Airline Services and Passenger Satisfaction Classification

MAULANA ABUL KALAM AZAD
UNIVERSITY OF TECHNOLOGY,
WEST BENGAL
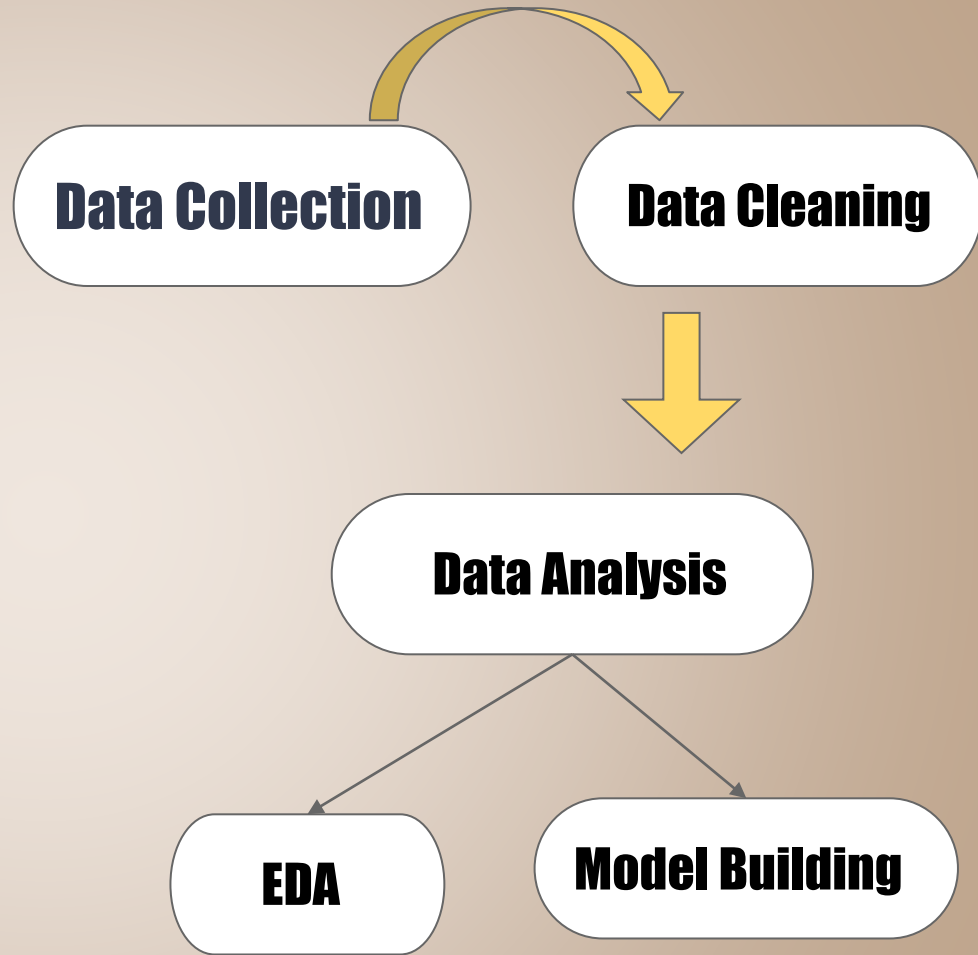
Utech

In Pursuit Of Knowledge And Excellence

Supervised by
Dr. Indrani Mukherjee

By
Brinta Debnath
Reg. No. –213001818010024
Department of Applied Statistics

Introduction

In the highly competitive airline industry, passenger satisfaction is a critical factor that directly influences an airline's success. Satisfied passengers are more likely to become loyal customers, repeat business, and generate positive word-of-mouth recommendations.

Here I built a model to classified customer satisfaction and dissatisfaction depending on airline services.

