Another Delay?!

Using KNN to predict departure times

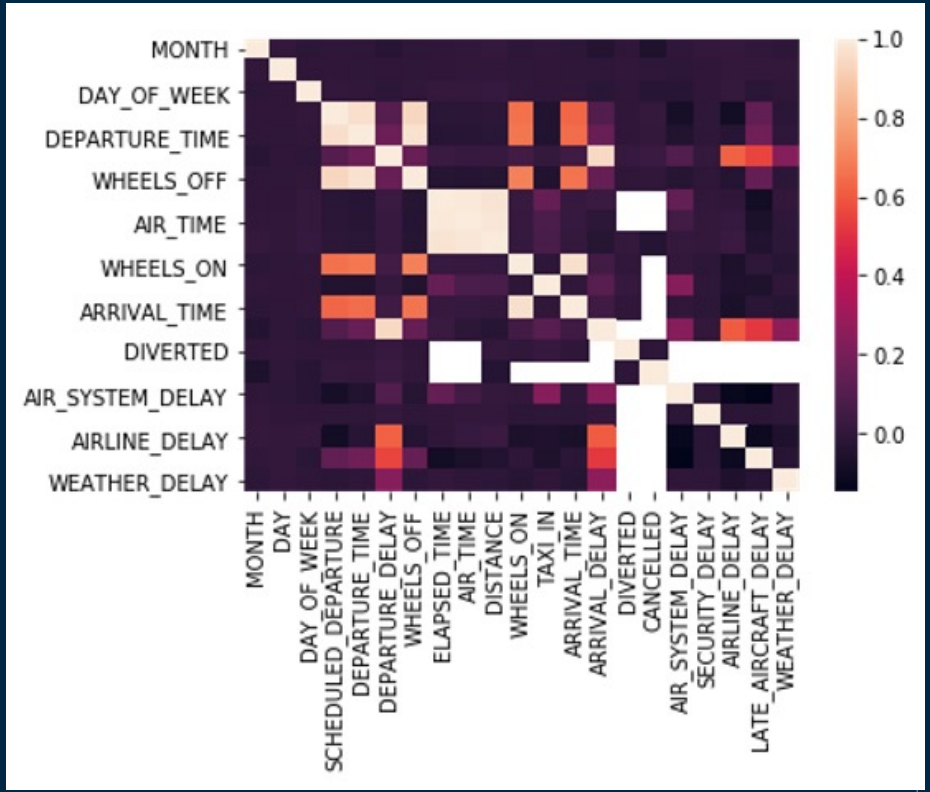
Akshar Shrivats
IB Computer Science HL 4B

The Data

- Extremely Large
- 3 individual files
- 31 total factors
- Published by the DOT
- Fairly recent

