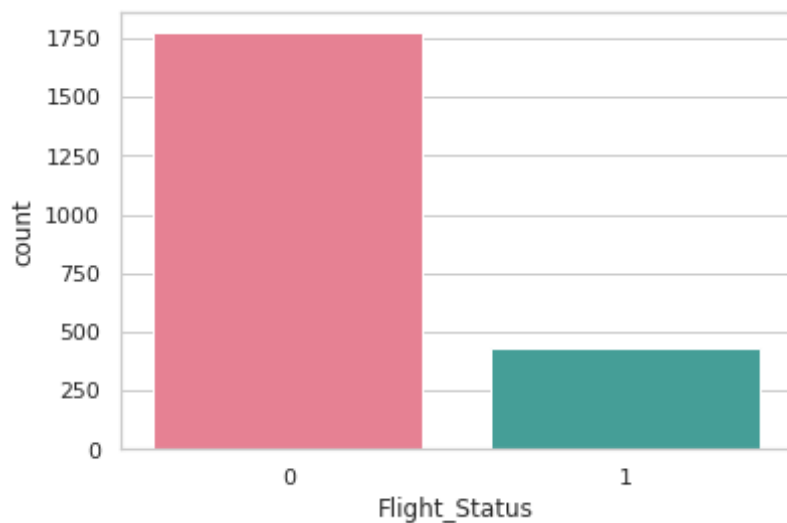


ASSIGNMENT 1

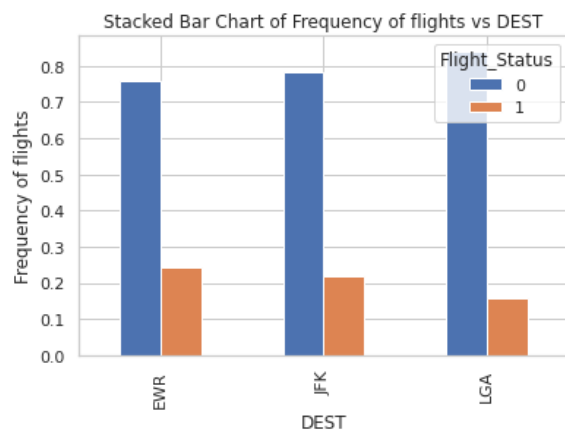
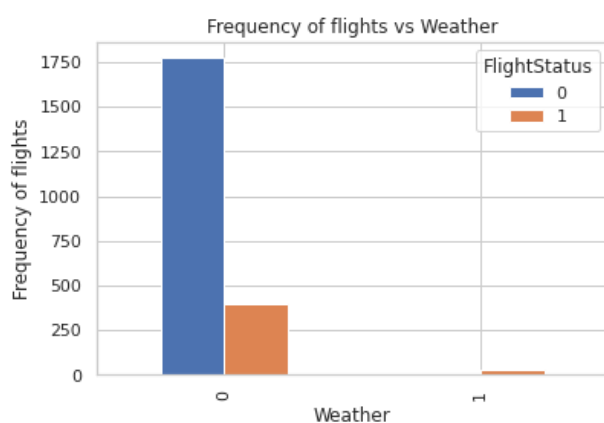
FLIGHT DELAY ANALYSIS

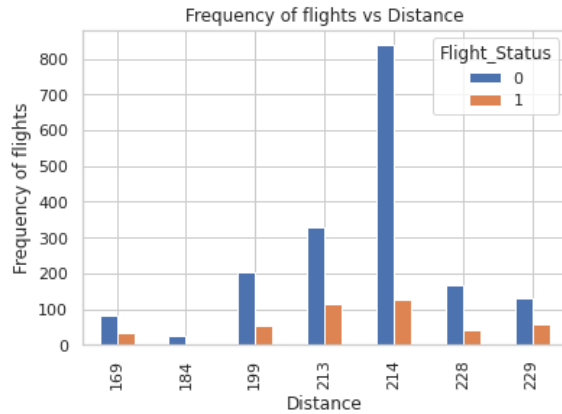
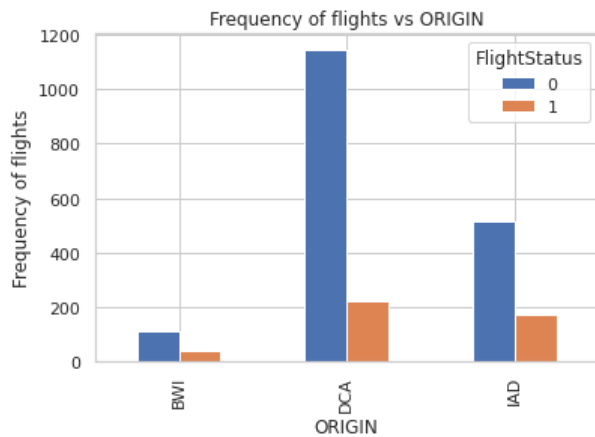
Q1. Exploratory Data Analysis