Data Source:

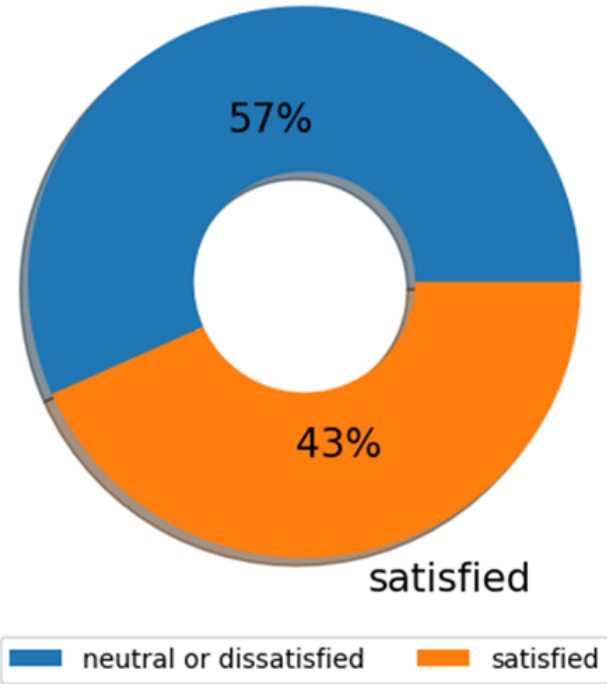https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

Data Preprocessing

Pie Chart of satisfied and dissatisfied Passengers

Here we can see that 57% of total passenger are not satisfied of their overall experience and 43% passengers are satisfied of the airline service.
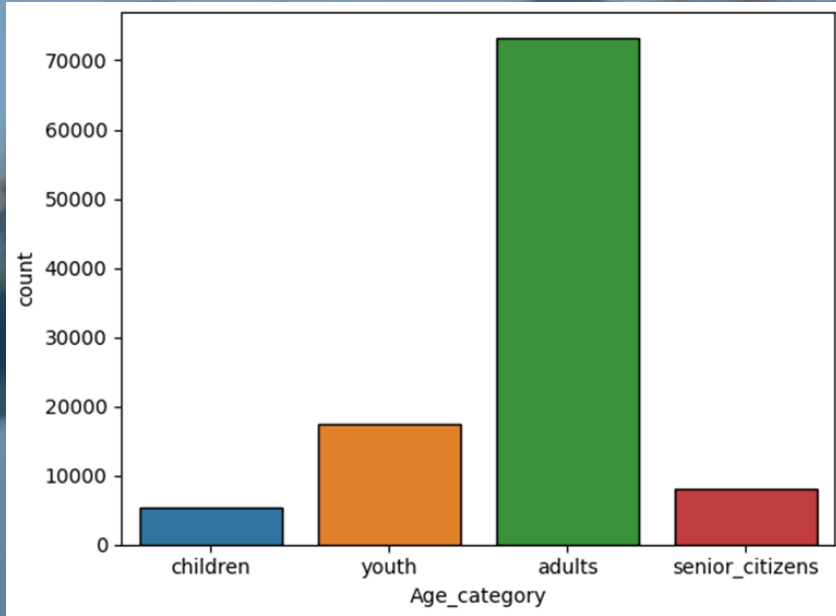
Bar chart of Age Category