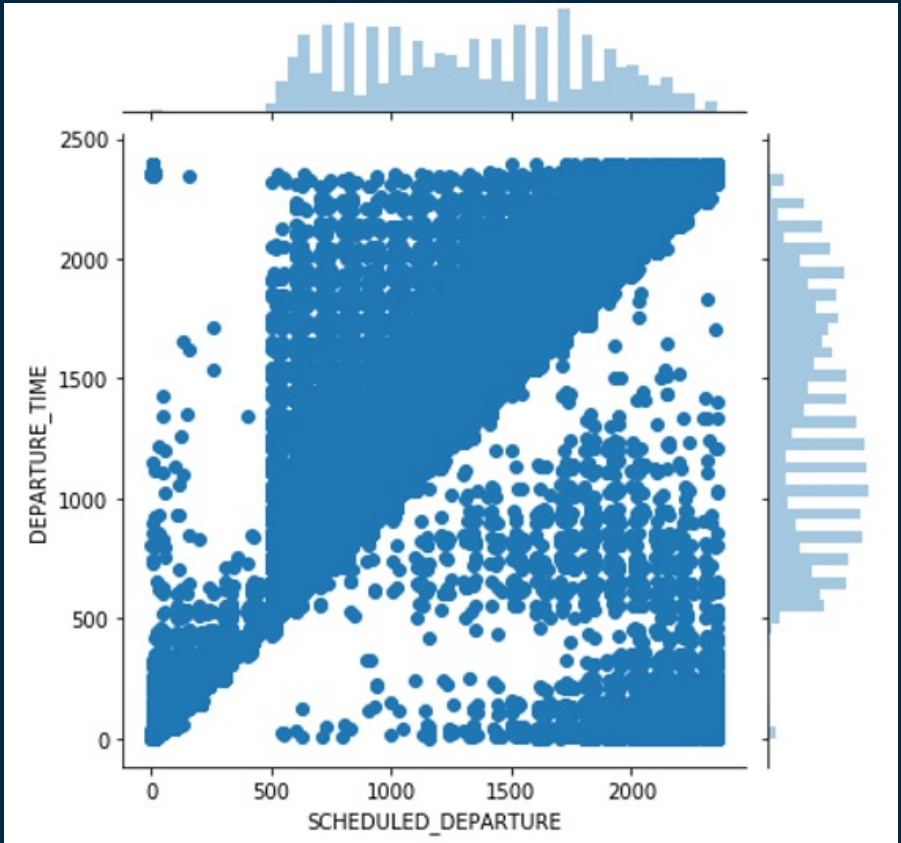
Data Visualization

A Correlation Plot to find trends visually

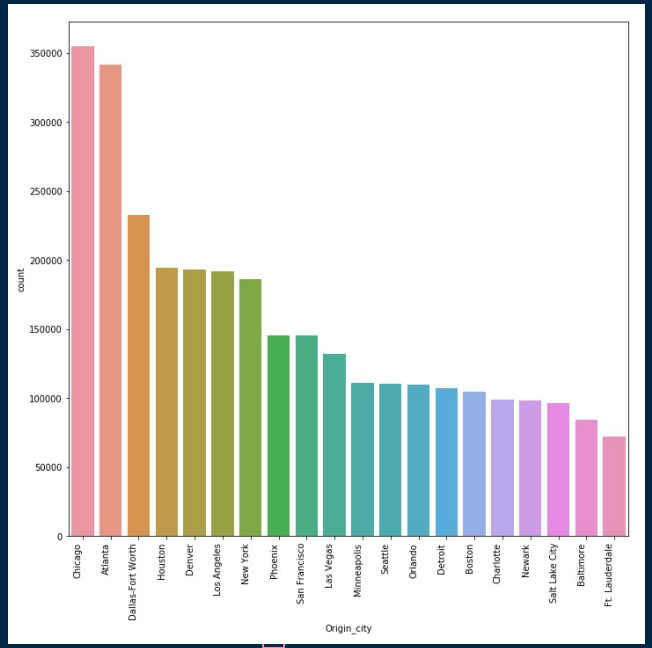


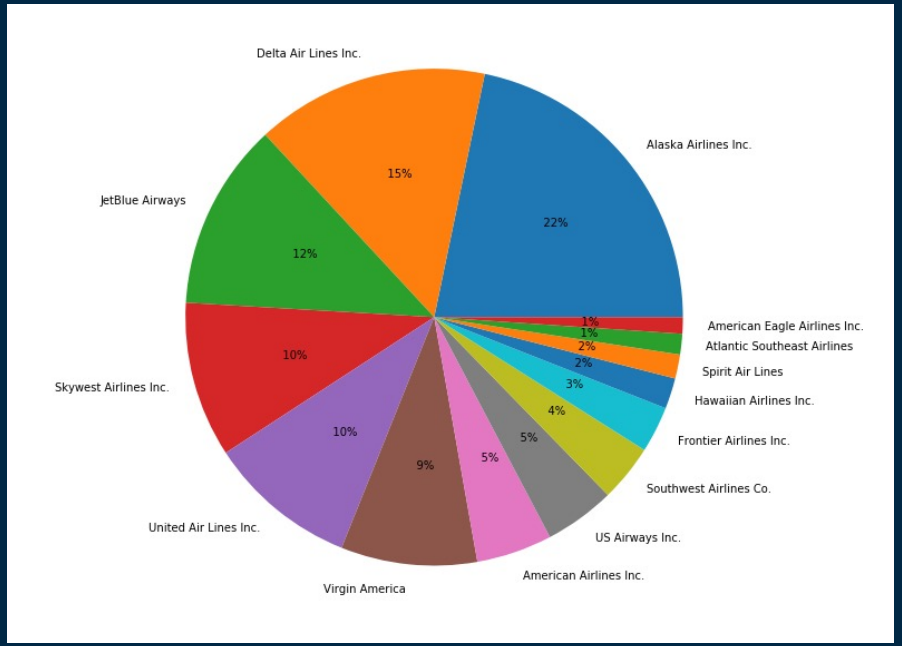
Data Visualization

A jointplot to see if my hypothesis was valid



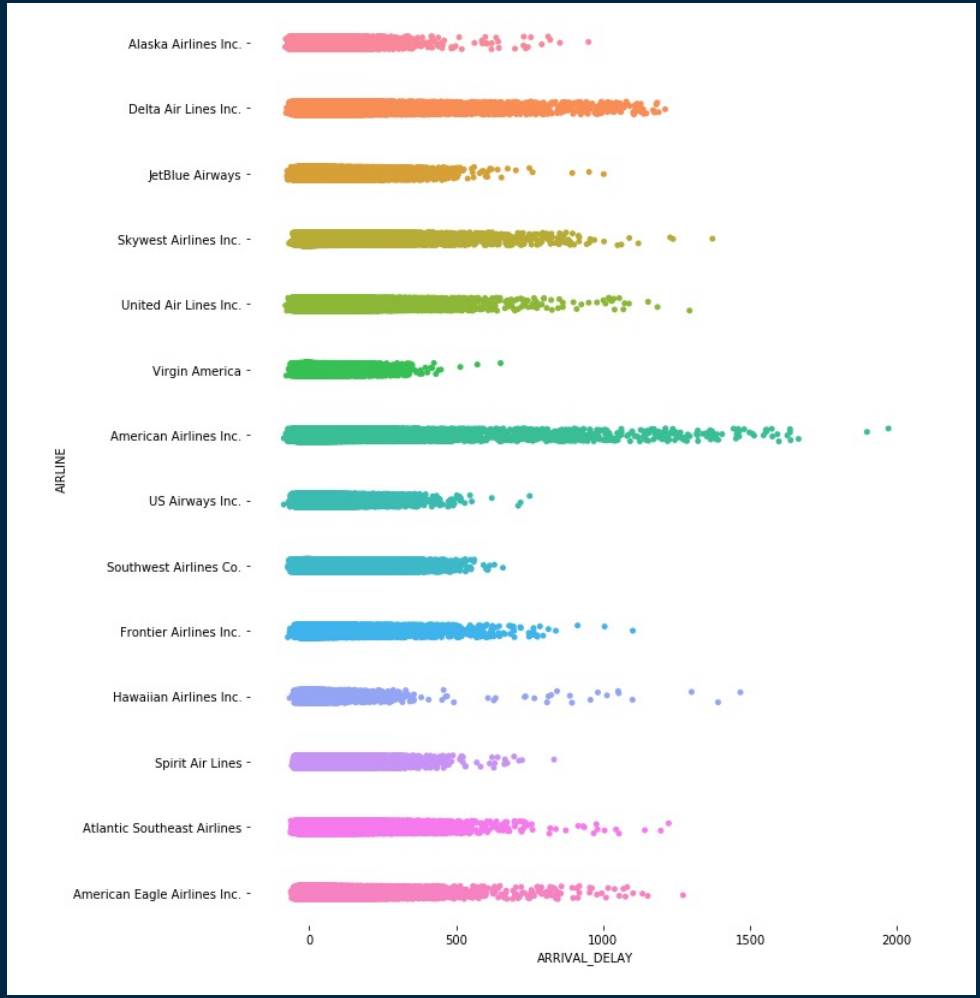
Data Visualization



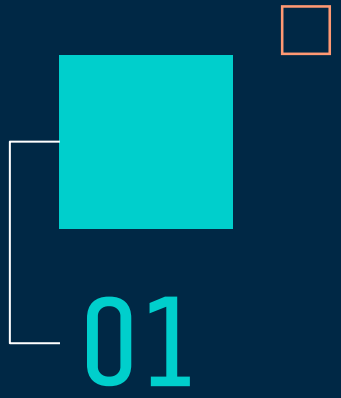


Data Visualization

What airlines had the longest delays?



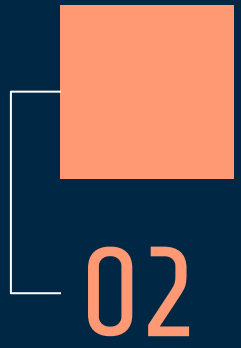
