PASSENGER SATISFACTION

Today, I'm going to present you passenger satisfaction prediction on their overall journey

Summary :
1) Context of the dataset
2) Cleaning the dataset
3) Tableau visualizations
4) SQL analyse
5) Results

1) Context of the dataset

- US Airline 2015 passengers satisfaction with their flight
- The number of rows and columns :
  - before cleaning: 103 904 rows and 25 columns
  - after cleaning: 75 119 rows and 24 columns

| id	gender	customer_type	age	type_of_travel	class	flight_distance	inflight_wifi_service	departure/arrival_time_convenient	ease_of_online_booking
5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	3
110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	2
24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	5
119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	3
111157	Female	Loyal Customer	26	Personal Travel	Eco	1180	3	4	2

gate_location	food_and_drink	online_boarding	seat_comfort	inflight_entertainment	on-board_service	leg_room_service	baggage_handling	checkin_service
3	1	3	1	1	1	5	3	1
2	5	5	5	5	4	3	4	4
5	2	2	2	2	2	5	3	1
3	4	5	5	3	3	4	4	3
1	1	2	1	1	3	4	4	4

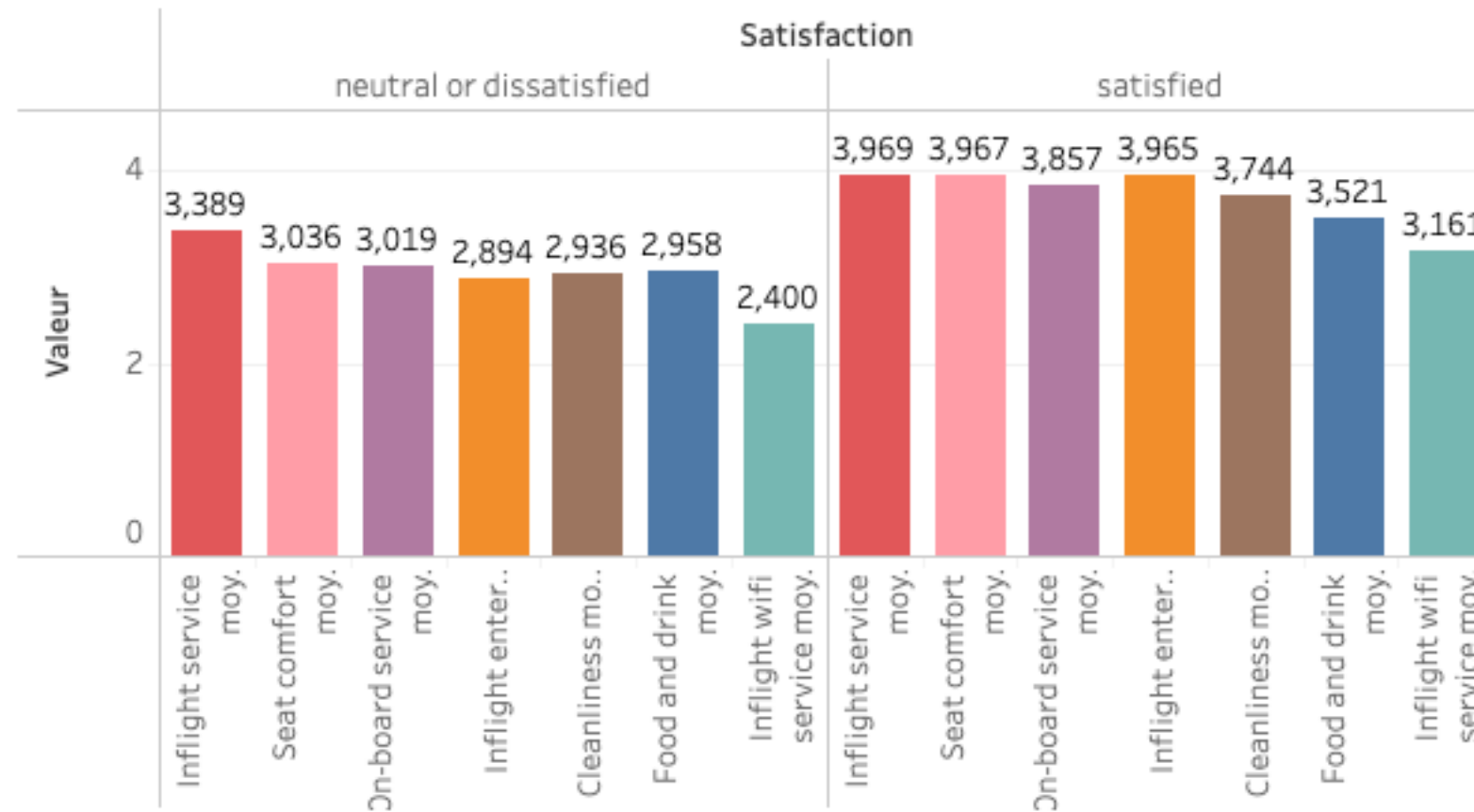| inflight_service	cleanliness	departure_delay_in_minutes	arrival_delay_in_minutes	satisfaction
4	1	1	6.0	neutral or dissatisfied
4	5	0	0.0	satisfied
4	2	11	9.0	neutral or dissatisfied
3	3	0	0.0	satisfied
4	1	0	0.0	neutral or dissatisfied