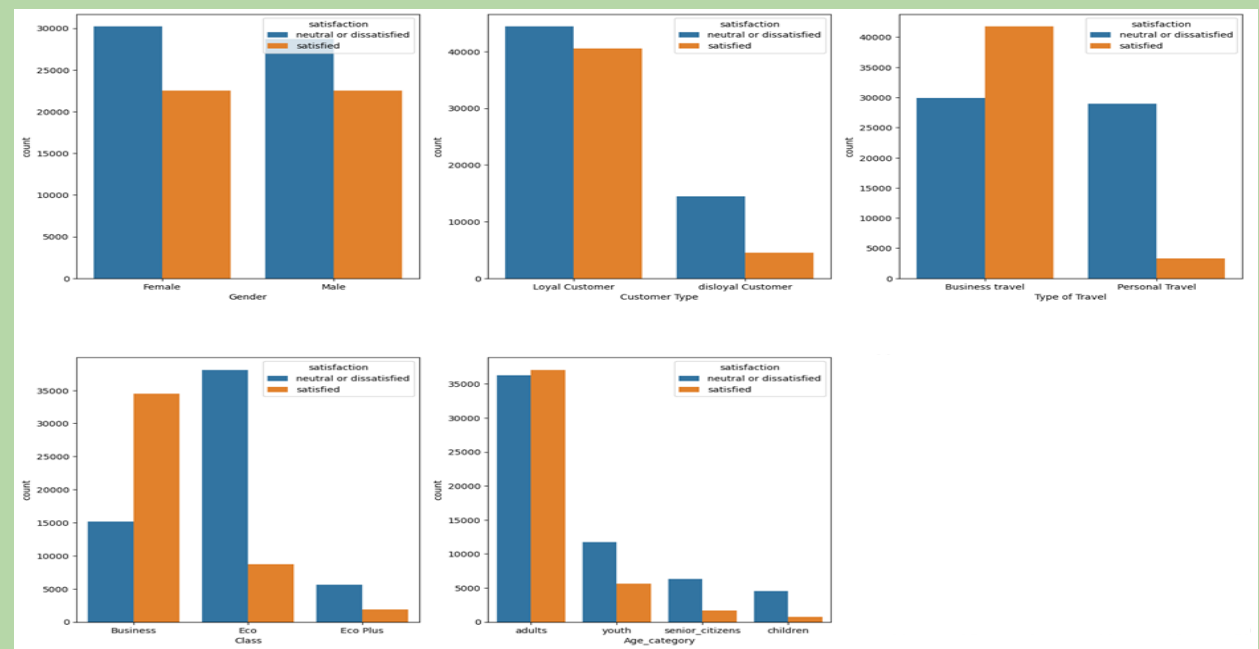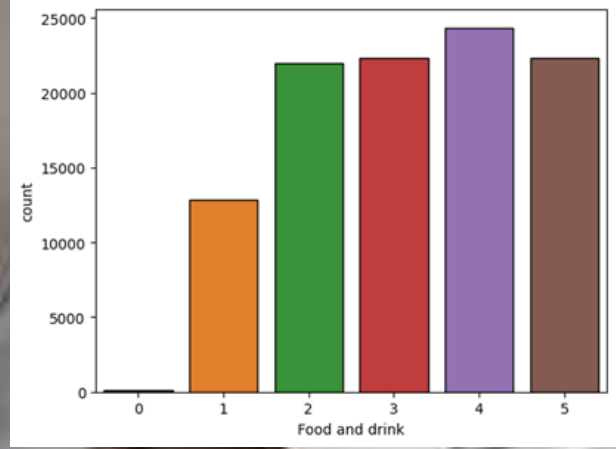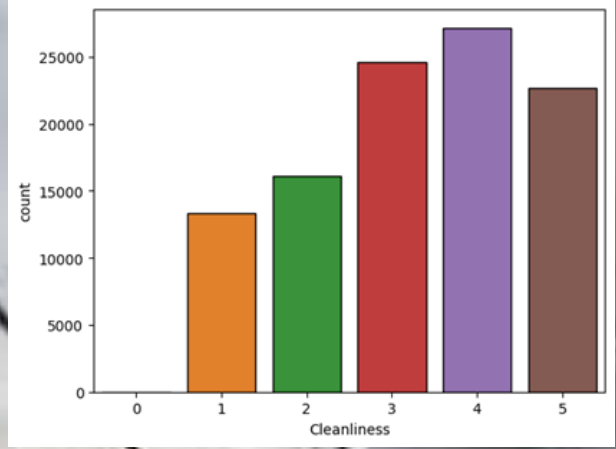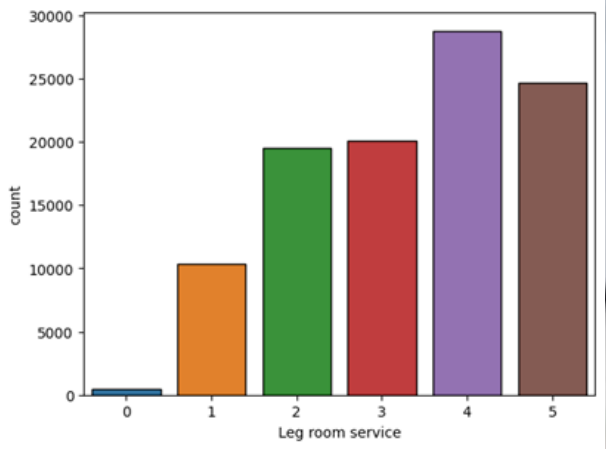Most of the passengers belongs to the adult group(age 26-50). So