

RESULTS:

Importing the libraries:

```
[5] #Importing required lib

import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import RandomizedSearchCV
import imblearn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, f1_score
```

Read the Dataset:

```
[7] plt.style.use('fivethirtyeight')

Reading csv data
dataset = pd.read_csv("../content/flight.csv")
dataset.head()
```

	DAY_OF_MONTH	DAY_OF_WEEK	OP_UNIQUE_CARRIER	OP_CARRIER_AIRLINE_ID	OP_CARRIER	TAIL_NUM	OP_CARRIER_FL_NUM	ORIGIN_AIRPORT_ID	ORIGIN_AIRPORT_SEQ_ID	ORIGIN	...	DEST	DEP_TIME	DEP_DELTA	DEP_TIME_BLK	ARR_TIME	ARR_DELTA	CANCELLED	REVERSED	DISTANCE	Unnamed: 21
0	1	2	9E	20363	9E	N868DC	3280	11953	1195302	GNV	...	ATL	601.0	0.0	0600-0609	722.0	0.0	0.0	0.0	300.0	NaN
1	1	2	9E	20363	9E	N348PQ	3281	13487	1348702	MSP	...	CVG	1359.0	0.0	1400-1409	1633.0	0.0	0.0	0.0	596.0	NaN
2	1	2	9E	20363	9E	N889BA	3282	11433	1143302	DTW	...	CVG	1216.0	0.0	1200-1209	1329.0	0.0	0.0	0.0	229.0	NaN
3	1	2	9E	20363	9E	N888BA	3283	15249	1524906	TLH	...	ATL	1521.0	0.0	1500-1509	1625.0	0.0	0.0	0.0	223.0	NaN
4	1	2	9E	20363	9E	N807AC	3284	10397	1039707	ATL	...	FSM	1647.0	0.0	1900-1909	1940.0	0.0	0.0	0.0	579.0	NaN

