



Final Project

Subervised By Dr Doaa Mahmoud

Represented By Menna And Mostafa



Airlines Customer Satisfaction



Meet Our Team



Menna

student at computer
science cairo university
AI intern at Samsung
Machine learning inter at
ITIDIA



Mostafa

student at computer
science cairo university
AI intern at Samsung

Agenda

Meet the Team

Contents of data

**Modeling and
Evaluation**

About

Feature Selection

Conclusion

**Problem
Statement**

**Exploratory data
analysis**

Recomendation

About :

- This data set contains a survey on air passenger satisfaction.
- it can affect whether a customer is likely to reuse the service and whether they are likely to recommend the service
- we want to predict the customer satisfaction.



The Problem

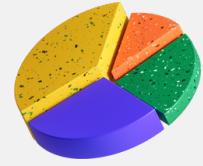


The main objective of this study is to identify the level of satisfaction of the passengers to know the quality of services provided by the airline companies and the key factors that derive customer satisfaction.

How can we improve our service and solve this problem ?



Data Features



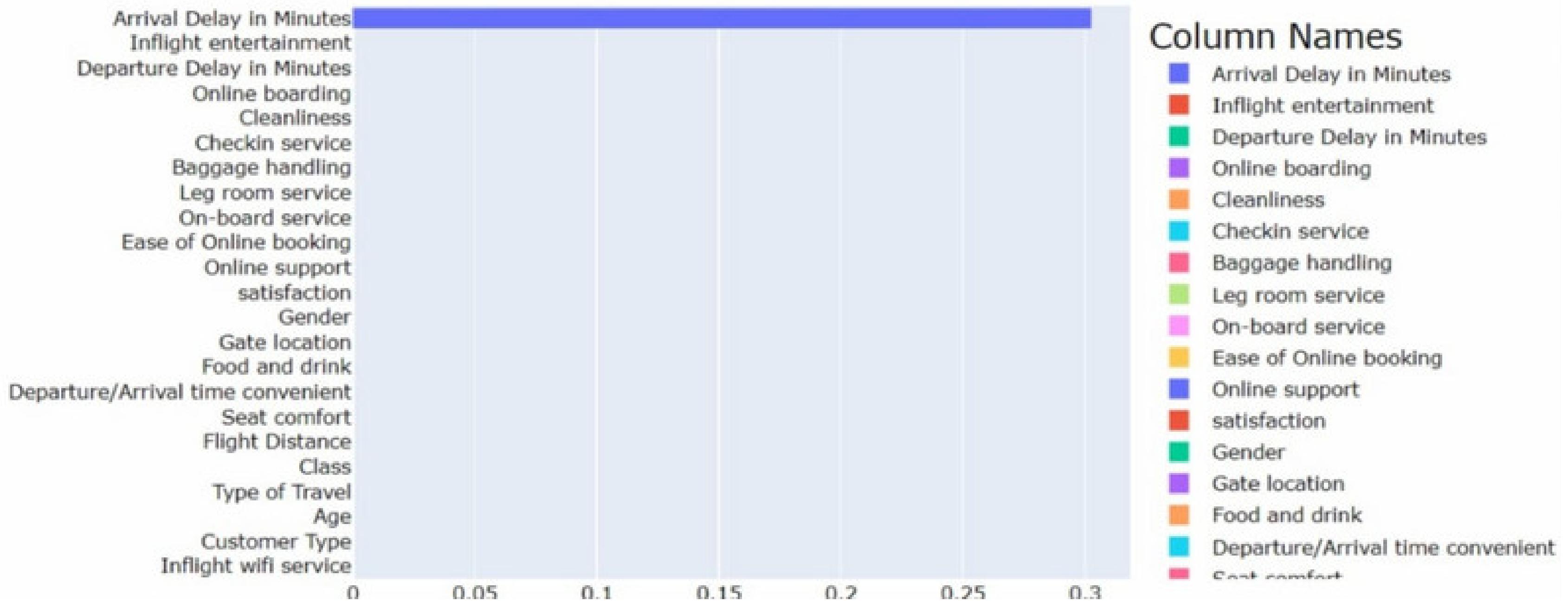
**the data contains 23 Features and
129880**



most of the features is numeric

Age
Flight Distance
Seat comfort
Departure/Arrival time convenient
Food and drink
Gate location
Inflight wifi service
Inflight entertainment
Online support
Ease of Online booking
On-board service
Leg room service
Baggage handling
Checkin service
Cleanliness
Online boarding
Departure Delay in Minutes
Arrival Delay in Minutes

Discovering the Missing values:



- Arrival Delay in Minutes feature has 393 missing values so we drop them

Detecting the outliers :



as we see Flight Distance feature contains the biggest number of outliers



we handled them using our detection function

Type of Travel	1.0
Class	1.0
Flight Distance	1185.0
Seat comfort	2.0
Departure/Arrival time convenient	2.0
Food and drink	2.0
Gate location	2.0
Inflight wifi service	2.0
Inflight entertainment	2.0
Online support	2.0
Ease of Online booking	3.0
On-board service	1.0
Leg room service	3.0
Baggage handling	2.0
Checkin service	1.0
Cleanliness	2.0
Online boarding	2.0
Departure Delay in Minutes	12.0
dtype: float64	