airline company should also focus on others group ages.
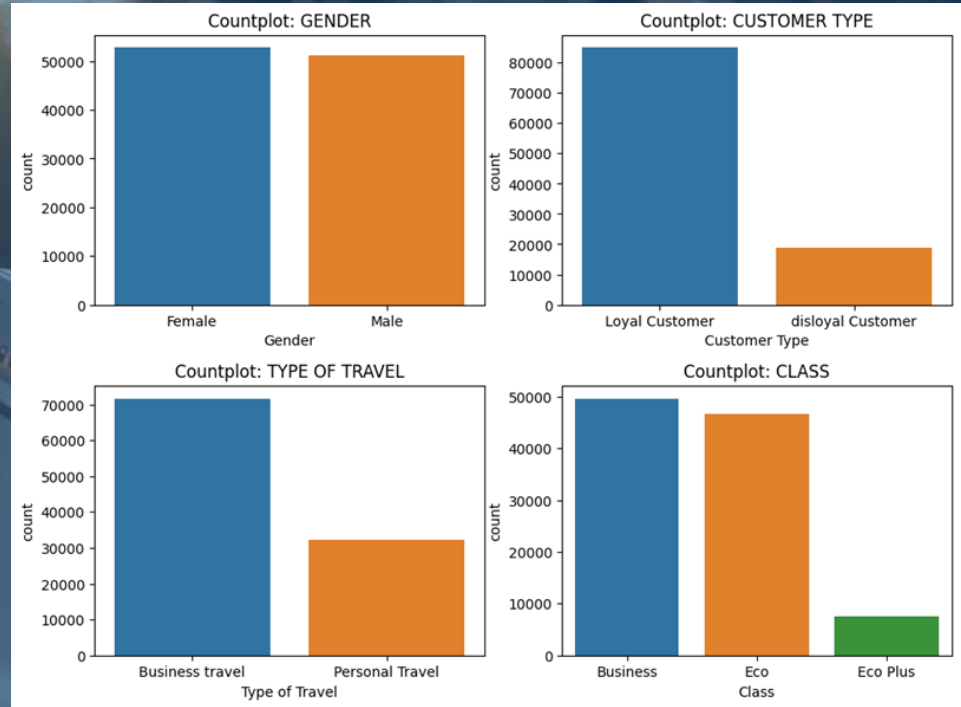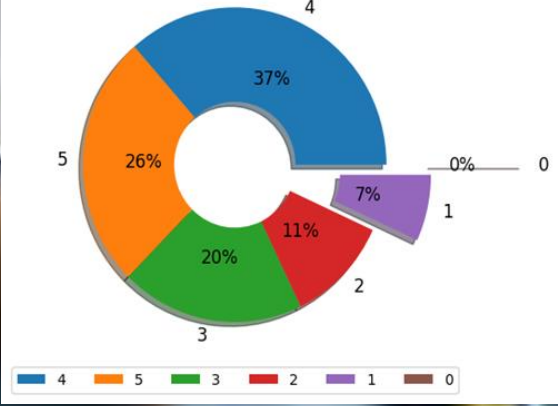
Bar charts for passenger satisfaction on Gender, Customer Type, Type of Travel & Class