2) Cleaning the dataset

- Normalizing the columns' name
- Cleaning rows from 1 column : "arrival_delay_in_minutes "
- Checking for duplicates
- Removing outliers from the following columns :
  - flight_distance
  - departure_delay_in_minutes
  - arrival_delay_in_minutes
- Saving the cleaned dataset to a new CSV file
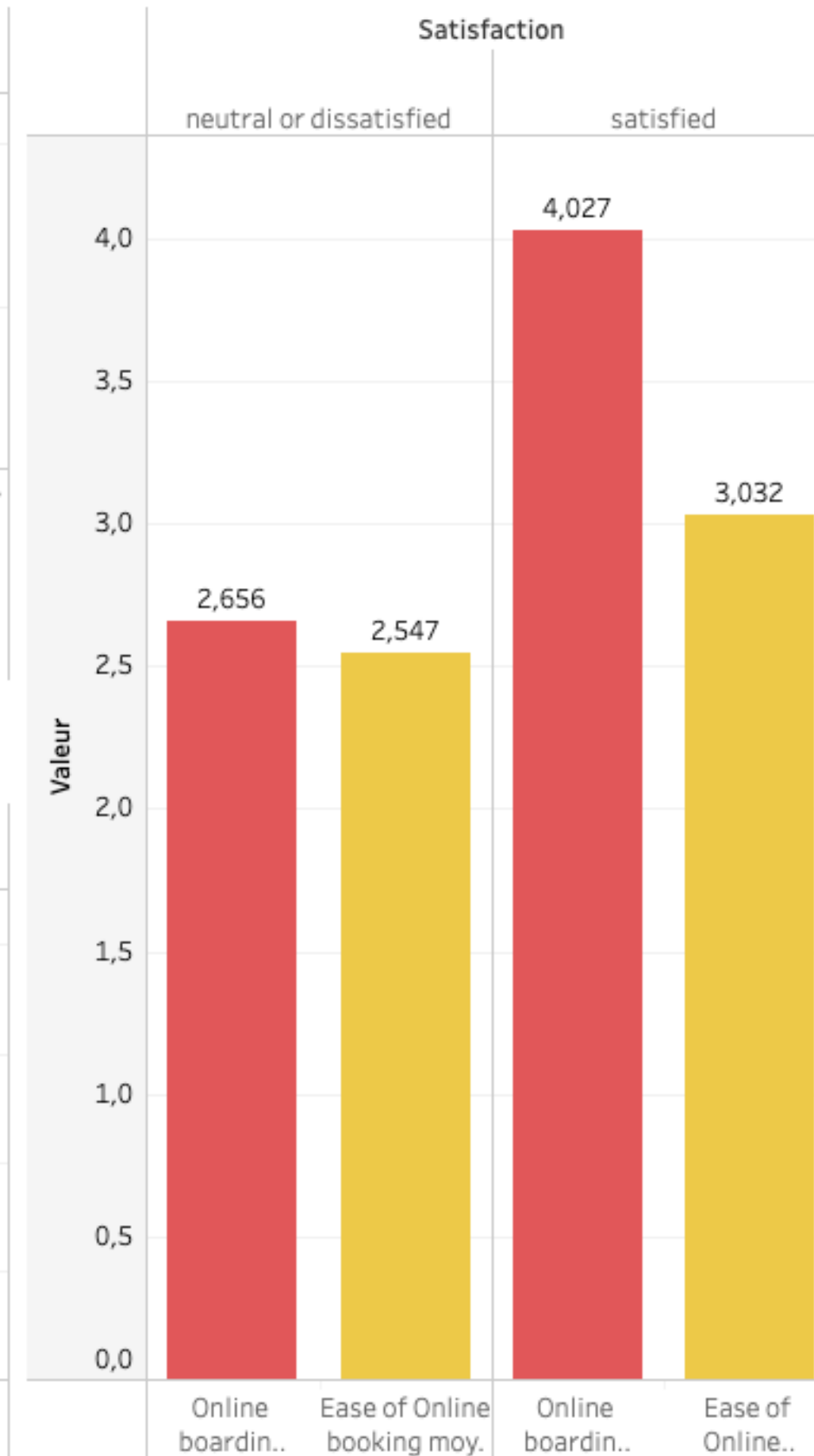
3) Tableau visualizations
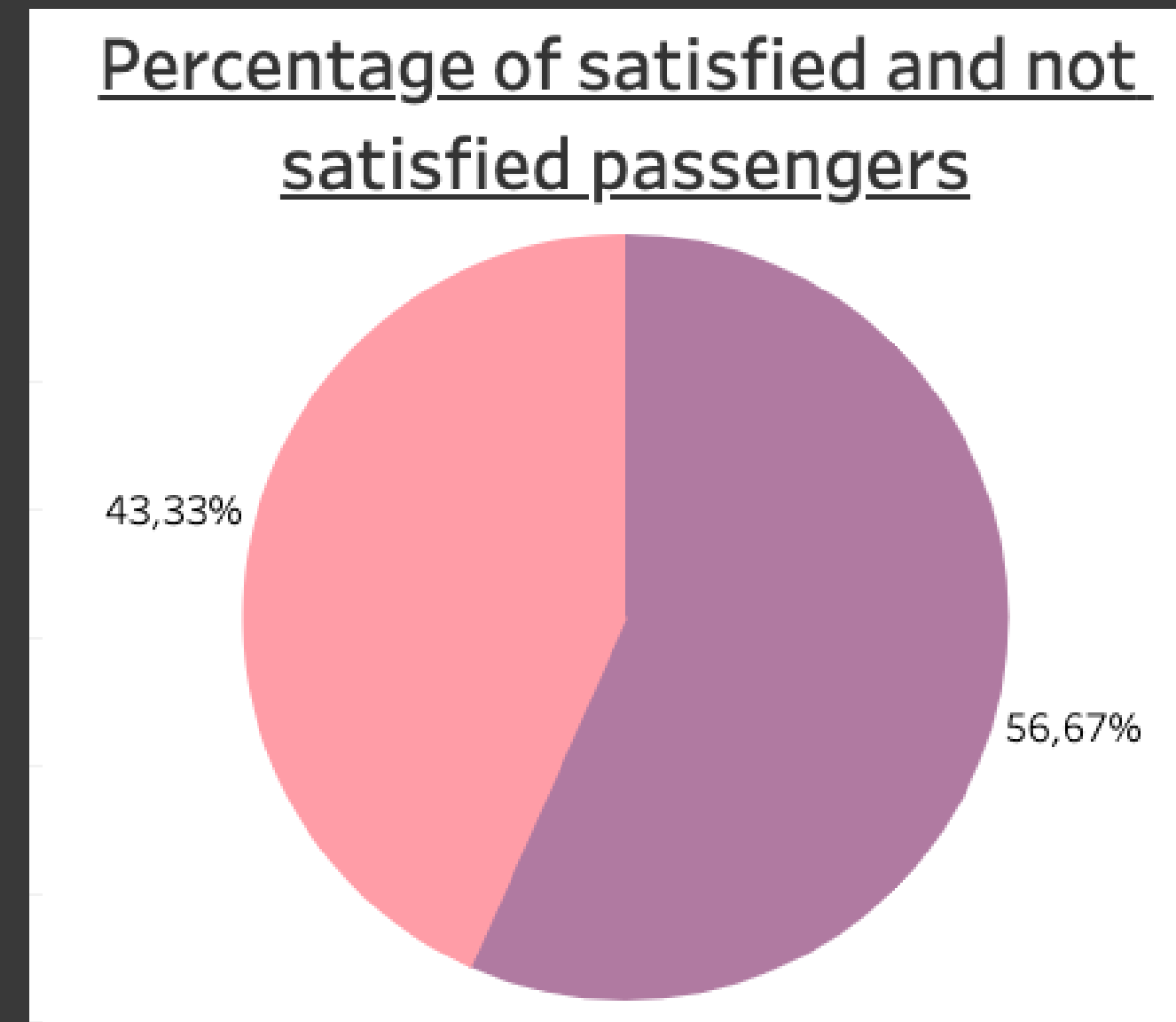