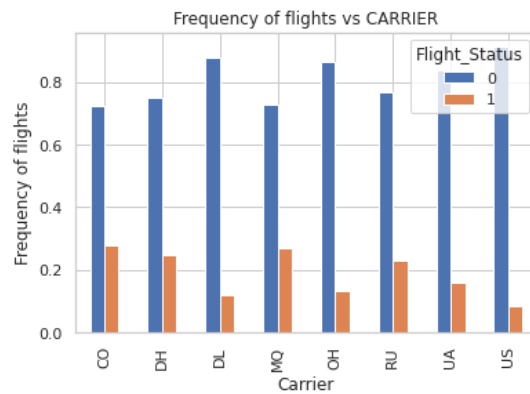
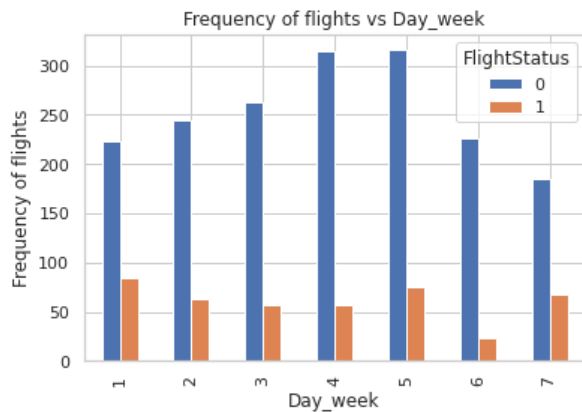
80.55% of the flights haven't been delayed.

GOOD CLASSIFIERS: Weather, destination (DEST), origin, and distance can be considered as good classifiers due to the lowest number of classes, and good variance.