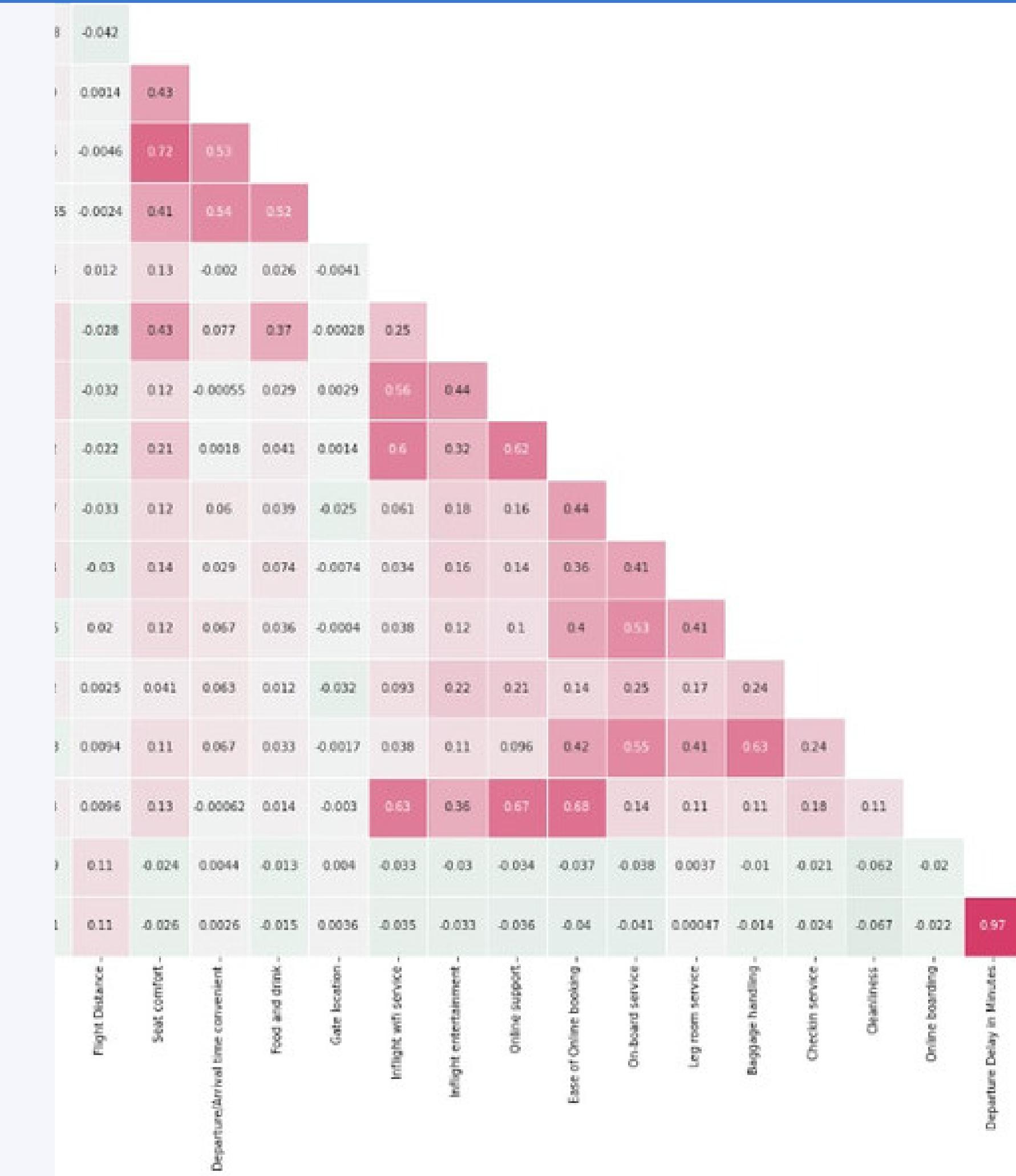
Correlation



Their is high correlation between :
 food-drink and seat comfort
 Ease of online booking and online boarding
 online support and online boarding
 inflight service and online boarding



the highest correlation is between Arrival Delay in Minutes and Departure Delay in Minutes



Feature Selection

- We used different methods to get the most important features
- we used chi square and the Random Forest Importance Method

```
Index(['Gender', 'Customer Type', 'Type of Travel', 'Class',
       'Inflight entertainment', 'Online support', 'Ease of Online booking',
       'On-board service', 'Leg room service', 'Online boarding'],
      dtype='object')
```

Label Encoding :

Before

	satisfaction	Gender	Customer Type
0	satisfied	Female	Loyal Customer
1	satisfied	Male	Loyal Customer
2	satisfied	Female	Loyal Customer
3	satisfied	Female	Loyal Customer
4	satisfied	Female	Loyal Customer

After

	satisfaction	Gender	Customer Type
0	1	0	0
1	1	1	0
2	1	0	0
3	1	0	0
4	1	0	0



How Many People is Satisfied and Dissatisfied ?