Passengers Satisfaction and dissatisfaction depending on the categorical features.

Charts of some flight services based on passengers rating
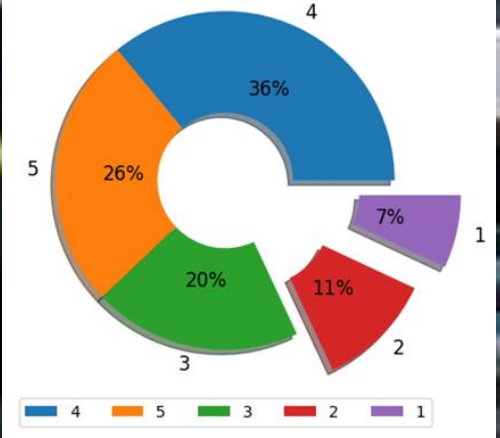
Passengers gave rating of different services provided by the airline. The rating is between 1-5.

For analysis I considered 1,2,3 rating as dissatisfied and 4,5 as satisfied.

Then plotted the features in pie chart.
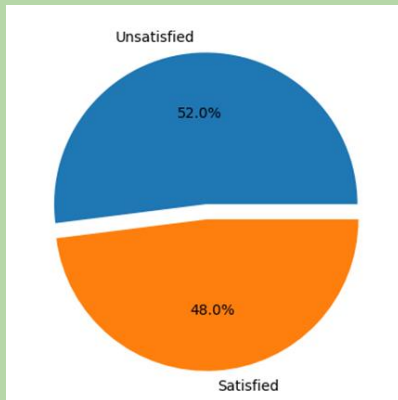

```python
z= ['Inflight wifi service', 'Departure/Arrival time convenient',
    'Ease of Online booking', 'Gate location', 'Food and drink',
    'Online boarding', 'Seat comfort', 'Inflight entertainment',
    'Leg room service', 'Baggage handling', 'Checkin service',
    'Inflight service', 'Cleanliness']


for i in z:
  x2 = df[i].value_counts()[1]
  x3 = df[i].value_counts()[2]
  x4 = df[i].value_counts()[3]
  x5 = df[i].value_counts()[4]
  x6 = df[i].value_counts()[5]
  Unsatisfied= x2+x3+x4
  Satisfied=x5+x6
  print(f'{i}_unsatisfied:',Unsatisfied)
  print(f'{i}_satisfied:',Satisfied)

  s=[Unsatisfied,Satisfied]
  explode = (0, 0.1)
  plt.pie(s,labels=["Unsatisfied","Satisfied"],autopct='%1.1f%%',explode=explode)
  plt.show()
  print("..............")
```
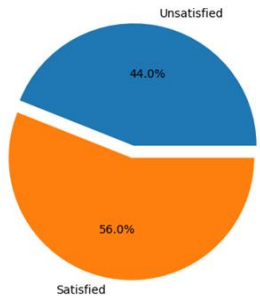
Seat comfort
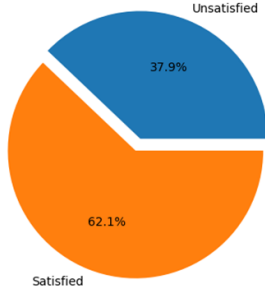
Baggage handling
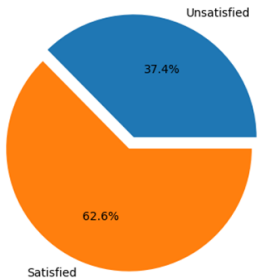
Inflight Service
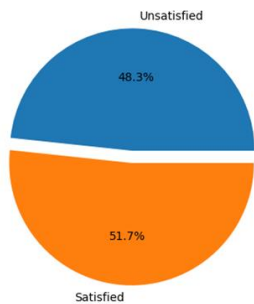
Leg Room

Satisfied % is less than dissatisfied %

Satisfied % is more than dissatisfied %
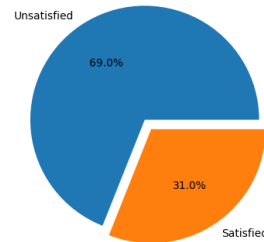
Inflight Wifi Service
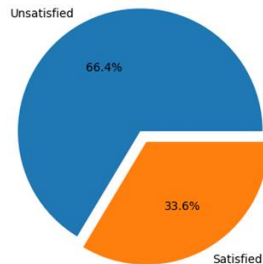
Online Ticket Booking