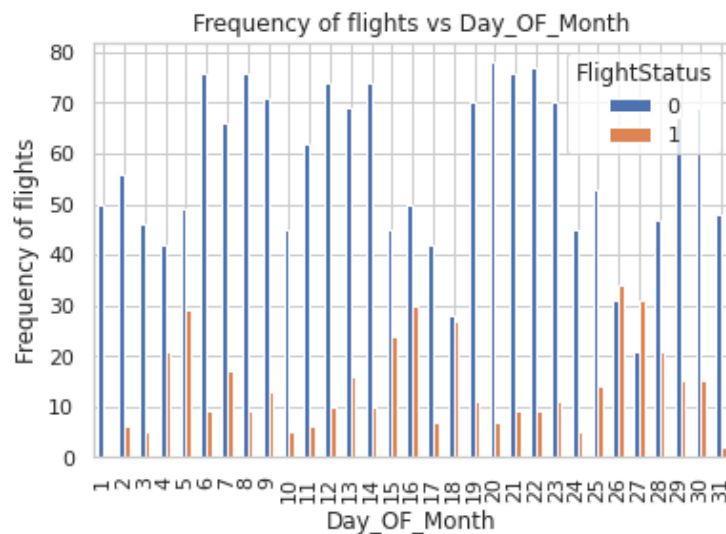




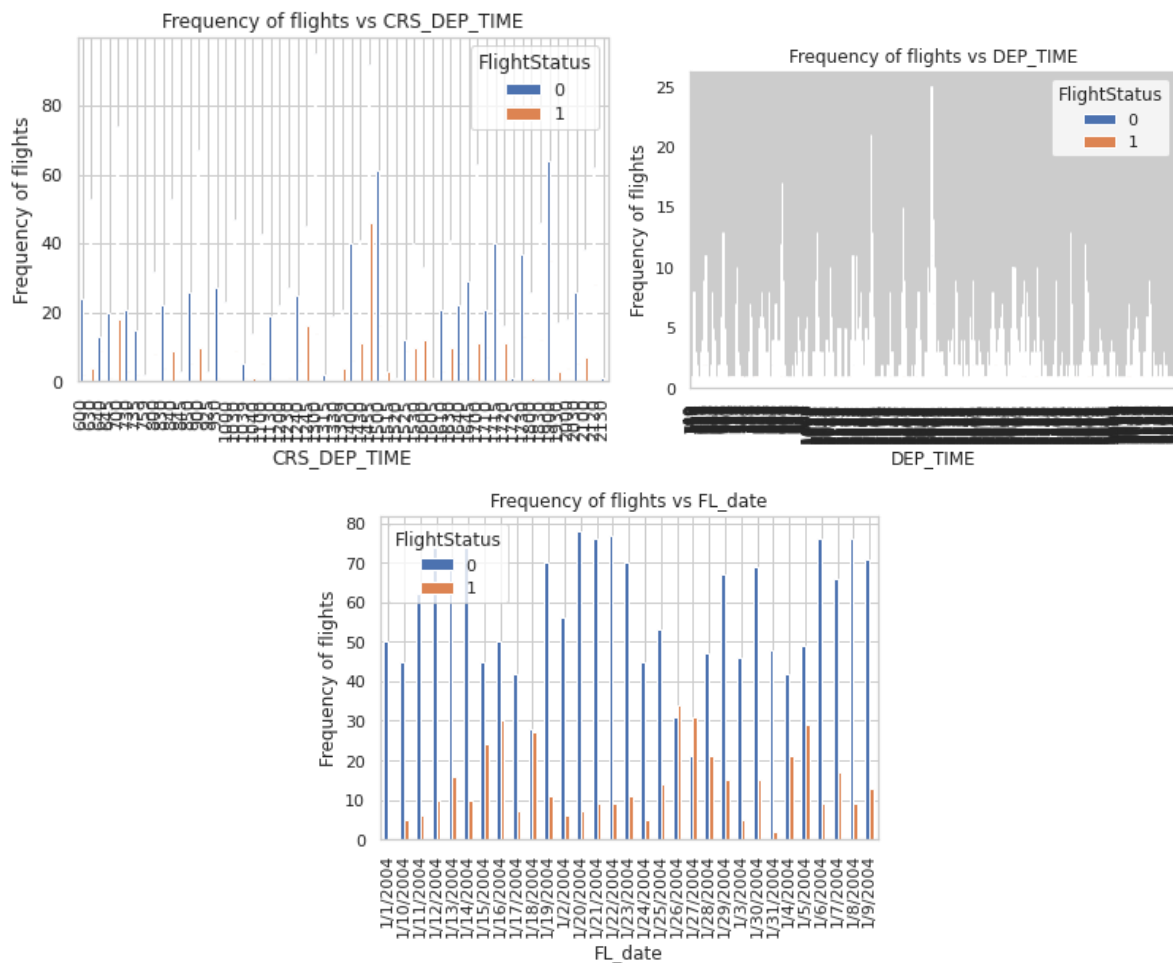
DECENT CLASSIFIERS: Day (Day_week) and Carrier can be considered as decent classifiers because of their mediate number of classes, and the variance of the values.



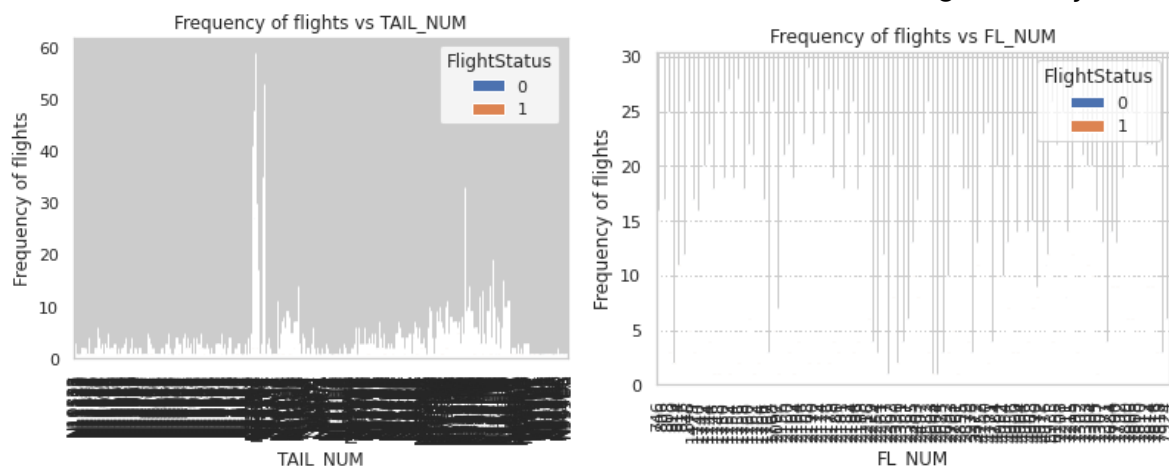
Day of the month in itself might not be a good classifier but if eliminate some of the days, the variance might make it useful.