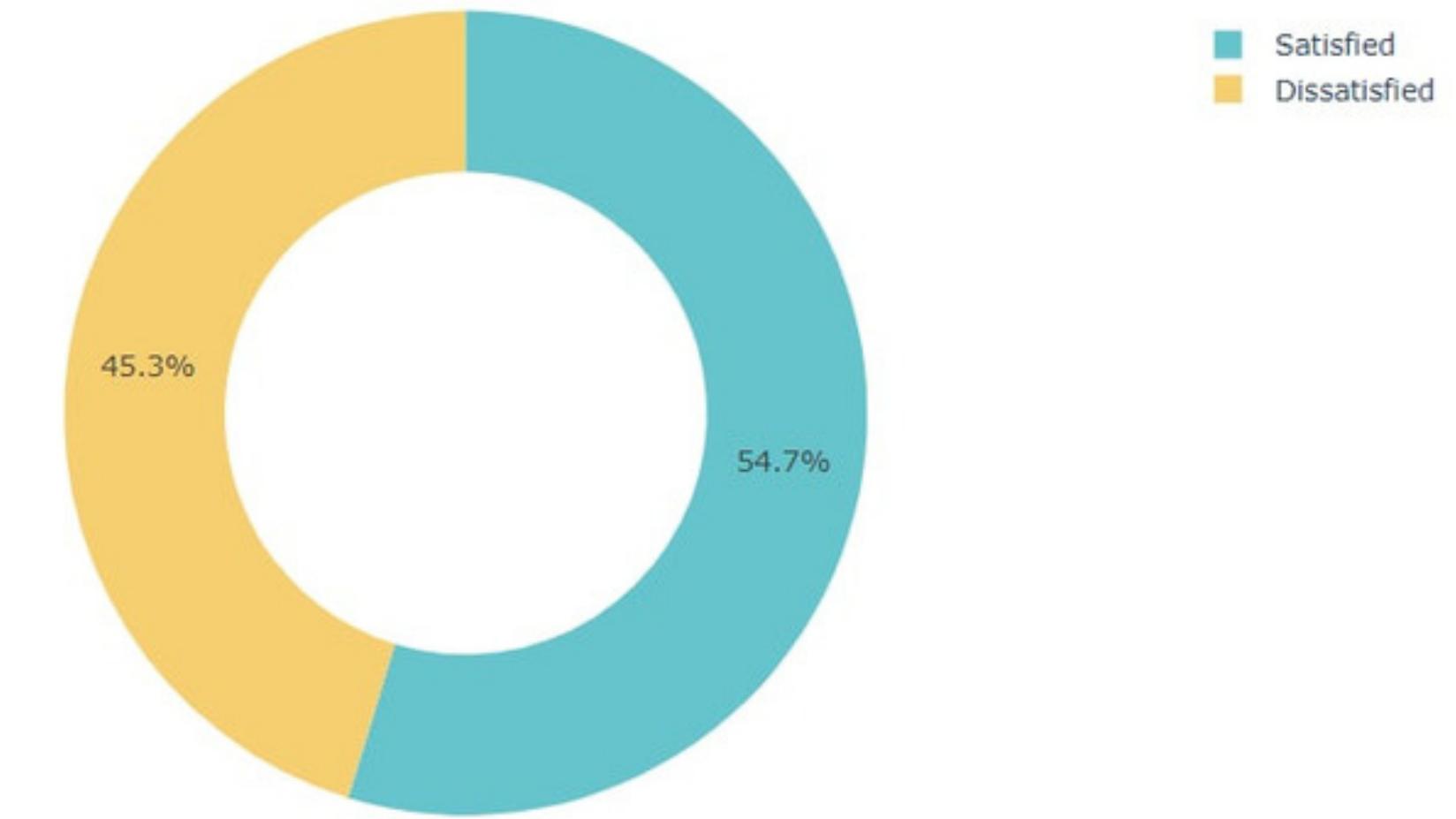


**45.3% is
Dissatisfied**



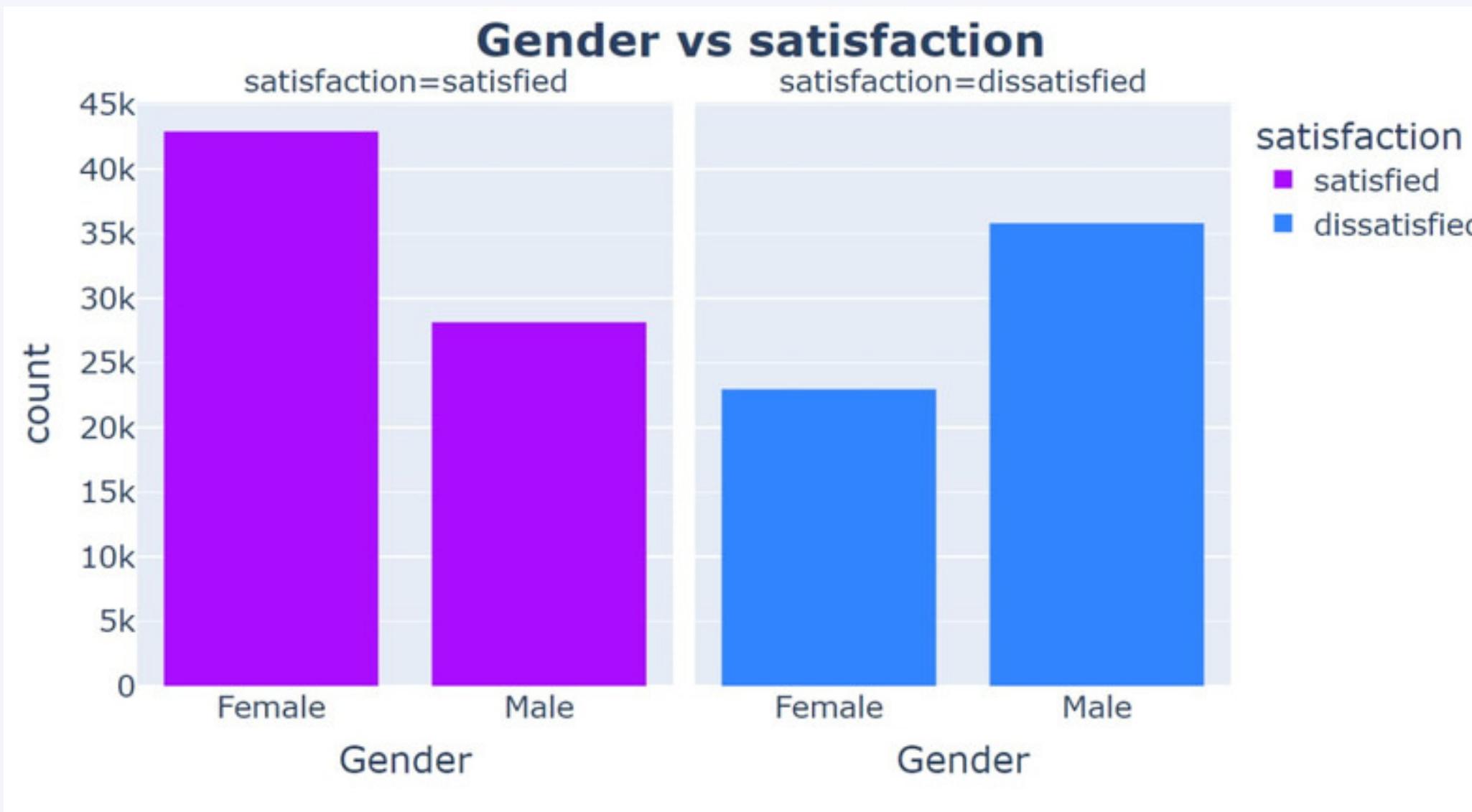
**54.7% is
Satisfied**

Satisfied And Dissatisfied Ratio



- The dissatisfied people is less than the satisfied

How many women is dissatisfied and how many men is satisfied ?



the percentage of the man that is dissatisfied is more the percentage of women



the number of males and females almost the same there is a slightly difference.



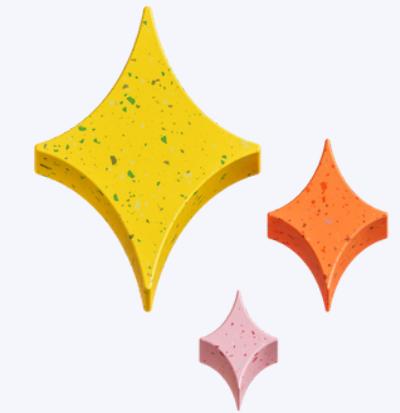