Gate Location

Food and Drink
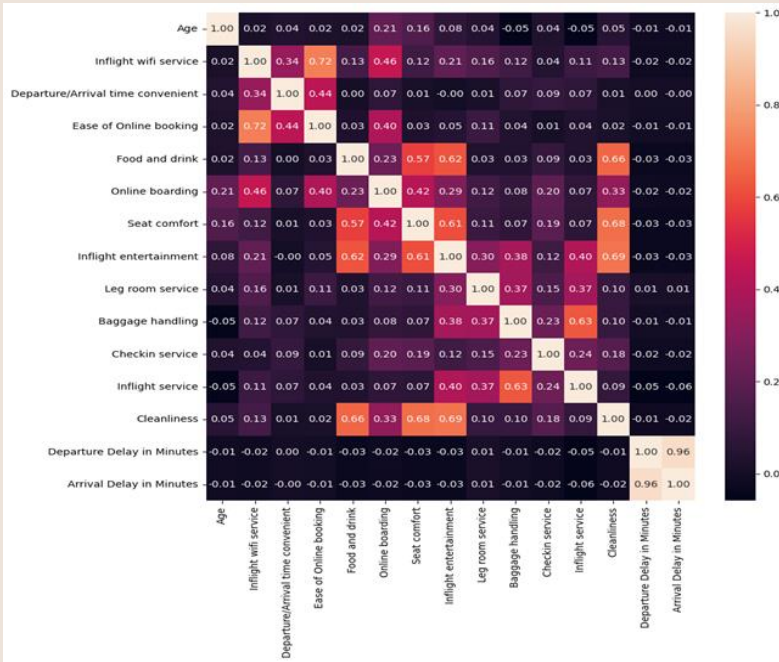
From the heat map we can see Departure delay in minutes and Arrival Delay in Minutes are highly correlated.

From the Scatter plot we can see with the increase of Departure delay in minutes, Arrival delay in minutes also increases.
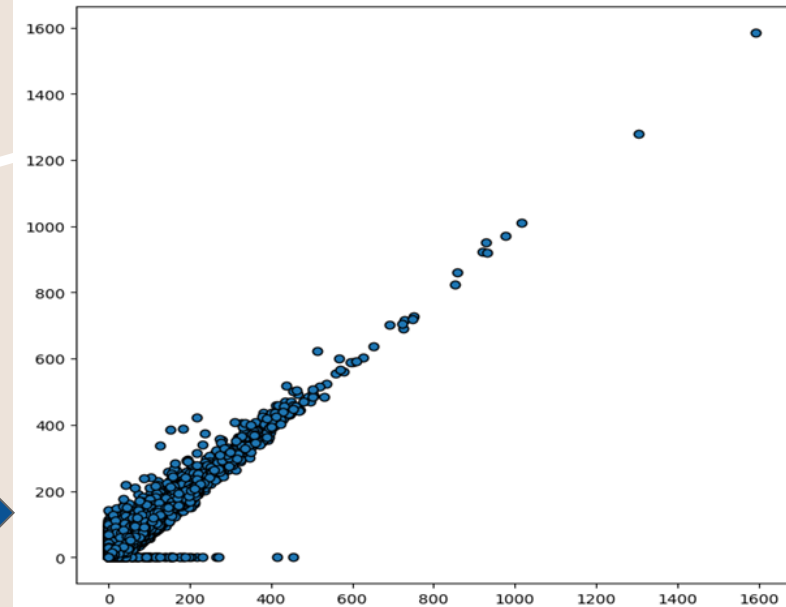
Model Building

```
feature: Gender
['Male', 'Female']
Categories (2, object): ['Female', 'Male']
[1 0]

feature: Customer Type
['Loyal Customer', 'disloyal Customer']
Categories (2, object): ['Loyal Customer', 'disloyal Customer']
[0 1]

feature: Type of Travel
['Personal Travel', 'Business travel']
Categories (2, object): ['Business travel', 'Personal Travel']
[1 0]

feature: Class
['Eco Plus', 'Business', 'Eco']
Categories (3, object): ['Business', 'Eco', 'Eco Plus']
[2 0 1]

feature: satisfaction
['neutral or dissatisfied', 'satisfied']
Categories (2, object): ['neutral or dissatisfied', 'satisfied']
[0 1]
```

Transforming the categorical variable in numerical variable

```python
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y,

print("X_train",len(X_train))
print("X_test",len(X_test))
print("Y_train",len(Y_train))
print("Y_test",len(Y_test))

X_train 72732
X_test 31172
Y_train 72732
Y_test 31172
```

Here 70% data are taken as training data and 30% data taken as test data.
And Satisfaction is the dependent variable, remaining features are independent variables.