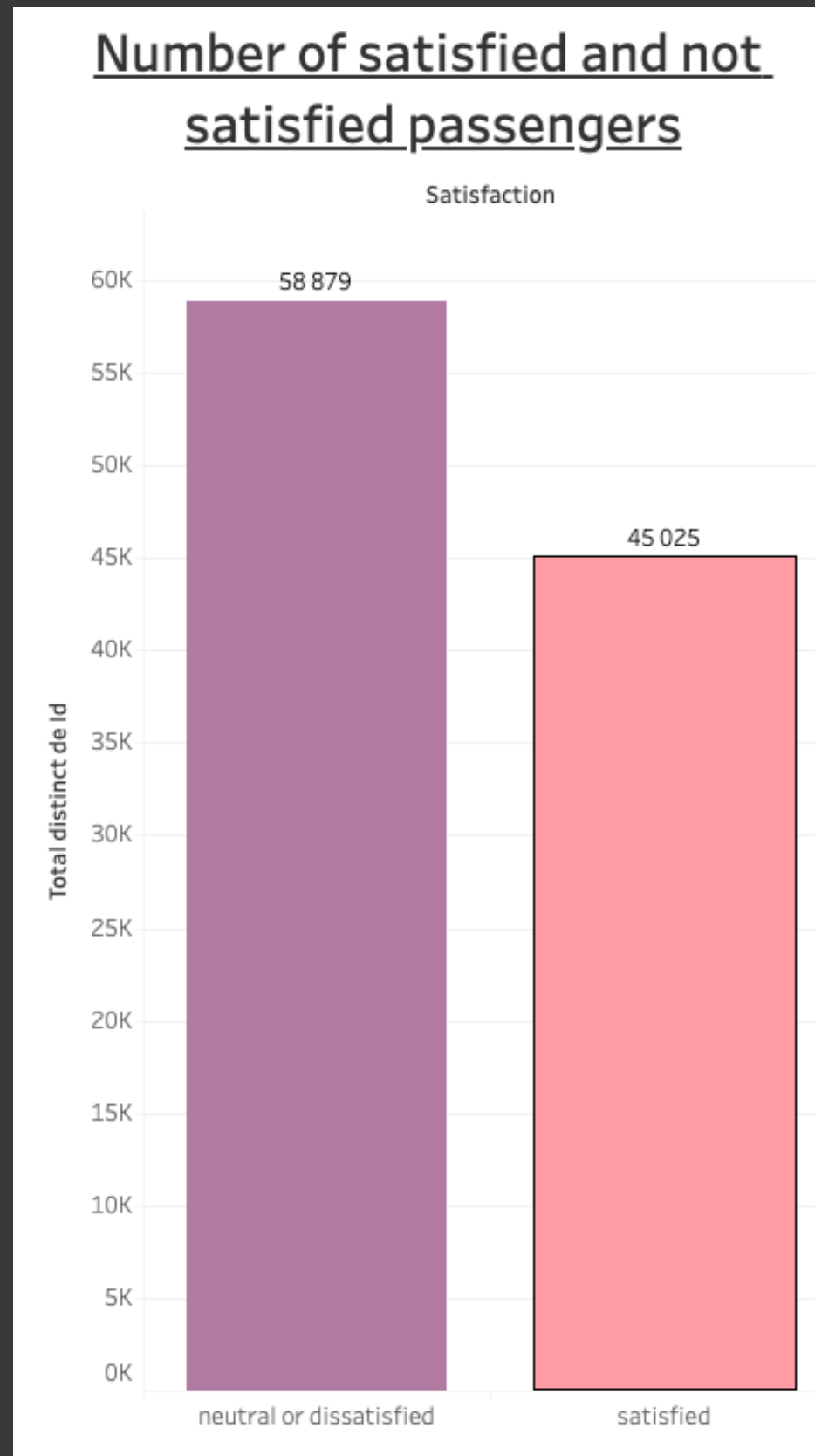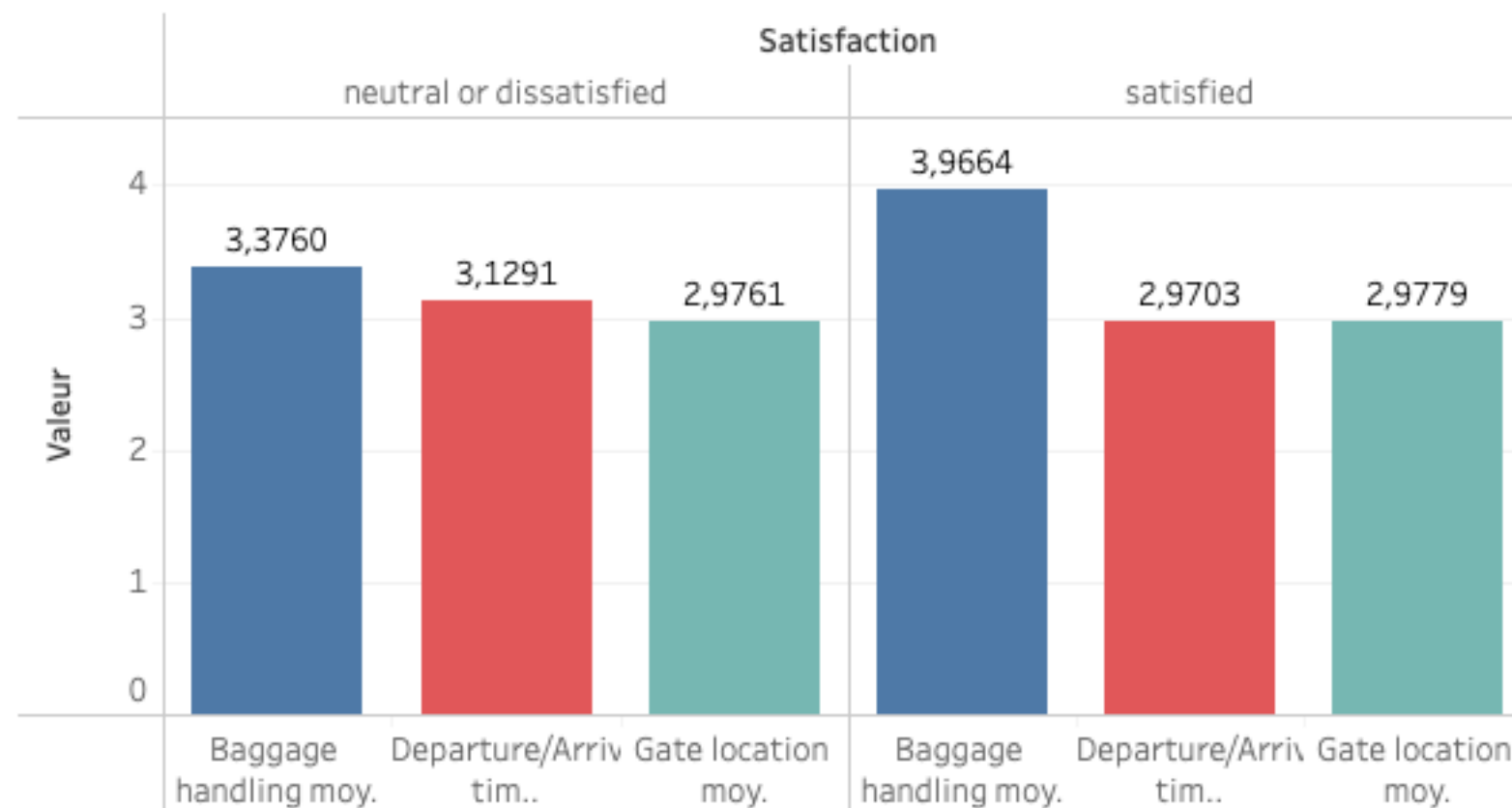
Average score regarding to the satisfaction

