Airline Passenger Satisfaction

ARWA ALSHANKITI – PROJECT 1-

Problem:

- As we all know we love to travel, but more importantly, the airlines we chose! To be satisfied with them.

- so, we want to see which are the most factors that effect that

Data :

- Data source from :

https://www.kaggle.com/johnddddddd/customer-satisfaction?select=satisfaction.xlsx

- The data contained 129,880 records with 24 features.

	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service
0	19556	Female	Loyal Customer	52	Business travel	Eco	160	5	4	3	...	5	5	5	5	2
1	90035	Female	Loyal Customer	36	Business travel	Business	2863	1	1	3	...	4	4	4	4	3
2	12360	Male	disloyal Customer	20	Business travel	Eco	192	2	0	2	...	2	4	1	3	2
3	77959	Male	Loyal Customer	44	Business travel	Business	3377	0	0	0	...	1	1	1	1	3
4	36875	Female	Loyal Customer	49	Business travel	Eco	1182	2	3	4	...	2	2	2	2	4

EDA :

```
df.info()
```

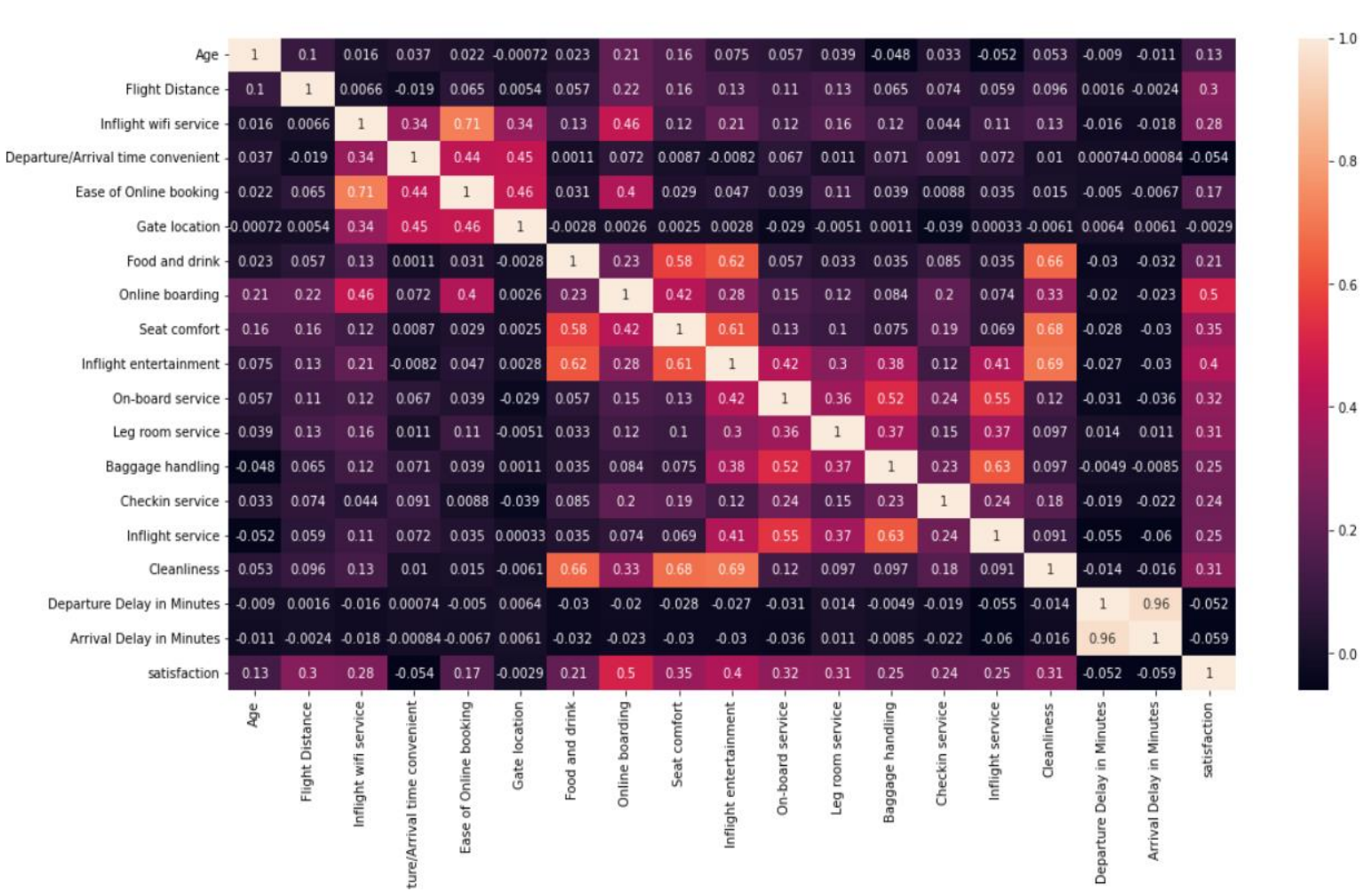
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129880 entries, 0 to 129879
Data columns (total 24 columns):
 #   Column                            Non-Null Count   Dtype  
---  ------                            --------------   -----  
 0   id                                129880 non-null  int64  
 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   Inflight wifi service             129880 non-null  int64  
 8   Departure/Arrival time convenient 129880 non-null  int64  
 9   Ease of Online booking            129880 non-null  int64  
 10  Gate location                     129880 non-null  int64  
 11  Food and drink                    129880 non-null  int64  
 12  Online boarding                   129880 non-null  int64  
 13  Seat comfort                      129880 non-null  int64  
 14  Inflight entertainment            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  Inflight service                  129880 non-null  int64  
 20  Cleanliness                       129880 non-null  int64  
 21  Departure Delay in Minutes        129880 non-null  int64  
 22  Arrival Delay in Minutes          129487 non-null  float64
 23  satisfaction                      129880 non-null  object 
dtypes: float64(1), int64(18), object(5)
memory usage: 23.8+ MB
```

EDA :

```
df.describe().transpose()
```

Food and drink	129487.0	3.204685	1.329905	0.0	2.0	3.0	4.0	5.0
Online boarding	129487.0	3.252720	1.350651	0.0	2.0	3.0	4.0	5.0
Seat comfort	129487.0	3.441589	1.319168	0.0	2.0	4.0	5.0	5.0
Inflight entertainment	129487.0	3.358067	1.334149	0.0	2.0	4.0	4.0	5.0
On-board service	129487.0	3.383204	1.287032	0.0	2.0	4.0	4.0	5.0
Leg room service	129487.0	3.351078	1.316132	0.0	2.0	4.0	4.0	5.0
Baggage handling	129487.0	3.631886	1.180082	1.0	3.0	4.0	5.0	5.0
Checkin service	129487.0	3.306239	1.266146	0.0	3.0	3.0	4.0	5.0
Inflight service	129487.0	3.642373	1.176614	0.0	3.0	4.0	5.0	5.0
Cleanliness	129487.0	3.286222	1.313624	0.0	2.0	3.0	4.0	5.0
Departure Delay in Minutes	129487.0	14.643385	37.932867	0.0	0.0	0.0	12.0	1592.0
Arrival Delay in Minutes	129487.0	15.091129	38.465650	0.0	0.0	0.0	13.0	1584.0

Correlation



ML Models

Train – test split

Models used :

- KNN
- Decision Tree
- Random Forest
- XGBoost

Result & conclusion

Accuracy:-

KNN: 0.751477005058501
Decision Tree: 0.9464030582692976
Random Forest: 0.9642429625053095


THANK YOU !