Logistic Regression

```
#Logistic model fit
from sklearn.linear_model import LogisticRegression
log_model=LogisticRegression()
log_model.fit(X_train,Y_train)
```

Model Fit
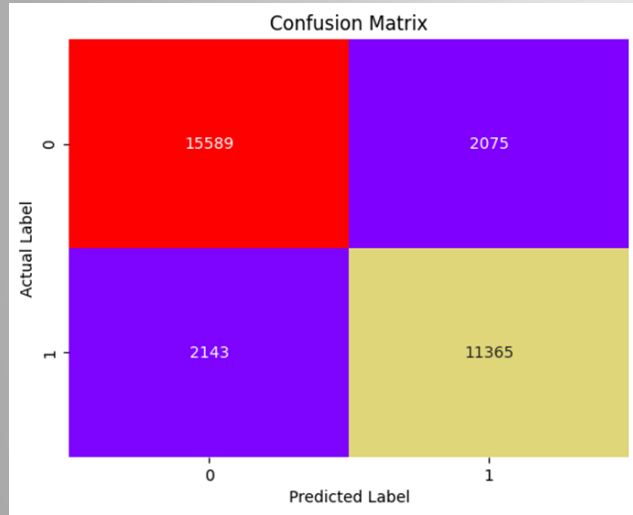
Confusion Matrix

Confusion Matrix

```
print(classification_report(Y_test,log_predict_y))

              precision    recall  f1-score   support

           0       0.88      0.88      0.88     17664
           1       0.85      0.84      0.84     13508

    accuracy                           0.86     31172
   macro avg       0.86      0.86      0.86     31172
weighted avg       0.86      0.86      0.86     31172
```

Accuracy Score : 0.8646862568972155

Decision Tree

Root Node

Sub-Tree

Decision Node

Decision Node

Decision Node

Leaf Node   Leaf Node   Leaf Node   Decision Node

Leaf Node   Leaf Node

Grid Search Method

```
param = {'criterion':['gini','entropy'],
         'max_depth':[20,22],
         'min_samples_split':[5,10],
         'min_samples_leaf':[3,5],
         'min_impurity_decrease':[0.0001,0.0002]}

grid_search_dt = GridSearchCV(estimator = dt_model, param_grid = param, cv = 10, n_jobs=-1,verbose=2)
grid_search_dt.fit(X_train,Y_train)
```

```
#Decision tree
from sklearn.tree import DecisionTreeClassifier
dt_model = DecisionTreeClassifier(criterion = 'gini', random_state=1)
dt_model.fit(X_train,Y_train)
```

```
          DecisionTreeClassifier
DecisionTreeClassifier(random_state=1)
```

```
dt_predict_y = dt_model.predict(X_test)
```

```
cm1 = confusion_matrix(Y_test, dt_predict_y)
cm1
```

```
array([[16777,   887],
       [  858, 12650]])
```

```
grid_search_dt.best_params_
```

```
{'criterion': 'entropy',
 'max_depth': 22,
 'min_impurity_decrease': 0.0001,
 'min_samples_leaf': 3,
 'min_samples_split': 5}
```

Accuracy Score : 0.95890542794

Accuracy Score : 0.9440202746054152

Random Forest

```
#RandomForest
from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(n_estimators = 100,max_features=6,random_state=0)
rf_model.fit(X_train, Y_train)

                    RandomForestClassifier
RandomForestClassifier(max_features=6, random_state=0)


rf_predict_y = rf_model.predict(X_test)


cm2 = confusion_matrix(Y_test, rf_predict_y)
cm2

array([[17318,   346],
       [  818, 12690]])
```
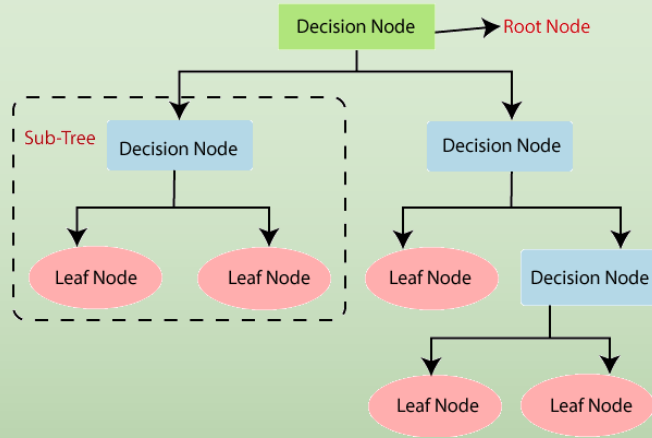
```
print(classification_report(Y_test,rf_predict_y))

              precision    recall  f1-score   support

           0       0.95      0.98      0.97     17664
           1       0.97      0.94      0.96     13508

    accuracy                           0.96     31172
   macro avg       0.96      0.96      0.96     31172
weighted avg       0.96      0.96      0.96     31172


rf_accuracy_score = accuracy_score(Y_test,rf_predict_y)
rf_accuracy_score
```