Average scores about plane's services

Satisfaction

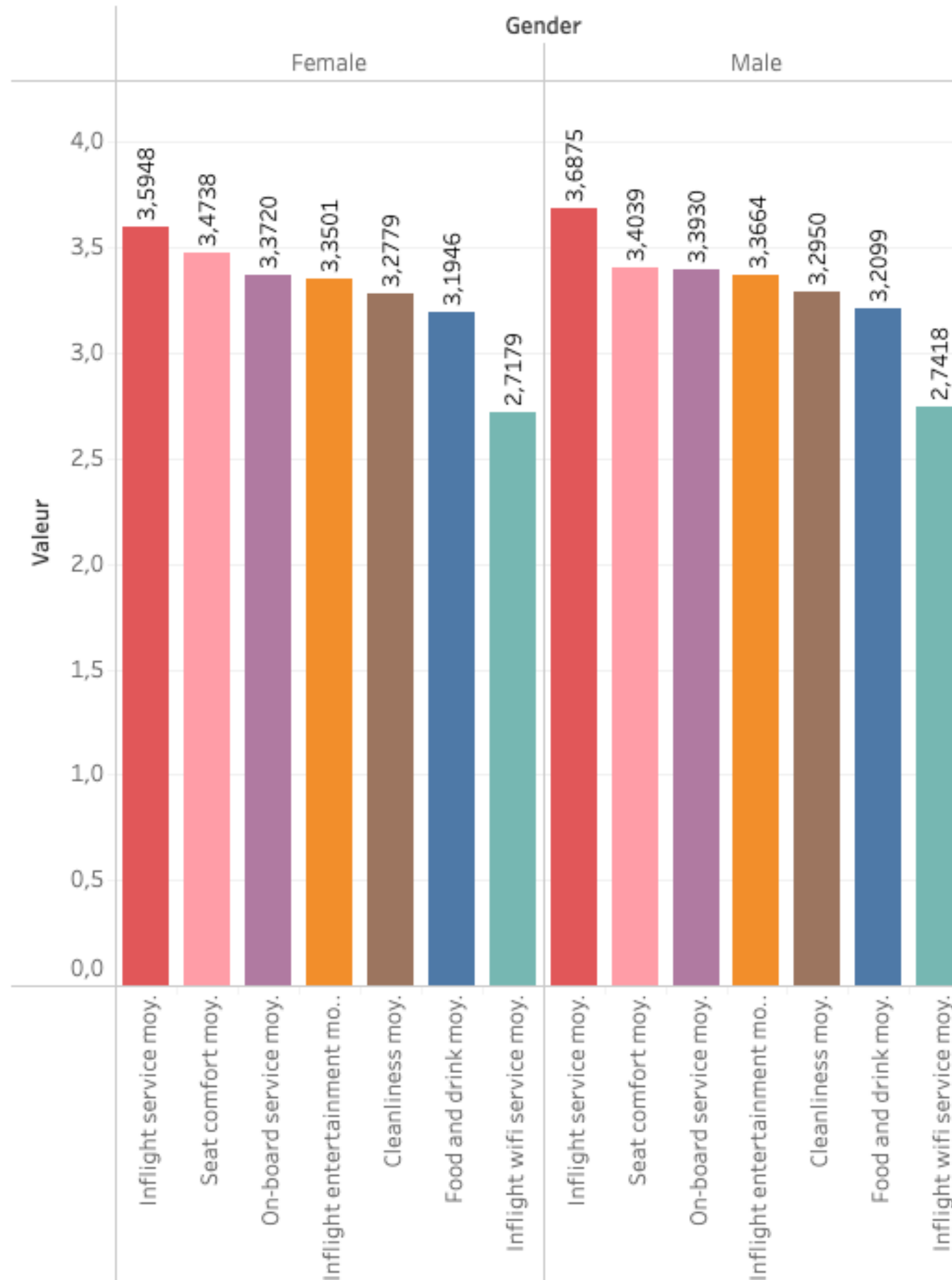| neutral or dissatisfied | satisfied |

| Inflight service moy. | Seat comfort moy. | On-board service moy. | Inflight enter.. | Cleanliness mo.. | Food and drink moy. | Inflight wifi service moy. | Inflight service moy. | Seat comfort moy. | On-board service moy. | Inflight enter.. | Cleanliness mo.. | Food and drink moy. | Inflight wifi service moy. |
| 3,389 | 3,036 | 3,019 | 2,894 | 2,936 | 2,958 | 2,400 | 3,969 | 3,967 | 3,857 | 3,965 | 3,744 | 3,521 | 3,161 |

Average scores about online services

Satisfaction

| neutral or dissatisfied | satisfied |

| Online boardin.. | Ease of Online booking moy. | Online boardin.. | Ease of Online.. |
| 2,656 | 2,547 | 4,027 | 3,032 |

Average scores about airport's services

Satisfaction

| neutral or dissatisfied | satisfied |

| Baggage handling moy. | Departure/Arriv tim.. | Gate location moy. | Baggage handling moy. | Departure/Arriv tim.. | Gate location moy. |
| 3,3760 | 3,1291 | 2,9761 | 3,9664 | 2,9703 | 2,9779 |

Average score regarding to passenger's gender

Average scores about plane's services

Gender

| Female | Male |

Valeur

Female:
- Inflight service moy.: 3,5948
- Seat comfort moy.: 3,4738
- On-board service moy.: 3,3720
- Inflight entertainment mo..: 3,3501
- Cleanliness moy.: 3,2779
- Food and drink moy.: 3,1946
- Inflight wifi service moy.: 2,7179

Male:
- Inflight service moy.: 3,6875
- Seat comfort moy.: 3,4039