Handling the Data



01

Deleting the junk

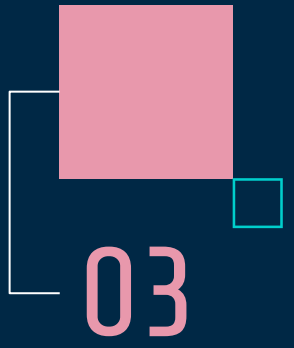
What factors don't matter?



02

Converting to Float

What do we do with weird strings?



03

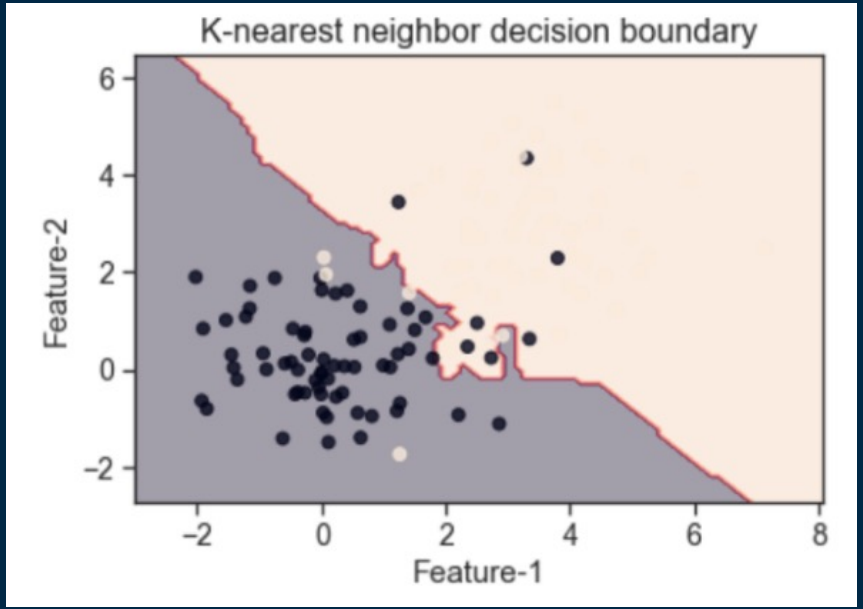
Normalizing

What stuff matters the most?

The Results

Visualizing the KNN boundaries

Unweighted Model: 88%
Weighted Model: 92%



Conclusion

- Arrival Time
- Airline Delays
- Air System Issues

THANK YOU