Accuracy Score : 0.9626587963.

Accuracy comparison for the models:

Models	Accuracy Score	Accuracy Percentage
Logistic Regression	0.8646862568972155	86.4%
Decision Tree	0.9440202746054152	94.4%
Decision Tree for best grid	0.9589054279481586	95.9%
Random Forest	0.9626587963557038	96.2%

Comparing all the accuracy score we can conclude that **Random Forest** is the best model compare to others.

Conclusion

In conclusion, airline passenger satisfaction is a multifaceted concept influenced by various factors.Meeting and exceeding passenger expectations is crucial for airlines to maintain a loyal customer.

1. People are depending on online services, online booking facility should be better.
2. Gate location from entry is another issue according to the passengers, airline company should also focus on that problem.
3. Airline company should develop the inflight wifi service and food & drink inflight.
4. Providing comfortable seating, sufficient legroom, and clean cabin conditions are essential for passenger comfort.

Here I built 3 models to classify the passenger's satisfaction and comparing the three model I conclude that Random Forest is the best model for this classification.

Limitation

1. Here I analyzed the satisfaction of the passengers depending on the ratings in various services of the airline. The problem is some passengers are satisfied with some services, and some are not satisfied with those services. So classifying satisfaction and dissatisfaction correctly depending on this is very difficult.

2. The data does not contain ticket price, which is a big part of airline services and customer satisfaction.

References

1. E Sezgen, KJ Mason, R Mayer -, 2019. Voice of Airline Passenger: A Text Mining Approach to Understand Customer Satisfaction. Journal of Air Transport Management. Volume 77, pp. 65-74.
2. S Tsafarakis, T Kokotas, A Pantouvakis – 2018. A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. Journal of Air Transport Management.
3. R Archana, MV Subha -2012. A STUDY ON SERVICE QUALITY AND PASSENGER SATISFACTION ON INDIAN AIRLINES. International Journal of Multidisciplinary Research Vol.2
4. Chang, Y. and Keller, M. R. (2002). "A survey analysis of service quality for domestic airlines", European Journal of Operational Research, Vol. 139, pp. 166-177.
5. Dennet, C.E., Ineson, M., Stone, G.J., and Colgate, M. (2000), "Pre-Bookable Services in the Chartered Airline Industry: Increasing Satisfaction through Differentiation", The Service Industries Journal, Vol. 20 (2), pp. 82-94.
6. Haynes, R., & Percy, J. (1994). "Perception paradox: Airline service quality issues", Proceedings of Decision Sciences Institute - Annual Meeting, Vol.3, pp. 1950-1952.
7. C Laming, K Mason – 2014. Customer Experience – an analysis of the concept and its performance in airline brands. Research in Transportation Business & Management. Volume 10, pp-15-25.
8. S Sanyal, MW Hisam – 2016, An Analysis of the Impact of Service Quality and Passenger Satisfaction on Passenger Preferences for Airlines: A Study of the Indian Aviation Sector. International Review of Management and Marketing, 2016, 6(2), 354-357.
9. R. Hussain, A. Al Nasser, and Y. K. Hussain, "Service quality and customer satisfaction of a UAE-based airline: An empirical investigation," J. Air Transp. Manag., vol. 42, pp. 167–175.
10. S. Tiernan, D. L. Rhoades, and B. Waguespack, "Airline service quality: Exploratory analysis of consumer perceptions and operational performance in the USA and EU," Manag. Serv. Qual. An Int. J., vol. 18, no. 3, pp. 212–224.

Thank You......