- On-board service moy.: 3,3930
- Inflight entertainment mo..: 3,3664
- Cleanliness moy.: 3,2950
- Food and drink moy.: 3,2099
- Inflight wifi service moy.: 2,7418

Average scores about airport's services

Gender

| Female | Male |

Valeur

Female:
- Baggage handlin..: 3,5883
- Departure/.. ti..: 3,0466
- Gate locatio..: 2,9765

Male:
- Baggage handlin..: 3,6767
- Departure/.. ti..: 3,0744
- Gate location..: 2,9773

Average scores about online services

Gender

| Female | Male |

Valeur

Female:
- Online boarding moy.: 3,3066
- Ease of Online booking moy.: 2,7473

Male:
- Online boarding moy.: 3,1925
- Ease of Online booking moy.: 2,7668

4) SQL analyse

```python
query = '''
    SELECT satisfaction, gender, count(id)
    FROM data
    group by satisfaction, gender
    '''


data = pd.read_sql_query(query, engine)
data.head()
```

	satisfaction	gender	count(id)
0	neutral or dissatisfied	Male	20024
1	satisfied	Female	16960
2	neutral or dissatisfied	Female	21137
3	satisfied	Male	16998

```python
query = '''
    SELECT class,gender, avg(l.Total_Score)
    FROM (
    SELECT class, gender, (inflight_wifi_service+"departure/arrival_time_convenient"+ease_of_online_booking+
    FROM data) l
    group by class, gender
    '''

data = pd.read_sql_query(query, engine)
data.head()
```

	class	gender	avg(l.Total_Score)
0	Business	Male	41.730610
1	Business	Female	41.586083
2	Eco	Female	36.676786
3	Eco	Male	36.938678
4	Eco Plus	Female	36.790294
5	Eco Plus	Male	37.154412

The scores are on 70.