Which type that has the most satisfied people ?



Loyal Customers much more than Disloyal Customers that looks good for our services.



Loyal Customers we have more dissatisfaction and that's a problem.



What is the ages of the customers who is satisfied and dissatisfied?



Most of Young people from 20-35 is not satisfied with the service



The old people that their age range is between 40 to 60 is satisfied with the service.



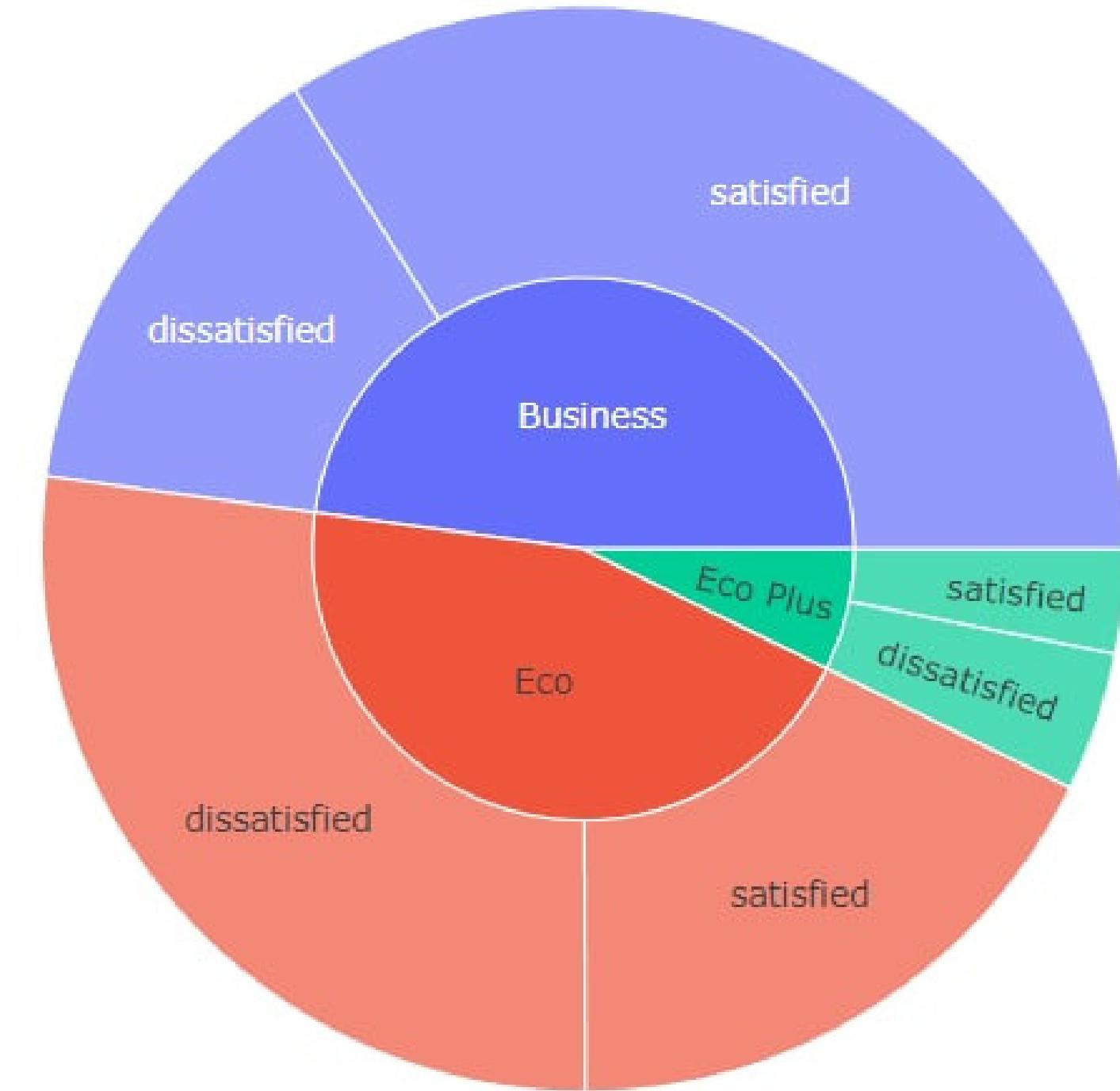
Our conclusion from that the company care more about old people