5 rows x 22 columns

Handling missing values:

```
# Checking data type
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 362354 entries, 0 to 362353
Data columns (total 22 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   DAY_OF_MONTH                          362354 non-null  int64
1   DAY_OF_WEEK                           362354 non-null  int64
2   OP_UNIQUE_CARRIER                    362354 non-null  object
3   OP_CARRIER_AIRLINE_ID                 362354 non-null  int64
4   OP_CARRIER                            362354 non-null  object
5   TAIL_NUM                               361670 non-null  object
6   OP_CARRIER_FL_NUM                     362354 non-null  int64
7   ORIGIN_AIRPORT_ID                      362353 non-null  float64
8   ORIGIN_AIRPORT_SEQ_ID                  362353 non-null  float64
9   ORIGIN                                  362353 non-null  object
10  DEST_AIRPORT_ID                        362353 non-null  float64
11  DEST_AIRPORT_SEQ_ID                    362353 non-null  float64
12  DEST                                    362353 non-null  object
13  DEP_TIME                               356549 non-null  float64
14  DEP_DEL15                             356548 non-null  float64
15  DEP_TIME_BLK                           362353 non-null  object
16  ARR_TIME                               356159 non-null  float64
17  ARR_DEL15                              355606 non-null  float64
18  CANCELLED                              362353 non-null  float64
19  DIVERTED                               362353 non-null  float64
20  DISTANCE                               362353 non-null  float64
21  Unnamed: 21                            0 non-null      float64
dtypes: float64(12), int64(4), object(6)
memory usage: 60.8+ MB
```

```
[16]
dataset = dataset.drop('Unnamed: 21', axis=1)
dataset.isnull().sum()

DAY_OF_MONTH          0
DAY_OF_WEEK            0
OP_UNIQUE_CARRIER    0
OP_CARRIER_AIRLINE_ID 0
OP_CARRIER            0
TAIL_NUM              684
OP_CARRIER_FL_NUM    0
ORIGIN_AIRPORT_ID      1
ORIGIN_AIRPORT_SEQ_ID  1
ORIGIN                  1
DEST_AIRPORT_ID        1
DEST_AIRPORT_SEQ_ID    1
DEST                    1
DEP_TIME               5805
DEP_DEL15              5806
DEP_TIME_BLK           1
ARR_TIME               6195
ARR_DEL15              6748
CANCELLED              1
DIVERTED               1
DISTANCE               1
dtype: int64
```

```
dataset = dataset[['DAY_OF_MONTH', 'DAY_OF_WEEK', 'DISTANCE', 'ARR_TIME', 'DEP_TIME', 'ORIGIN_AIRPORT_SEQ_ID', 'ORIGIN_AIRPORT_ID', 'OP_CARRIER_FL_NUM', 'OP_CARRIER_AIRLINE_ID', 'CANCELLED', 'DIVERTED', 'ARR_DEL15', 'DEP_DEL15']]
dataset.isnull().sum()

DAY_OF_MONTH      0
DAY_OF_WEEK       0
OP_UNIQUE_CARRIER 0
OP_CARRIER_AIRLINE_ID 0
OP_CARRIER        0
TAIL_NUM          684
OP_CARRIER_FL_NUM 0
ORIGIN_AIRPORT_ID  1
ORIGIN_AIRPORT_SEQ_ID 1
ORIGIN             1
DEST_AIRPORT_ID    1
DEST_AIRPORT_SEQ_ID 1
DEST              1
DEP_TIME           5885
DEP_DEL15          5886
DEP_TIME_BLK       1
ARR_TIME           6195
ARR_DEL15          6748
CANCELLED          1
DIVERTED           1
DISTANCE           1
dtype: int64
```

```
dataset[dataset.isnull().any(axis=1)].head(10)

DAY_OF_MONTH  DAY_OF_WEEK  OP_UNIQUE_CARRIER  OP_CARRIER_AIRLINE_ID  OP_CARRIER  TAIL_NUM  OP_CARRIER_FL_NUM  ORIGIN_AIRPORT_ID  ORIGIN_AIRPORT_SEQ_ID  ORIGIN  ...  DEST_AIRPORT_SEQ_ID  DEST  DEP_TIME  DEP_DEL15  DEP_TIME_BLK  ARR_TIME  ARR_DEL15  CANCELLED  DIVERTED  DISTANCE
387           1           2           AA           19805           AA           N966NN           178           14027.0           1402702.0  ...  1363007.0  ORD  NaN      NaN      0600-0659      NaN      NaN      1.0      0.0      1143.0
403           1           2           AA           19805           AA           N901SD           189           11503.0           1103036.0  ...  1129806.0  DFW  NaN      NaN      0700-0759      NaN      NaN      1.0      0.0      721.0
871           1           2           OH           20387           OH           N672NN           5103          11057.0           1105703.0  ...  1419306.0  PHS  1993.0      0.0      1990-1999      NaN      NaN      1.0      0.0      488.0
1260          1           2           B6           20409           B6           N613JB           368           19800.0           1980003.0  ...  1247805.0  JFK  2203.0      0.0      2100-2159      709.0      NaN      0.0      1.0      2485.0
1833          1           2           B6           20409           B6           N184JB           1210          12451.0           1245102.0  ...  1072102.0  BOS  1201.0      0.0      1200-1259      1810.0      NaN      0.0      1.0      1810.0
1888          1           2           B6           20409           B6           N627JB           2338          19800.0           1980003.0  ...  1072102.0  BOS  2040.0      1.0      2000-2059      610.0      NaN      0.0      1.0      2601.0
1882          1           2           B6           20409           B6           N70LJB           2358          19800.0           1980003.0  ...  1247805.0  JFK  1500.0      1.0      1400-1459      43.0      NaN      0.0      1.0      2485.0
1878          1           2           B6           20409           B6           N699JB           2451          10721.0           1072102.0  ...  1320402.0  MCO  934.0      0.0      0900-0959      1456.0      NaN      0.0      1.0      1121.0
1873          1           2           EV           20366           EV           N17984           4187          12448.0           1244807.0  ...  1225603.0  IAH  NaN      NaN      0900-0959      NaN      NaN      1.0      0.0      351.0
1876          1           2           EV           20366           EV           N17984           4189          12068.0           1228603.0  ...  1244807.0  JAN  NaN      NaN      0700-0759      NaN      NaN      1.0      0.0      351.0
10 rows x 21 columns

[11] dataset['OP_CARRIER_AIRLINE_ID'].mode()

0      58193
Name: OP_CARRIER_AIRLINE_ID, dtype: int64

[12] Import math
for index, row in dataset.iterrows():
    dataset.loc[index, 'OP_CARRIER_AIRLINE_ID'] = math.floor(row['OP_CARRIER_AIRLINE_ID']/200)
dataset.head()
```