5) Results

Model	Score	Results
Logistic Regression	0,97	As we can see, the score is really high. I checked through a matrix, to understand better this score. This matrix shows that 97,7% of the preditions are "True" and 2,3% are "False".
Random Forest	0,97	This is obtained with a max_depth = 1 as a parameter. The main paramters are the following : random_state=0, max_depth=1, max_features='sqrt', min_samples_leaf=5, min_samples_split=5, n_estimators=250
	0,99	This is obtained with a max_depth = 2 as a parameter.
	1	This is obtained with a max_depth = 3 as a parameter. After 3, the score is always 1.
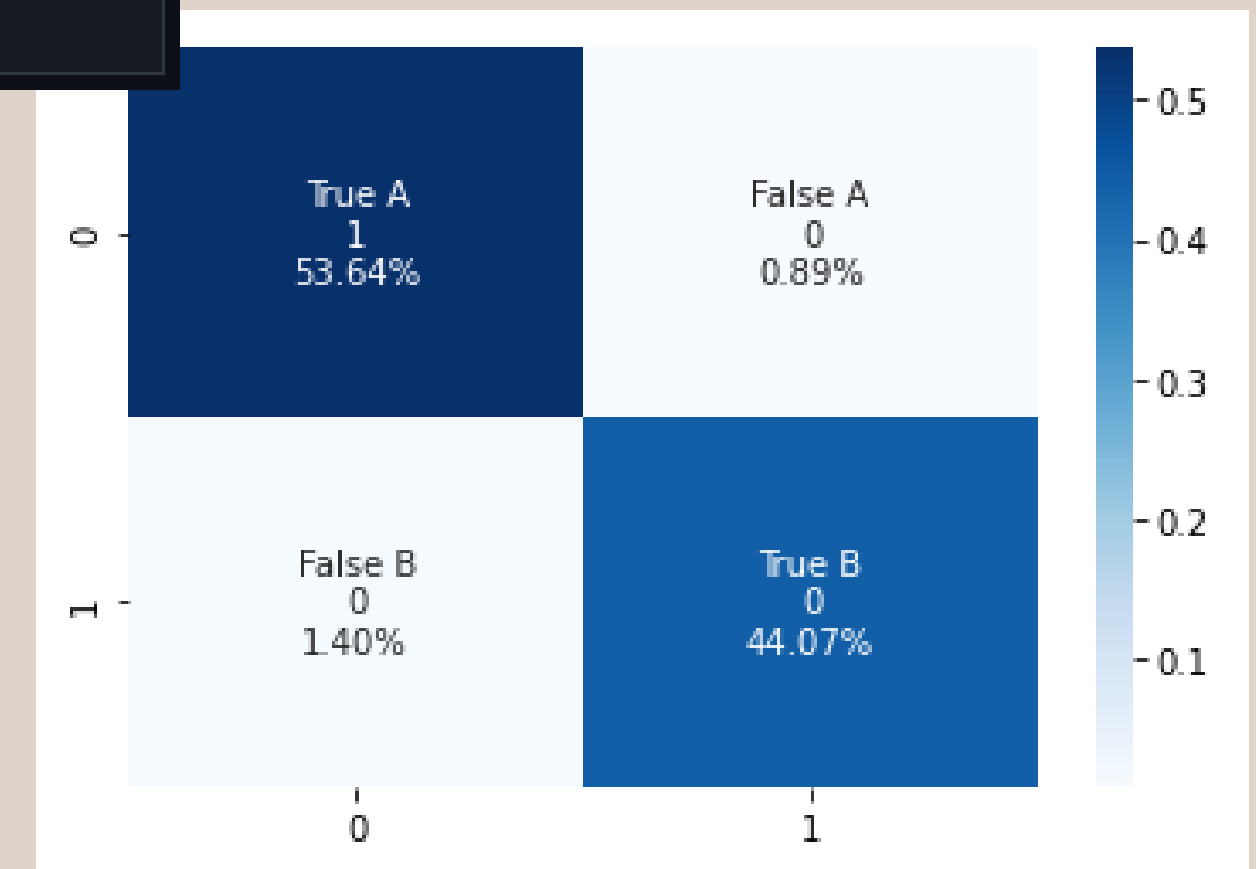
I would like to thank Abhi and his team for this incredible experience full of learning and enjoy !