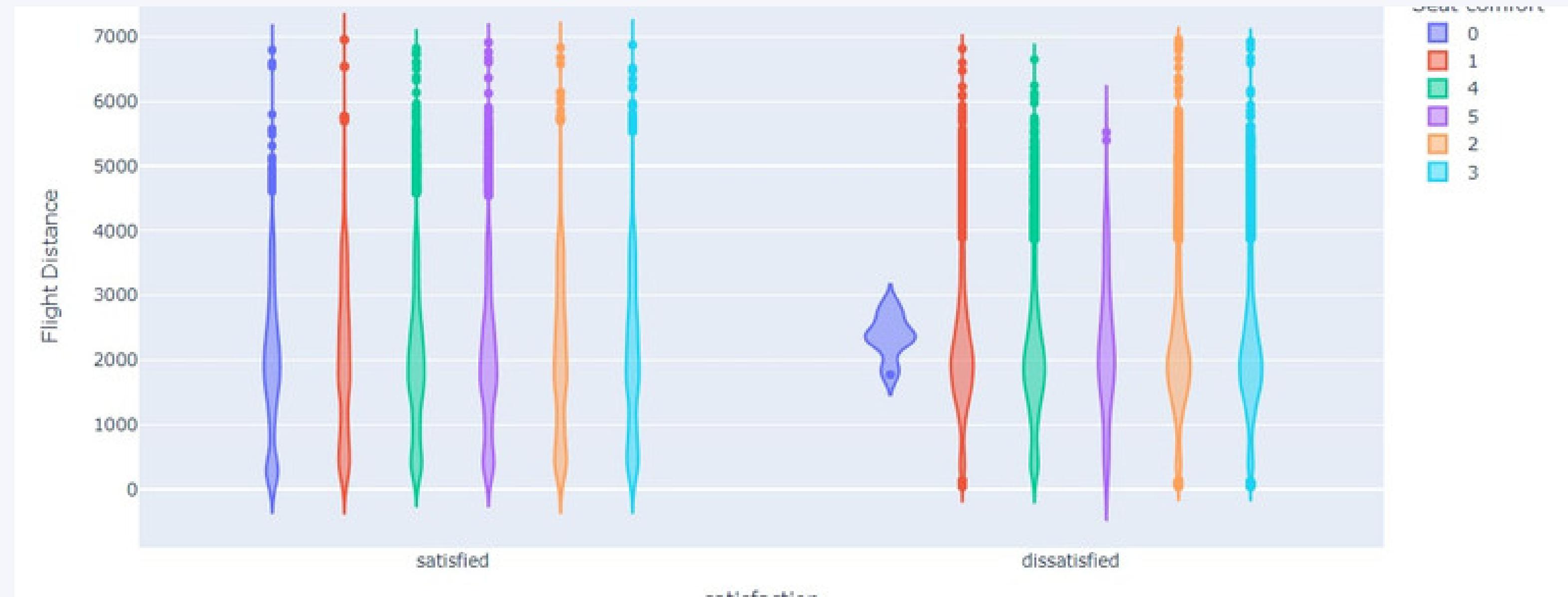
Which class has more dissatisfied people ?

