268

False Positive: 464

False Negative: 827

Precision: 0.9731434855588355

Recall: 0.9531179138321996

F1-Score: 0.9630266059512559

Modelling – Random Forest

hasil analisis:

1. True Positif = 16813 yg berarti diprediksi puas/satisfied dan benar satisfied sebanyak 16813.
2. True Negative = 14268 yg berarti diprediksi tidak puas ternyata tidak puas sebanyak 14268.
3. False Positive = 464 yg berarti diprediksi puas ternyata tidak puas sebanyak 464.
4. False Negative = 827 yg berarti diprediksi tidak puas dan ternyata tidak puas sebanyak 827.

kita juga temukan:

1. Accuracy = 96%. yg berarti perbandingan TP dan TN terhadap semua data tingkat akurasi sebanyak 96%.
2. Precision = 97% yg berarti diantara semua customer yg diprediksi puas ternyata yg sesungguhnya puas sebanyak 97%.
3. Recall = 95% yg berarti diantara semua customer yg diprediksi puas yg berhasil diprediksi puas sebanyak 95%.

Kita ingin memprediksi kepuasan/satisfied customer untuk rencana bisnis kedepannya. Dari hasil diatas ingin menghilangkan False Positive, yaitu kita prediksi satisfied ternyata dissatisfied.

Insight dan Recommendation

Insight

- untuk semua jenis class (Eco, Business, Eco Pluss) Female memiliki tingkat satisfaction lebih tinggi dari Male. tapi pada class business perbedaan Satisfaction female dan Male hanya selisih sedikit saja.
- untuk Type of Travel (Personal Travel dan Business Travel), pada Personal Travel Female memiliki tingkat satisfaction jauh lebih tinggi dari pada Male. dan untuk Business Travel Male memiliki tingkat satisfaction sedikit lebih tinggi daripada Female.
- untuk Customer Type (Loyal Customer) Female memiliki tingkat satisfaction lebih tinggi daripada Male. dan untuk Disloyal Customer Male memiliki tingkat satisfaction lebih tinggi daripada Female.
- tingkat satisfaction sepertinya semakin menurun sebanding dengan lamanya Delay(Departure maupun Arrival)
- Female memiliki jumlah satisfied lebih banyak dibanding Male, dan sebaliknya Male memiliki jumlah Dissatisfied lebih banyak

Recommendation

- Tingkatkan service untuk Male pada jenis Class Eco dan Eco Plus. misal promo yg menarik untuk tgl tertentu
- Tingkatkan service untuk Female pada jenis Class Business. misal promo yg menarik
- Tingkatkan service untuk Male pada Personal Travel, dan untuk Female pada Business Travel. misal promo yg menarik
- Tingkatkan pelayanan service secara keseluruhan untuk meminimalisir sekecil mungkin Delay baik departure maupun arrival.
- secara keseluruhan minimal harus mempertahankan satisfied Female dan harus meningkatkan service untuk Male supaya satisfied Male bisa lebih meningkat