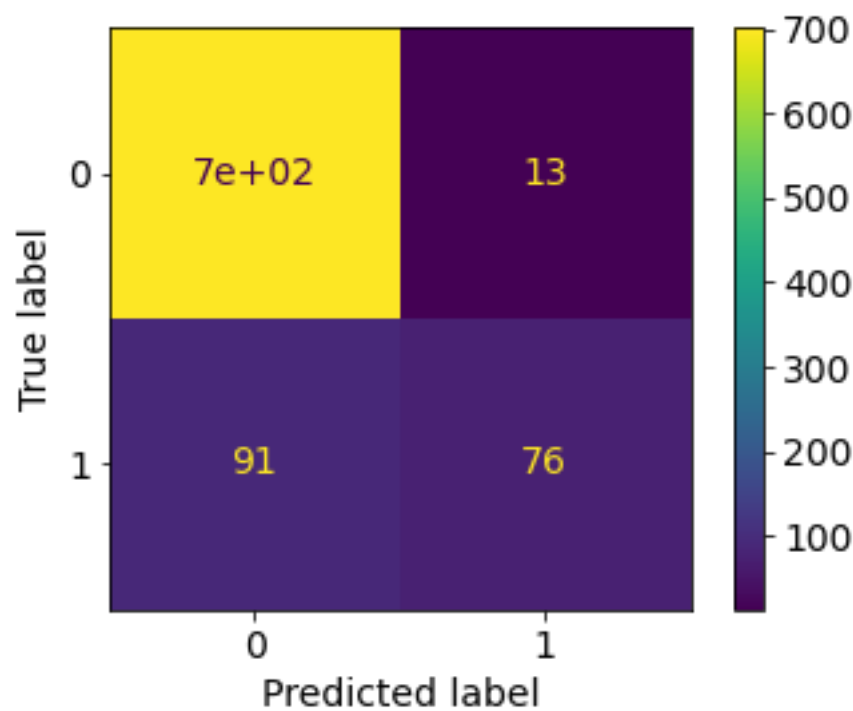
USABLE CLASSIFIERS: These classifiers aren't the ideal classes to be considered but still add to the accuracy, and have noticeable influence on the flight delays hence can be used



BAD CLASSIFIERS: All of these variables are bad classifiers due to the huge number of classes, and not a lot of variance. (This indicated that these classifiers have no real influence on the flight delays)



Q2. Logistic model results:



	precision	recall	f1-score	support
0	0.89	0.98	0.93	714
1	0.85	0.46	0.59	167
accuracy			0.88	881
macro avg	0.87	0.72	0.76	881
weighted avg	0.88	0.88	0.87	881

Accuracy:88.19%

Q3. A feature having higher influence on the probability of the flight being delayed has a higher coefficient than other features. Hence comparing these values and assigning higher values of any increasing function is a good idea in this case. Exponential and sigmoid functions are commonly used to serve this purpose.

The accuracy of the model highly depends on the number of features selected. In logistic regression, the sigmoid function is used, which minimizes negative values and maximizes the influence of the more important features.

Q4. As seen in Q1.

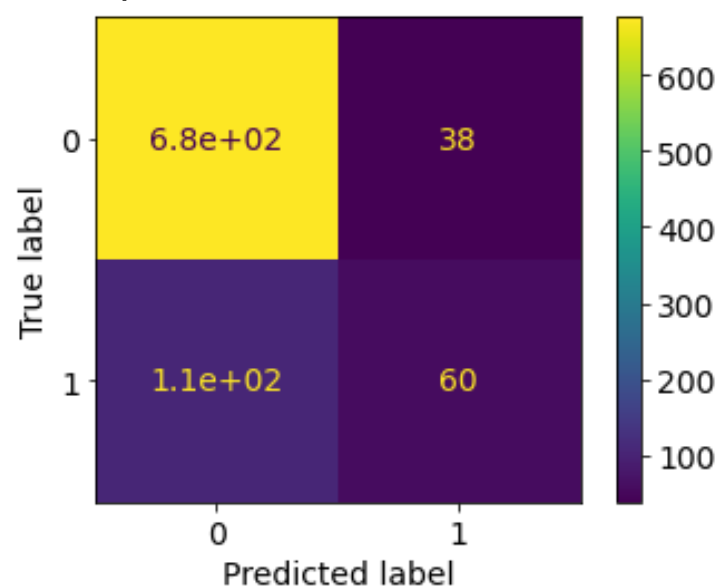
Good variables: Weather, destination (DEST), origin, and distance

Decent/Average: Day (Day_week), Day of the month and Carrier

Usable: departure time, CRS Dept. time, Flight date

Not Usable: Tail number and Flight number

Q5. Removing the less important features,(I considered that flight date was least important of the considered features) we get



	precision	recall	f1-score	support
0	0.86	0.95	0.90	714
1	0.61	0.36	0.45	167
accuracy			0.84	881
macro avg	0.74	0.65	0.68	881
weighted avg	0.82	0.84	0.82	881

We see that the accuracy now drops to 83.54% which is lower than before so it is a better idea to keep the flight date as a feature.

Q6. Ideal conditions:

Good weather, 1st day of the month, US Airlines, at 3:00PM

Bonus Question answers:

- 1) H.OM.E.R., P.L.A.T.O., and V.I.R.G.I.L.
- 4) C3-P0 AND R2-D2