65%

Dissatisfied people is
in class Eco

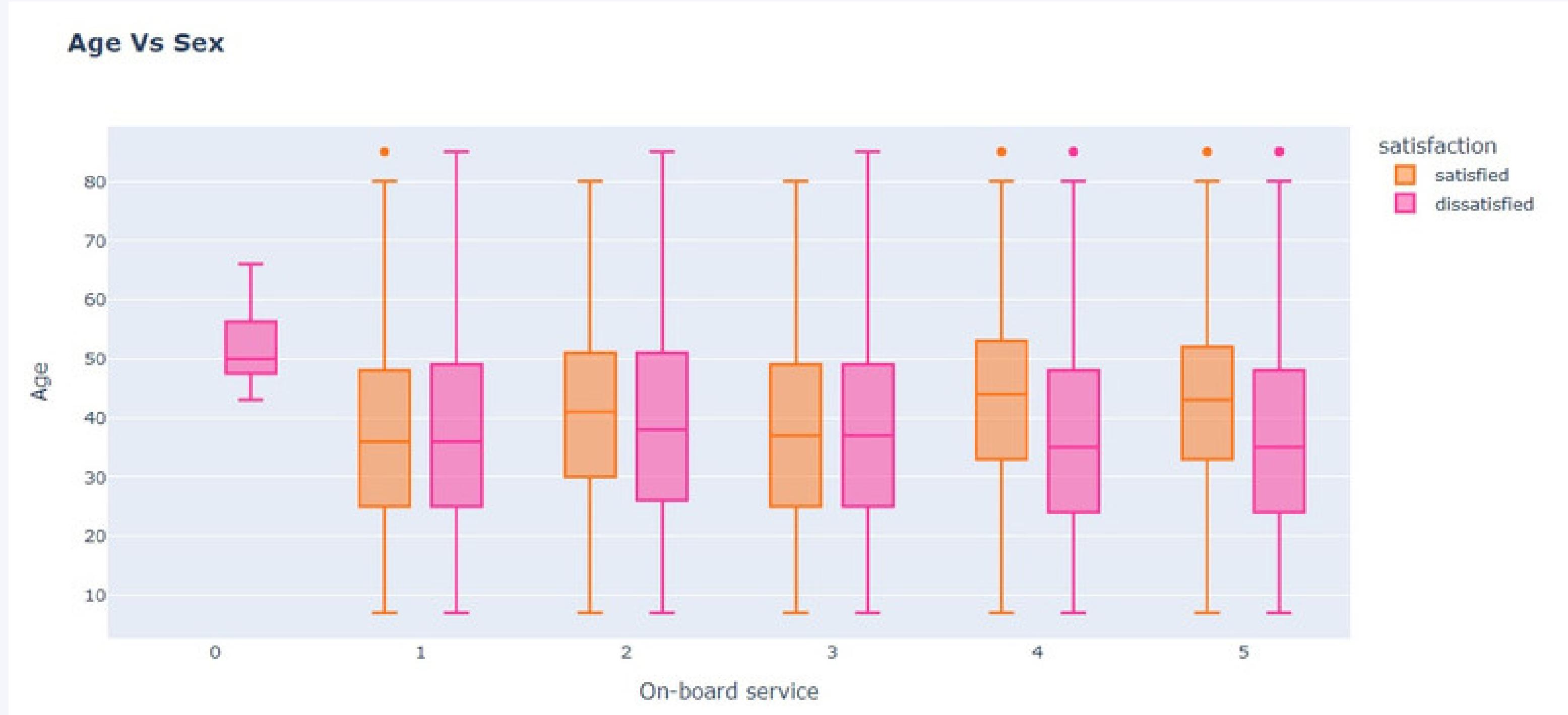


Which factors that affect the dissatisfied people ?



- The most dissatisfied people their flight distance is between 1800 to 2800

What is the ages of people that dissatisfied with On-board service ?



- Ages of people that dissatisfied with On-board service their age concentrated between 43 to 66 years old

Modeling and Evaluation

After splitting the data into train and test split we tasted our data with 8 algorithms :

- logistic regression
- support vector machine
- Decision Tree
- Random Forest
- XGB Classifier
- Adaboost
- Naive Bayes Classifier



- each model with its accuracy

```
Using model: Logistic Regression
```

```
Trainning Score: 0.8237843096072427
```

```
Test Score: 0.822496195610121
```

```
Acc Train: 0.8237843096072427
```

```
Acc Test: 0.822496195610121
```

```
Using model: Support Vector Machine
```

```
Trainning Score: 0.9236925593889215
```

```
Test Score: 0.9211008227799128
```

```
Acc Train: 0.9236925593889215
```

```
Acc Test: 0.9211008227799128
```

```
Using model: Decision Tree
```

```
Trainning Score: 0.9620618374362999
```

```
Test Score: 0.9294575842769079
```

```
Acc Train: 0.9620618374362999
```

```
Acc Test: 0.9294575842769079
```

```
Using model: Random Forest
```

```
Trainning Score: 0.9620618374362999
```

```
Test Score: 0.9342033994480411
```

```
Acc Train: 0.9620618374362999
```

```
Acc Test: 0.9342033994480411
```

```
Using model: XGBClassifier
```

```
Trainning Score: 0.9445740247393962
```

```
Test Score: 0.9375822135100977
```

```
Acc Train: 0.9445740247393962
```

```
Acc Test: 0.9375822135100977
```

