DEVELOPING A FLIGHT DELAY MODEL USING MACHINE LEARNING

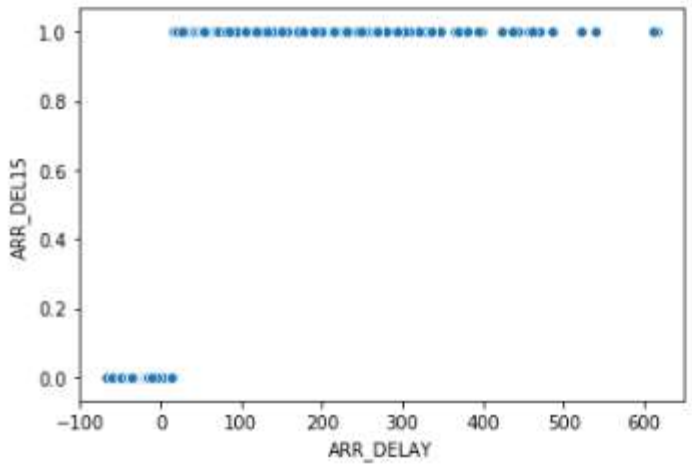
TEAM ID: PNT2022TMID27775

Data Visualization

- Data visualization is where a given data set is presented in a graphical format. It helps the detection of patterns, trends and correlations that might go undetected in text-based data.Understanding your data and the relationship present within it is just as important as any algorithm used to train your machine learning model. In fact, even the most sophisticated machine learning models will perform poorly on data that wasn't visualized and understood properly.
- To visualize the dataset we need libraries called Matplotlib and Seaborn.The Matplotlib library is a Python 2D plotting library which allows you to generate plots, scatter plots, histograms, bar charts etc.
- Let's visualize our data using Matplotlib and Seaborn library.

```
sns.scatterplot(x='ARR_DELAY',y='ARR_DEL15',data=flight_data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x2857c574208>
```

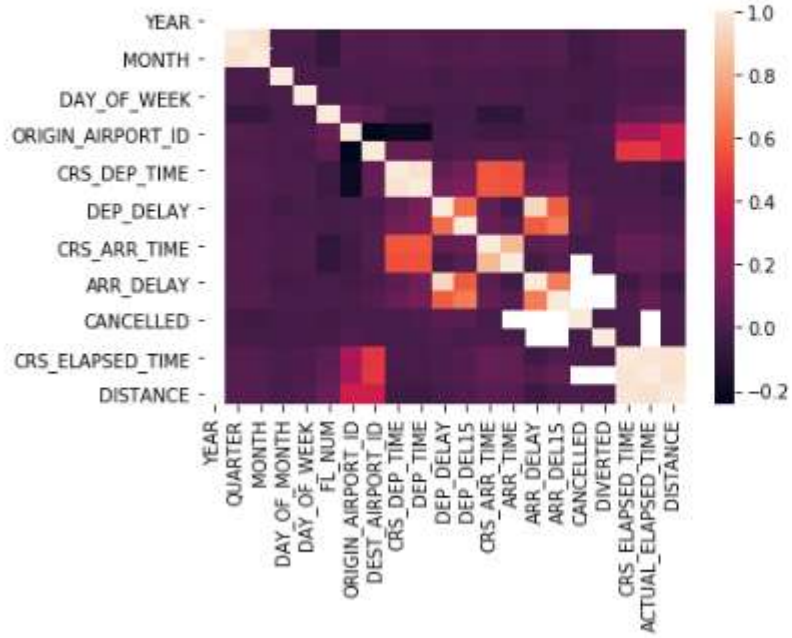


- Plot the Cat Plot between ARR_DELAY& ARR_DEL15.
- Correlation is a statistical relationship between two variables and it could be positive, meaning both variables move in the same direction, or negative, meaning that when one variable's value increases, the other variables' values decrease.
- With the help of seaborn heatmap we will be plotting the heatmap and for finding the correlation between variables we have corr() available

- If you observe the heatmap, lighter the colour the correlation between that two variables will be high.

```
sns.heatmap(dataset.corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x2857a89a7c8>
```



- And correlation plays a very important role for extracting the correct features for building our model.