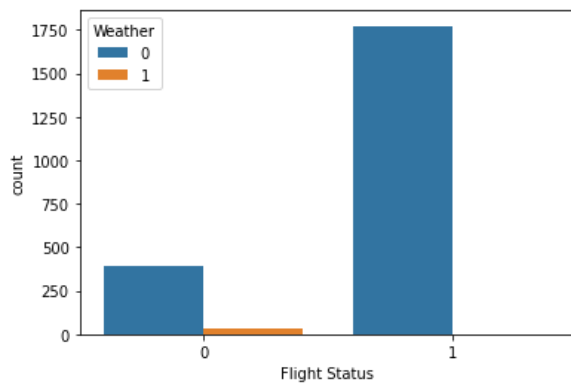
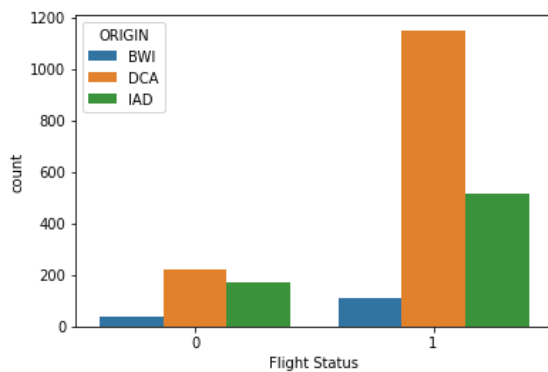


Q1.



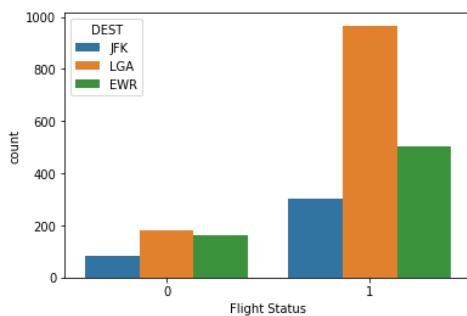
We can conclude that:

1. If there is weather related delay then flights are confirmed delayed
2. If there is no weather-related delay then flights are mostly on time i.e delay is due to some technical faults and delays.



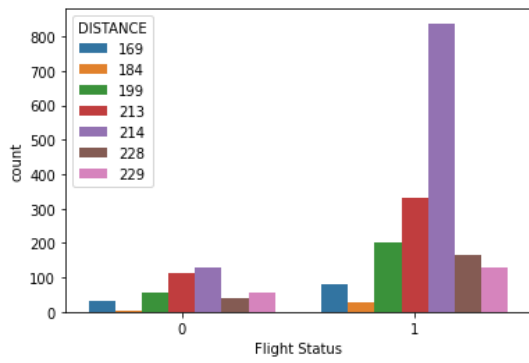
We can conclude that:

1. Most flights are from “DCA” origin and are mostly on time
2. Least flights are from “BWI” and significant are delayed



We can conclude that:

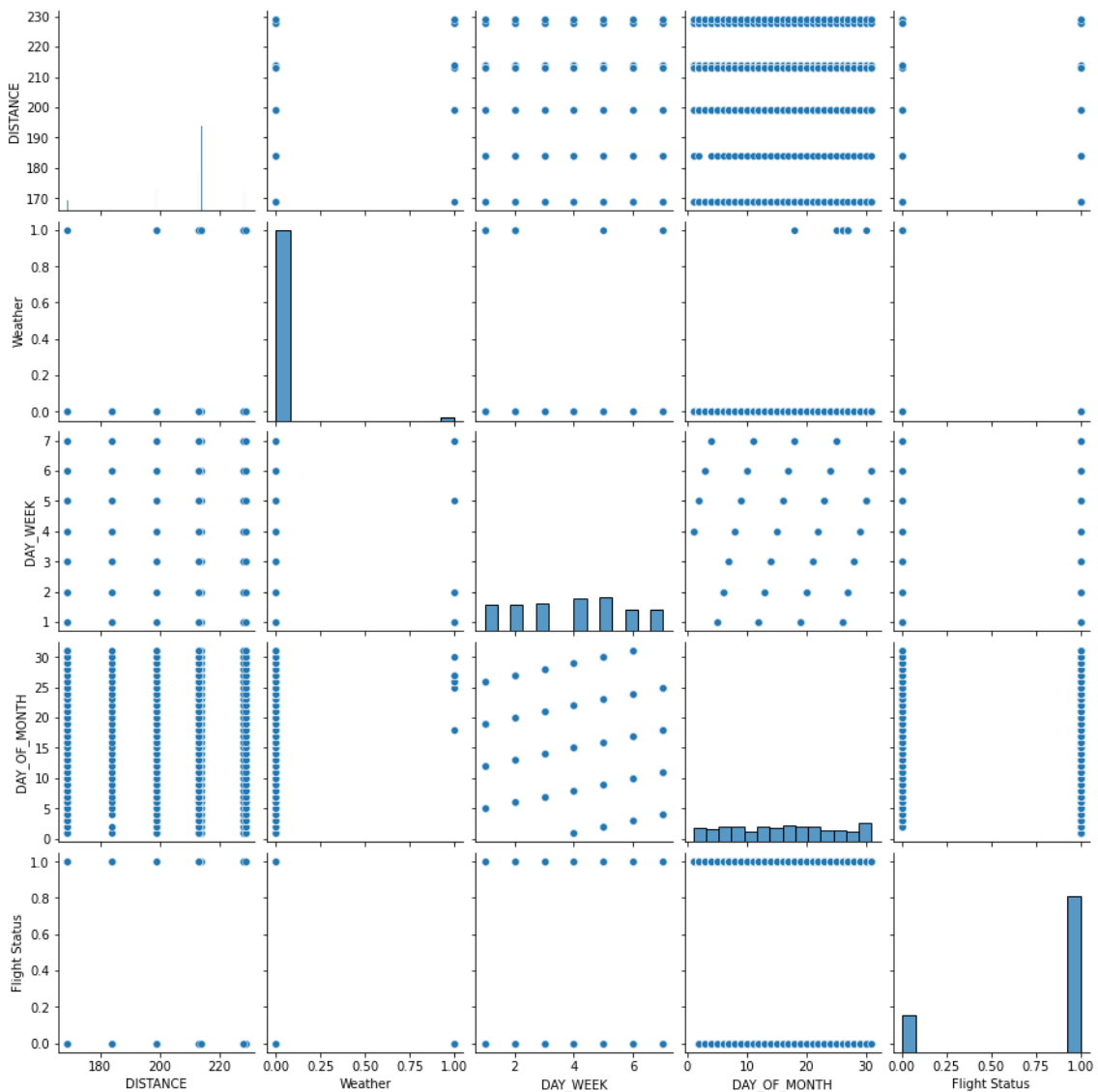
1. Most flights have destination as “DCA”
2. Least flights have destination as “BWI”