Decision Tree Classifier

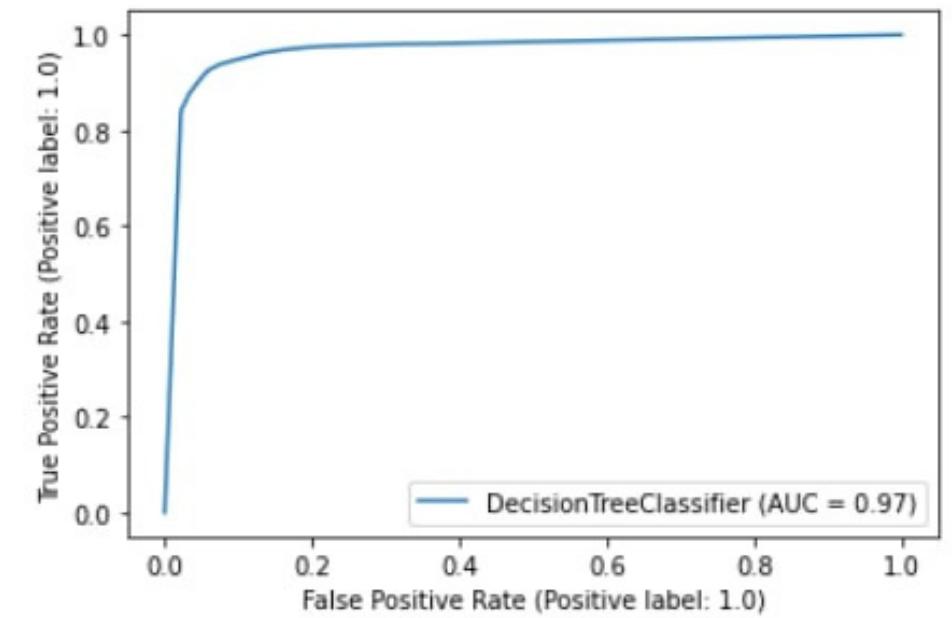
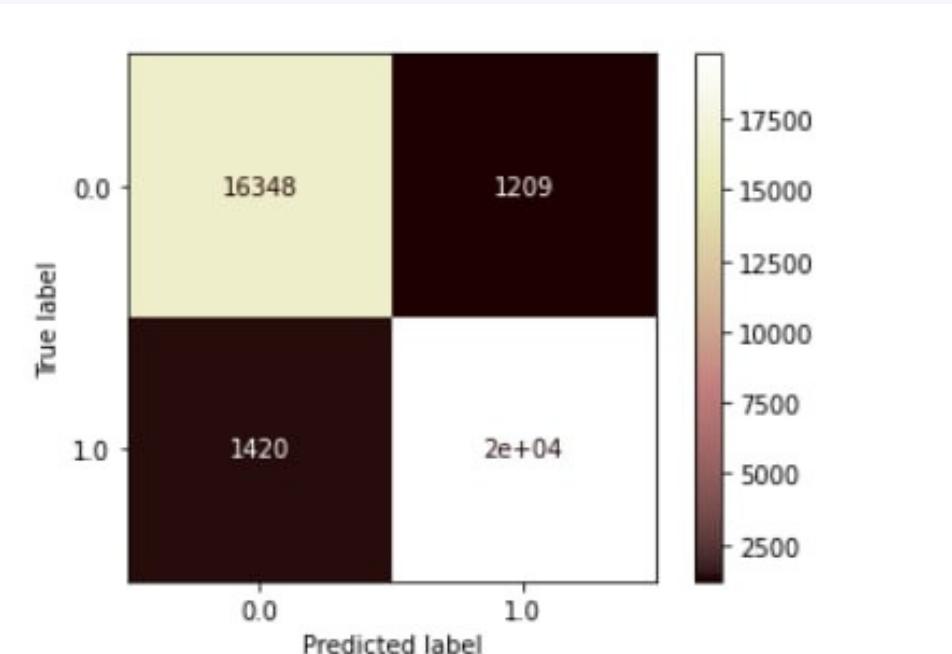
- we evaluated the models using F1 score, recall and precision

Accuracy = 0.9321915864950607

ROC Area under Curve = 0.9321008243810961

Time taken = 0.3052675724029541

	precision	recall	f1-score	support
0.0	0.92008	0.93114	0.92558	17557
1.0	0.94244	0.93306	0.93773	21214
accuracy			0.93219	38771
macro avg	0.93126	0.93210	0.93165	38771
weighted avg	0.93231	0.93219	0.93222	38771



Ada Boost

- we evaluated the models using F1 score, recall and precision

Accuracy = 0.8897113822186686

ROC Area under Curve = 0.8888833614381761

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0.0	0.87681	0.88010	0.87845	17557
-----	---------	---------	---------	-------

1.0	0.90046	0.89766	0.89906	21214
-----	---------	---------	---------	-------

accuracy			0.88971	38771
----------	--	--	---------	-------

macro avg	0.88864	0.88888	0.88876	38771
-----------	---------	---------	---------	-------

weighted avg	0.88975	0.88971	0.88973	38771
--------------	---------	---------	---------	-------