Handling Categorical Values:

```
[22] from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
dataset['OP_CARRIER_AIRLINE_ID'] = le.fit_transform(dataset['OP_CARRIER_FL_NUM'])
dataset['OP_CARRIER_AIRLINE_ID'] = le.fit_transform(dataset['OP_CARRIER_AIRLINE_ID'])

[23] dataset.head(5)

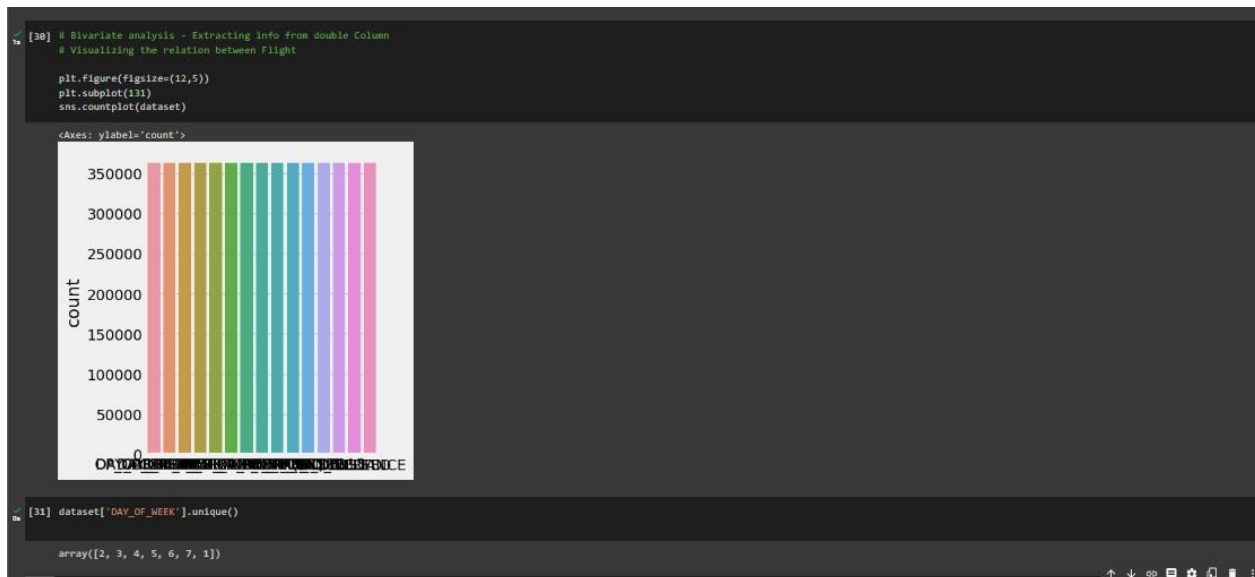
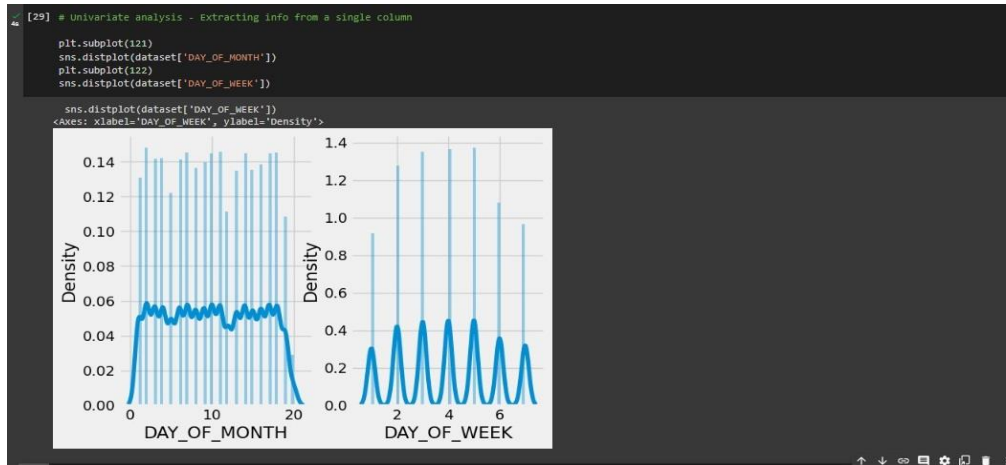
DAY_OF_MONTH  DAY_OF_WEEK  OP_UNIQUE_CARRIER  OP_CARRIER_AIRLINE_ID  OP_CARRIER  TAIL_NUM  OP_CARRIER_FL_NUM  ORIGIN_AIRPORT_ID  ORIGIN_AIRPORT_SEQ_ID  ORIGIN  ...  DEST_AIRPORT_SEQ_ID  DEST  DEP_TIME  DEP_DEL15  DEP_TIME_BLK  ARR_TIME  ARR_DEL15  CANCELLED  DIVERTED  DISTANCE
0           1           2           9E           3279           9E           N888BC           3280          11953.0           1195302.0  ...  1039707.0  ATL  001.0      0.0      0900-0959      722.0      0.0      0.0      0.0      300.0
1           1           2           9E           3279           9E           N348PC           3281          13487.0           1348702.0  ...  1119302.0  CVG  1359.0      0.0      1400-1459      1633.0      0.0      0.0      0.0      596.0
2           1           2           9E           3277           9E           N888GA           3282          11433.0           1143302.0  ...  1119302.0  CVG  1215.0      0.0      1200-1259      1329.0      0.0      0.0      0.0      229.0
3           1           2           9E           3278           9E           N888GA           3283          15249.0           1524906.0  ...  1039707.0  ATL  1521.0      0.0      1500-1559      1626.0      0.0      0.0      0.0      223.0
4           1           2           9E           3279           9E           N887AC           3284          10397.0           1039707.0  ...  1177801.0  FBM  1847.0      0.0      1900-1959      1940.0      0.0      0.0      0.0      578.0
5 rows x 21 columns
```

```
[24] dataset['ORIGIN'].unique()

array(['GNV', 'MSP', 'DTW', 'TLH', 'ATL', 'DAY', 'JAN', 'LGA', 'JAX',
       'BWI', 'CAK', 'PHL', 'JFK', 'AGS', 'LIT', 'IND', 'EWR', 'CAE',
       'OMA', 'TRI', 'BOS', 'MGM', 'MDY', 'RKE', 'TVC', 'FSM', 'BHM',
       'PIA', 'CVG', 'GTR', 'SDF', 'BNA', 'SAT', 'LFT', 'MSN', 'DSN',
       'EVR', 'ABE', 'DCA', 'BWI', 'ILM', 'TYS', 'LEX', 'CLE', 'ELM',
       'EMM', 'PMF', 'SHV', 'CSG', 'TUL', 'BDL', 'DFW', 'RIC', 'RAP',
       'PIT', 'CRW', 'GPT', 'CHA', 'OMA', 'IAD', 'RDU', 'SAV', 'GRR',
       'BTU', 'LAN', 'OAJ', 'AEX', 'CHO', 'CHS', 'HRL', 'ORF', 'HOT',
       'HSV', 'MCO', 'BTR', 'LAX', 'ORD', 'PHX', 'MLA', 'CHH', 'SFO',
       'SEA', 'STL', 'CLT', 'LAS', 'OAK', 'BUR', 'FLL', 'SJO', 'ALB',
       'PMH', 'SMF', 'MHT', 'TPA', 'EGE', 'OGG', 'HNL', 'MSY', 'IAH',
       'KOA', 'DEN', 'SAN', 'TUS', 'SJC', 'PBI', 'SNA', 'GRI', 'SLC',
       'FAT', 'ENR', 'JAC', 'MCI', 'PMS', 'ELP', 'AUS', 'ABQ', 'PSP',
       'MEM', 'POX', 'ACT', 'ABI', 'DRT', 'GSO', 'BIL', 'AVP', 'CLL',
       'CID', 'CRP', 'GRK', 'AZO', 'TYR', 'FAR', 'GSP', 'MFE', 'AMA',
       'SMD', 'MTJ', 'LAM', 'XNA', 'CIT', 'MKE', 'SGF', 'AVL', 'AZA',
       'PBD', 'SFB', 'SRO', 'PIE', 'ILV', 'PBG', 'BFI', 'USA', 'SBN',
       'BGR', 'FNT', 'IAO', 'IDA', 'BIS', 'GFM', 'PVD', 'MLB', 'TOL',
       'MYR', 'HOB', 'HPH', 'FAY', 'SYR', 'ART', 'HVN', 'LGB', 'BND',
       'RSW', 'ROC', 'HOU', 'STX', 'ONT', 'BUR', 'SMF', 'DAB', 'BQN',
       'PSE', 'ORH', 'STT', 'VPS', 'ICT', 'OKC', 'LBB', 'LRO', 'BRO',
       'HOB', 'SBA', 'TTN', 'COS', 'PSM', 'ISP', 'BZN', 'LIT', 'ITO',
       'ECP', 'HHH', 'LBE', 'ACY', 'FLG', 'ASE', 'HDN', 'ROM', 'YUM',
       'GRB', 'GJT', 'FSD', 'FWA', 'ROM', 'ATW', 'BFL', 'EUG', 'SAF',
       'HFR', 'JLN', 'SGJ', 'COS', 'HEI', 'PIB', 'ROT', 'DAL', 'STS',
       'SRP', 'HRV', 'GEG', 'RST', 'VLD', 'HRS', 'SUN', 'PSC', 'MLI',
       'FCA', 'DLH', 'ERI', 'DWN', 'ABY', 'BQK', 'MDW', 'GTF', 'LSE',
       'ESC', 'CPR', 'SCE', 'HLN', 'LWK', 'TSM', 'MLU', 'MSO', 'ROA',
       'CRK', 'EAU', 'LMB', 'SHD', 'MKG', 'HVS', 'SLN', 'EAR', 'UIN',
       'CKB', 'RKS', 'PUB', 'PAH', 'CGI', 'VEL', 'CNY', 'LBL', 'BFF',
       'SPI', 'DVL', 'JMS', 'LAR', 'PRC', 'GCC', 'LBF', 'ACV', 'RDD',
       'DTH', 'MMH', 'MAF', 'GBO', 'LMS', 'PBG', 'MKT', 'ABR', 'APN',
       'PLN', 'BTT', 'BRD', 'BTH', 'CDC', 'CIU', 'EKO', 'JRT', 'TWF',
       'HIB', 'RHI', 'BGR', 'ITH', 'JNL', 'STJ', 'GGG', 'SUX', 'TXK',
       'ALO', 'DBQ', 'ADQ', 'ANC', 'BET', 'BRM', 'SCC', 'FAT', 'KTN',
       'JNU', 'SIT', 'WRG', 'PSG', 'OME', 'OTZ', 'LCH', 'LCK', 'HTS',
       'GUC', 'SPS', 'COU', 'BPT', 'ORD', 'GUM', 'SPN', 'SCK', 'PVU',
       'RFD', 'GCK', 'STC', 'SMK', 'YAK', 'CDV', 'ADK', 'HGR', 'OWB',
       'OGS', 'PPG', 'OGD', 'LYH', nan], dtype=object)
```

```
[25]
# creating dummy dataframe for categorical values
dataset_cat = dataset.select_dtypes(include='object')
dataset_cat.head()
```

	OP_UNIQUE_CARRIER	OP_CARRIER	TAIL_NUM	ORIGIN	DEST	DEP_TIME_BLK
0	9E	9E	N8888C	GNV	ATL	0800-0859
1	9E	9E	N348PQ	MSP	CVG	1400-1459
2	9E	9E	N890QA	DTW	CVG	1200-1259
3	9E	9E	N8888A	TLH	ATL	1500-1559
4	9E	9E	N8974C	ATL	FSM	1900-1959



```
[12] dataset = pd.get_dummies(dataset, columns=['DAY_OF_MONTH', 'DAY_OF_WEEK'])
dataset.head()

OP_UNIQUE_CARRIER  OP_CARRIER_AIRLINE_ID  OP_CARRIER_TAIL_NUM  OP_CARRIER_FL_NUM  ORIGIN_AIRPORT_ID  ORIGIN_AIRPORT_SEQ_ID  ORIGIN_CITY_STATE_AIRPORT_ID  DEST_AIRPORT_ID  DEST_AIRPORT_SEQ_ID  ...  DAY_OF_MONTH_18  DAY_OF_MONTH_19  DAY_OF_MONTH_20  DAY_OF_WEEK_1  DAY_OF_WEEK_2  DAY_OF_WEEK_3  DAY_OF_WEEK_4  DAY_OF_WEEK_5  DAY_OF_WEEK_6  DAY_OF_WEEK_7
0           DL              3275           DL  N0882C              2280              11935.0              1195302.0  DMV              10397.0              1039707.0  ...              0              0              0              0              1              0              0              0              0              0
1           DL              3276           DL  N1430PC              2281              10447.0              1044702.0  MSP              111602.0              111602.0  ...              0              0              0              0              1              0              0              0              0              0
2           DL              3277           DL  N1808A              2282              11432.0              1143202.0  DTW              11199.0              1119902.0  ...              0              0              0              0              1              0              0              0              0              0
3           DL              3278           DL  N1808A              2283              10340.0              1034002.0  TLH              10397.0              1039707.0  ...              0              0              0              0              1              0              0              0              0              0
4           DL              3279           DL  N1807AC              2284              10397.0              1039707.0  ATL              10778.0              1077801.0  ...              0              0              0              0              1              0              0              0              0              0

0 rows x 40 columns

[13] x = dataset.iloc[:, 0:8].values
y = dataset.iloc[:, 8:9].values

[14] x
array([[ 'DL', 3275, 'DL', ..., 11935.0, 1195302.0, 'DMV'],
       [ 'DL', 3276, 'DL', ..., 10447.0, 1044702.0, 'MSP'],
       [ 'DL', 3277, 'DL', ..., 11432.0, 1143202.0, 'DTW'],
       ...,
       [ 'DL', 1893, 'DL', ..., 14679.0, 1467901.0, 'SAN'],
       [ 'DL', 1894, 'DL', ..., 14771.0, 1477104.0, 'SFO'],
       [ 'DL', 1895, 'DL', ..., NaN, NaN, NaN]], dtype=object)

[15] from sklearn.preprocessing import OneHotEncoder
oh = OneHotEncoder()
z=oh.fit_transform(x[:,4:5]).toarray()
t=oh.fit_transform(x[:,5:6]).toarray()
x=np.delete(x,[4,7],axis=1)
```

```
[35] from sklearn.preprocessing import OneHotEncoder
oh = OneHotEncoder()
z=oh.fit_transform(x[:,4:5]).toarray()
t=oh.fit_transform(x[:,5:6]).toarray()
#x=np.delete(x,[4,7],axis=1)

[36] z
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])

[37] t
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 1.]])

[38] x=np.delete(x,[4,5],axis=1)
```

Exploratory Data Analysis:

Descriptive statistical:

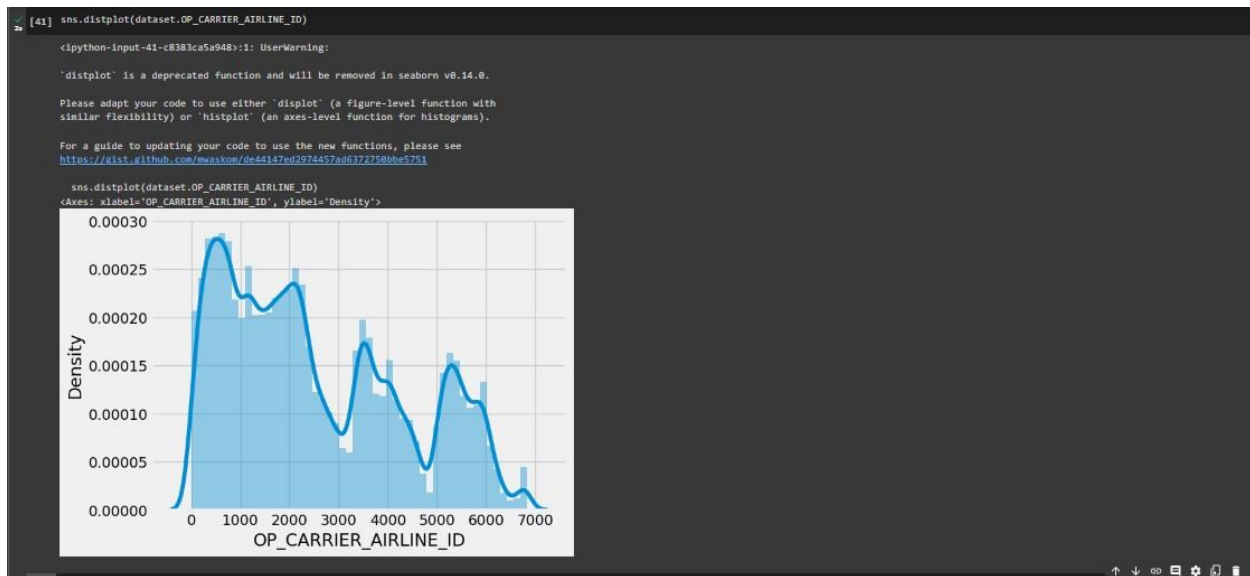
```
[16] dataset.describe()

OP_UNIQUE_CARRIER  OP_CARRIER_AIRLINE_ID  OP_CARRIER_TAIL_NUM  OP_CARRIER_FL_NUM  ORIGIN_AIRPORT_ID  ORIGIN_AIRPORT_SEQ_ID  ORIGIN_CITY_STATE_AIRPORT_ID  DEST_AIRPORT_ID  DEST_AIRPORT_SEQ_ID  ...  DAY_OF_MONTH_18  DAY_OF_MONTH_19  DAY_OF_MONTH_20  DAY_OF_WEEK_1  DAY_OF_WEEK_2  DAY_OF_WEEK_3  DAY_OF_WEEK_4  DAY_OF_WEEK_5  DAY_OF_WEEK_6  DAY_OF_WEEK_7
count  36254.000000  36254.000000  36254.000000  36254.000000  3.62539e+02  3.62539e+02  36254.000000  36254.000000  36254.000000  ...  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000  36254.000000
mean    2545.211222    3022.070941    12661.483779    1.38823e+01    13882.429583    1.38824e+01    1388.384522    0.127232    4165.226877    0.107888  ...    0.022148    0.011117    0.011686    0.110286    0.127420    0.142618    0.146466    0.144637    0.138465    0.115841
std     1871.632674    1627.391483    10716.483682    1.51946e+02    496.728294    1.51922e+02    496.728294    0.361091    522.208294    0.373287  ...    0.338468    0.196791    0.194734    0.310234    0.308366    0.306466    0.279266    0.275825    0.330105    0.320202
min      0.000000      1.000000    10135.000000    1.01305e+01    10135.000000    1.01305e+01    1.000000    0.000000    1.000000    0.000000  ...    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%     1613.000000    1871.000000    11321.000000    1.13210e+01    11321.000000    1.13210e+01    1136.000000    0.000000    4163.000000    0.000000  ...    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
50%     2114.000000    2122.000000    12661.000000    1.38863e+01    12661.000000    1.38863e+01    1388.000000    0.000000    4163.000000    0.000000  ...    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
75%     2624.000000    2626.000000    14227.000000    1.42270e+01    14227.000000    1.42270e+01    1427.000000    0.000000    4163.000000    0.000000  ...    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
max     4854.000000    7636.000000    40214.000000    1.40214e+02    40214.000000    1.40214e+02    4048.000000    1.000000    3402.000000    1.000000  ...    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000    1.000000

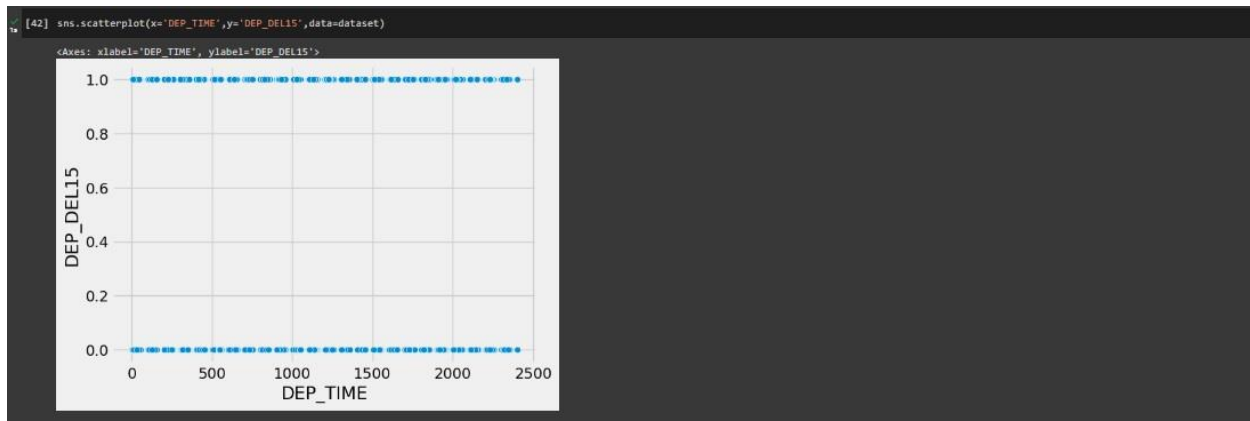
0 rows x 40 columns
```

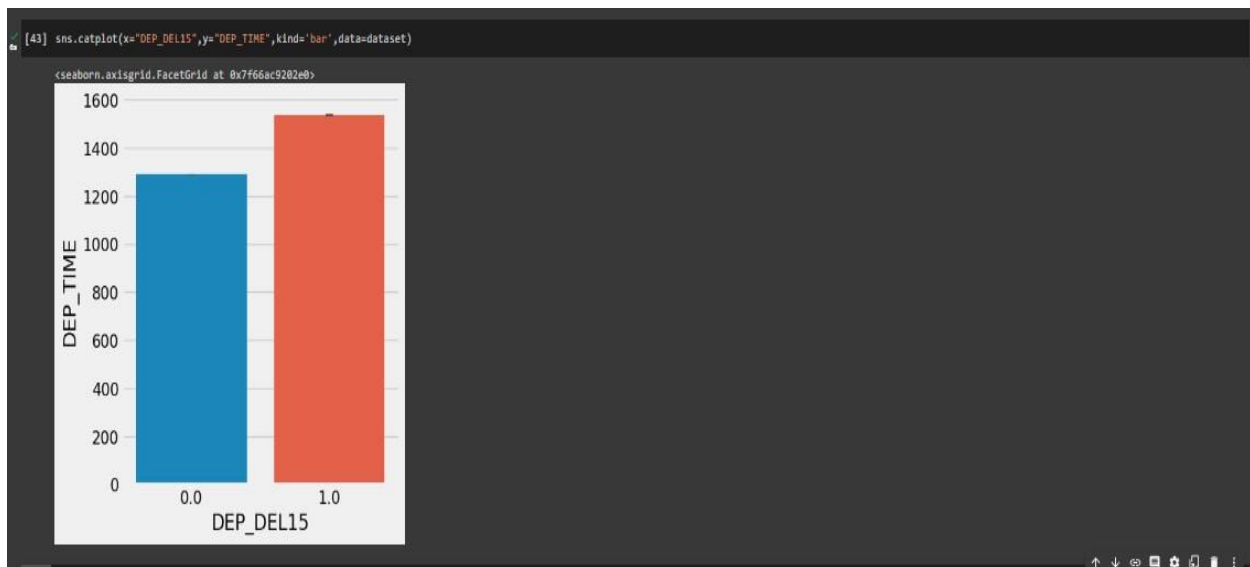
Visual analysis:

Univariate analysis:

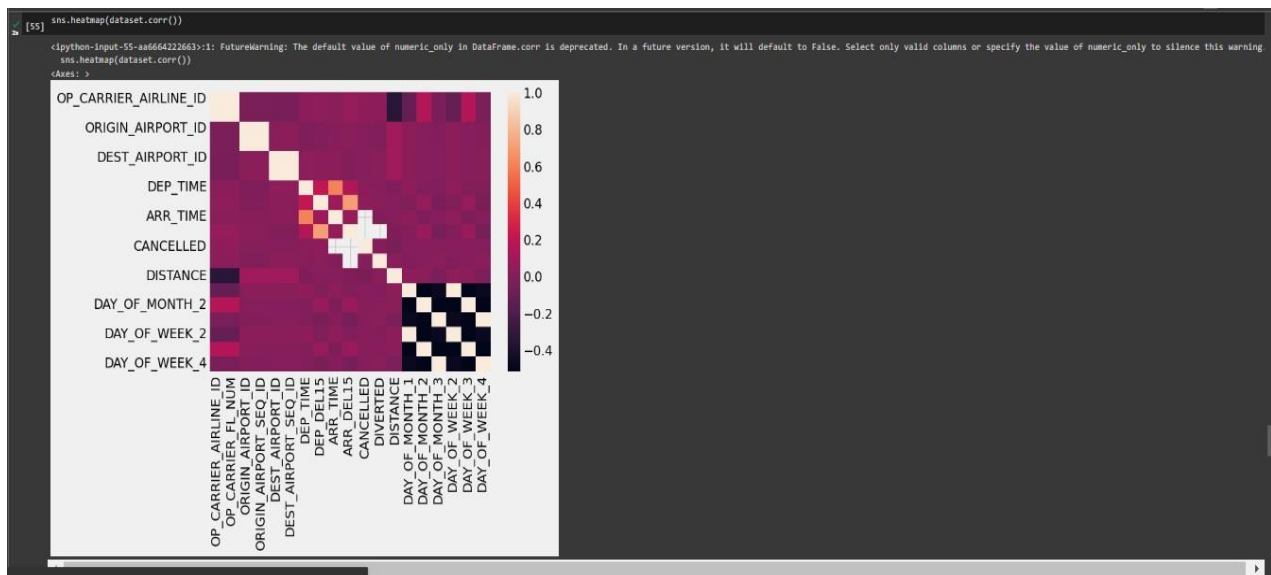


Bivariate analysis:





Multivariate analysis:



Splitting data into train and test:


```
[46] dataset = dataset.drop(columns=['OP_CARRIER_DELAY_15', 'OP_CARRIER_DELAY_30'])
dataset.head()
```

	OP_NASDAQ_CARRIER	OP_CARRIER	TAIL_NUM	OP_CARRIER_15_MIN	OP_CARRIER_30_MIN	ORIGIN_AIRPORT_SEQ_ID	ORIGIN	DEST_AIRPORT_SEQ_ID	DEST	...	OP_CARRIER_DELAY_15_MIN	OP_CARRIER_DELAY_30_MIN	OP_CARRIER_DELAY_15_MIN	OP_CARRIER_DELAY_30_MIN	OP_CARRIER_DELAY_15_MIN	OP_CARRIER_DELAY_30_MIN	OP_CARRIER_DELAY_15_MIN	OP_CARRIER_DELAY_30_MIN
0	06	06	1603AC	3300	11402	116333	DFW	10307	10307	ATL	...	0	0	0	0	0	0	0
1	06	06	1603PO	3301	11407	116333	MSP	11602	11602	DFW	...	0	0	0	0	0	0	0
2	06	06	1603AA	3302	11402	116333	DFW	11602	11602	DFW	...	0	0	0	0	0	0	0
3	06	06	1603AA	3303	11406	116333	TXL	10307	10307	ATL	...	0	0	0	0	0	0	0
4	06	06	1607AC	3304	11307	116333	ATL	11701	11701	PWM	...	0	0	0	0	0	0	0

Index: 15002 columns

Model Building:

```
[57] x = dataset.iloc[:, 0:8].values
y = dataset.iloc[:, 8:9].values
```

```
[58] from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)
```

```
[59] from sklearn.model_selection import train_test_split
train_x, test_x, train_y, test_y = train_test_split(dataset.drop('ARR_DELAY_15', axis=1), dataset['ARR_DELAY_15'], test_size=0.2, random_state=0)
```

```
x_test.shape
```

(11305, 8)

```
[61] x_train.shape
```

(45217, 8)

```
[62] y_test.shape
```

(11305, 1)

```
[63] y_train.shape
```

(45217, 1)