We can conclude that:

1. Most flights have distance as “214 km”
2. Least flights have destination as “184”

## Pair Plot



Q2. Look in the code file

Q3.

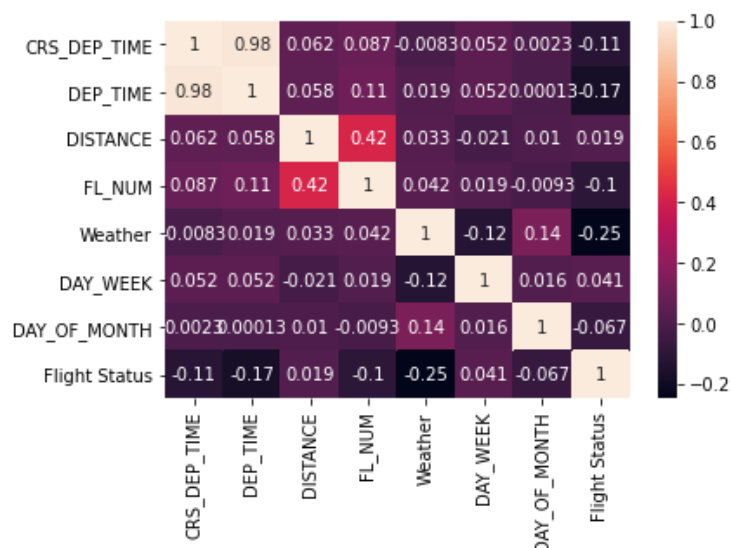
```
Precision: 0.60  
f1-score: 0.16  
accuracy: 80.54%
```

```
confusion matrix = [[ 21, 146],  
                    [ 2, 712]]
```

It is

	Predicted no	Predicted yes	
Actual no	24	241	265
Actual yes	16	1040	1056
	40	1281	

Q4.

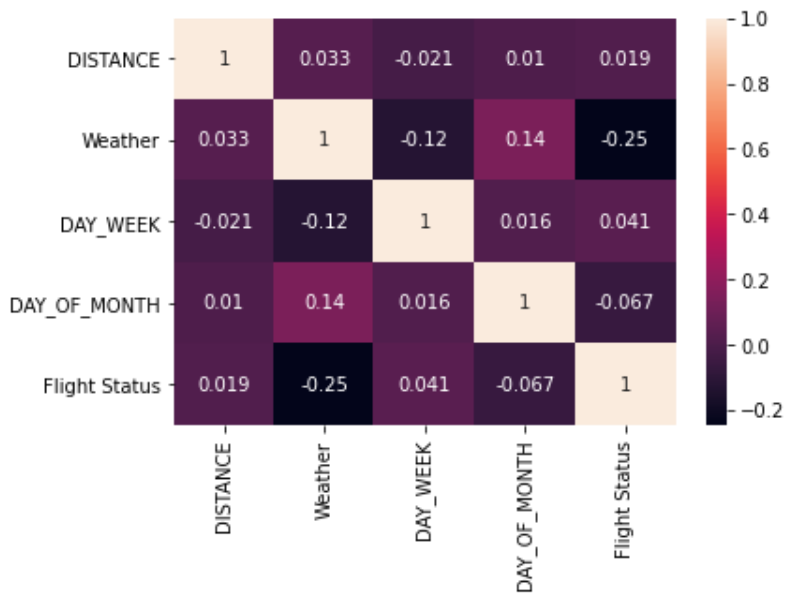


From this correlation heatmap we can see that none of the variable/ feature are correlated with each other as almost all values are  $< 0.03$  except "CRS\_DRP\_TIME" and "DEP\_TIME" so they can't be used as a variables/ feature to predict model as it can overfit the data.

But I have further used the importance for the feature and concluded that only

CARRIER, DEST, DISTANCE, ORIGIN, Weather, DAY\_WEEK, DAY\_OF\_MONTH are to be used as they are mostly responsible for the delay in flight.

And further FL\_NUM and more are not much relevant to this model.



This shows the correlation of my new model which looks quite good

Q5.

```
Precision: 0.86
f1-score: 0.16
accuracy: 81.45%
```

```
confusion matrix = [[ 24, 241],
                    [ 4, 1052]]
```

It is

	Predicted no	Predicted yes	
Actual no	24	241	265
Actual yes	4	1052	1056
	28	1293	

Bonus

Q1. Ultron and Veronica are AI made by tony stark

Q5. Name is R2-D2 and C3PO

<https://github.com/Shivansh1910/ds.git>