Models Comparison

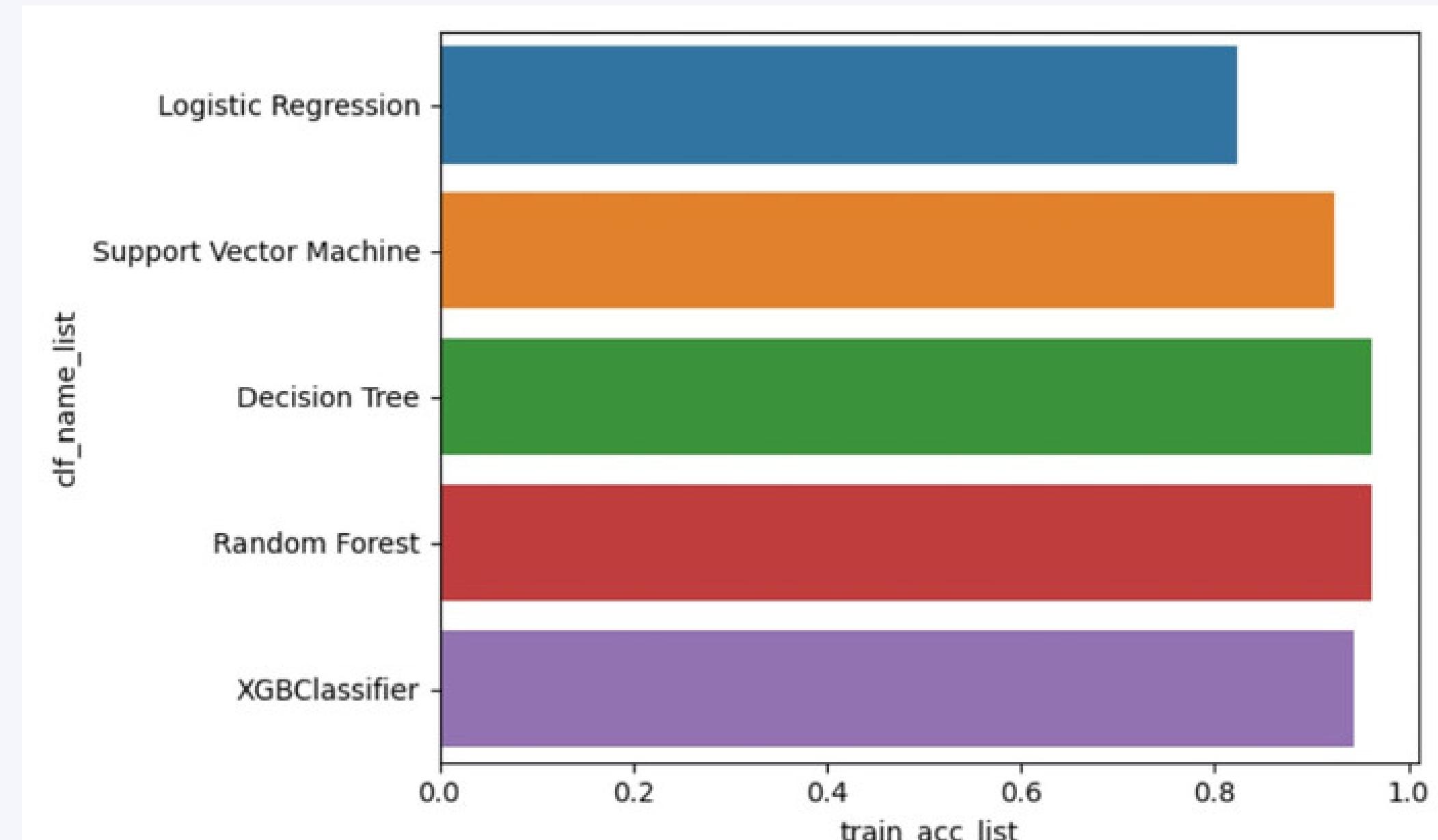
The highest accuracy is :

92%

**Decision Tree
Classifier**

93%

Random Forest



Conclusion

- **the airline should tackle such issues as a priority to improve services.**
- **There are also aspects such as ticket price, flight location also missing from this dataset**

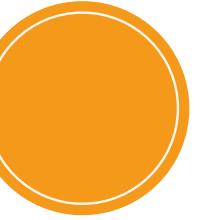


Recomended solutions

- **About business, I will suggest to the company to improve the services that have low score for the Eco class:**
 - **food_and_drink**
 - **online_boarding**
 - **seat_comfort**
 - **inflight_entertainment**
 - **on_board_service**
 - **leg_room_service**



Future work



we would love to know more information about our
data like the departure delay caused by bad
weather and was their seatmate is rude or not





GET CONTACT



EMAIL ADDRESS mostafamenna304@gmail.com



PHONE NUMBER **01063557277**



EMAIL ADDRESS mostafa.ahraf154@gmail.com



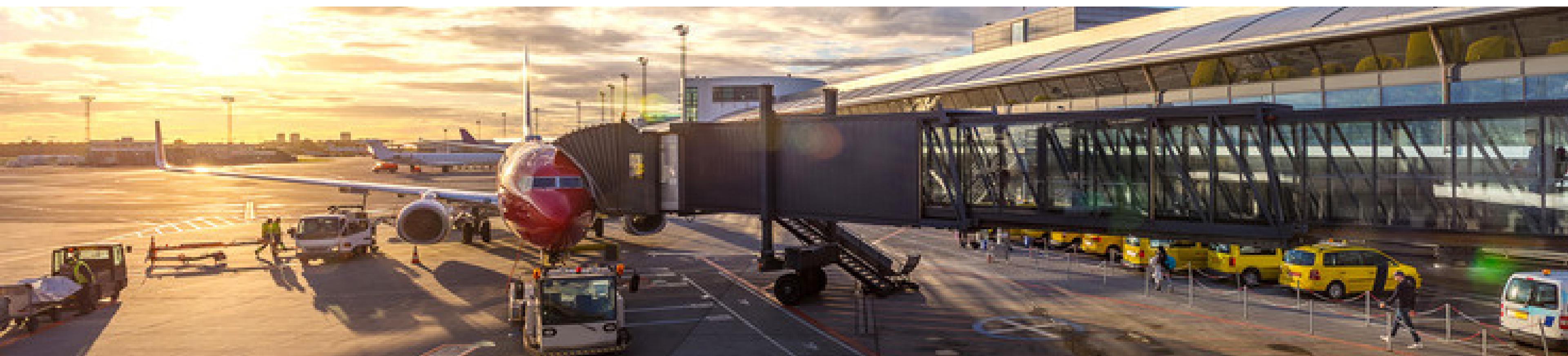
PHONE NUMBER **01283033359**



LINKEDIN [HTTPS://WWW.LINKEDIN.COM/IN/MENNA-MOSTAFA-1AB0311B7](https://www.linkedin.com/in/menna-mostafa-1ab0311b7)



LINKEDIN <https://www.linkedin.com/in/mostafa-ashraf-17913a211/>



THANK YOU!



SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations