

DEVELOPING A FLIGHT DELAY MODEL USING MACHINE LEARNING

TEAM ID: PNT2022TMID27775

Dropping Un Necessary Columns

- Remove the column unnamed from the dataset.

```
dataset = dataset.drop('Unnamed: 25', axis=1)
dataset.isnull().sum()
```

```
YEAR                0
QUARTER             0
MONTH              0
DAY_OF_MONTH        0
DAY_OF_WEEK         0
UNIQUE_CARRIER     0
TAIL_NUM            0
FL_NUM             0
ORIGIN_AIRPORT_ID   0
ORIGIN              0
DEST_AIRPORT_ID     0
DEST               0
CRS_DEP_TIME        0
DEP_TIME           107
DEP_DELAY           107
DEP_DEL15           107
CRS_ARR_TIME        0
ARR_TIME           115
ARR_DELAY           188
ARR_DEL15           188
CANCELLED           0
DIVERTED            0
CRS_ELAPSED_TIME    0
ACTUAL_ELAPSED_TIME 188
DISTANCE            0
dtype: int64
```

- Filter the dataset to eliminate columns that aren't relevant to a predictive model.

```
dataset = dataset[["FL_NUM", "MONTH", "DAY_OF_MONTH", "DAY_OF_WEEK", "ORIGIN", "DEST", "CRS_ARR_TIME", "DEP_DEL15", "ARR_DEL15"]]
dataset.isnull().sum()

FL_NUM      0
MONTH       0
DAY_OF_MONTH 0
DAY_OF_WEEK 0
ORIGIN      0
DEST        0
CRS_ARR_TIME 0
DEP_DEL15   107
ARR_DEL15   188
dtype: int64
```

- Replace the missing value 0's & 1's

```
dataset = dataset.fillna({'ARR_DEL15': 1})
dataset = dataset.fillna({'DEP_DEL15': 0})
dataset.iloc[177:185]
```

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
177	2834	1	9	6	MSP	SEA	852	0.0	1.0
178	2839	1	9	6	DTW	JFK	1724	0.0	0.0
179	86	1	10	7	MSP	DTW	1632	0.0	1.0
180	87	1	10	7	DTW	MSP	1649	1.0	0.0
181	423	1	10	7	JFK	ATL	1600	0.0	0.0
182	440	1	10	7	JFK	ATL	849	0.0	0.0
183	485	1	10	7	JFK	SEA	1945	1.0	0.0
184	557	1	10	7	MSP	DTW	912	0.0	1.0

- The math.floor() method rounds a number DOWN to the nearest integer, if necessary, and returns the result.

```
import math

for index, row in dataset.iterrows():
    dataset.loc[index, 'CRS_ARR_TIME'] = math.floor(row['CRS_ARR_TIME'] / 100)
dataset.head()
```

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
0	1399	1	1	5	ATL	SEA	21	0.0	0.0
1	1476	1	1	5	DTW	MSP	14	0.0	0.0
2	1597	1	1	5	ATL	SEA	12	0.0	0.0
3	1768	1	1	5	SEA	MSP	13	0.0	0.0
4	1823	1	1	5	SEA	DTW	6	0.0	0.0