Final project Hector Napier

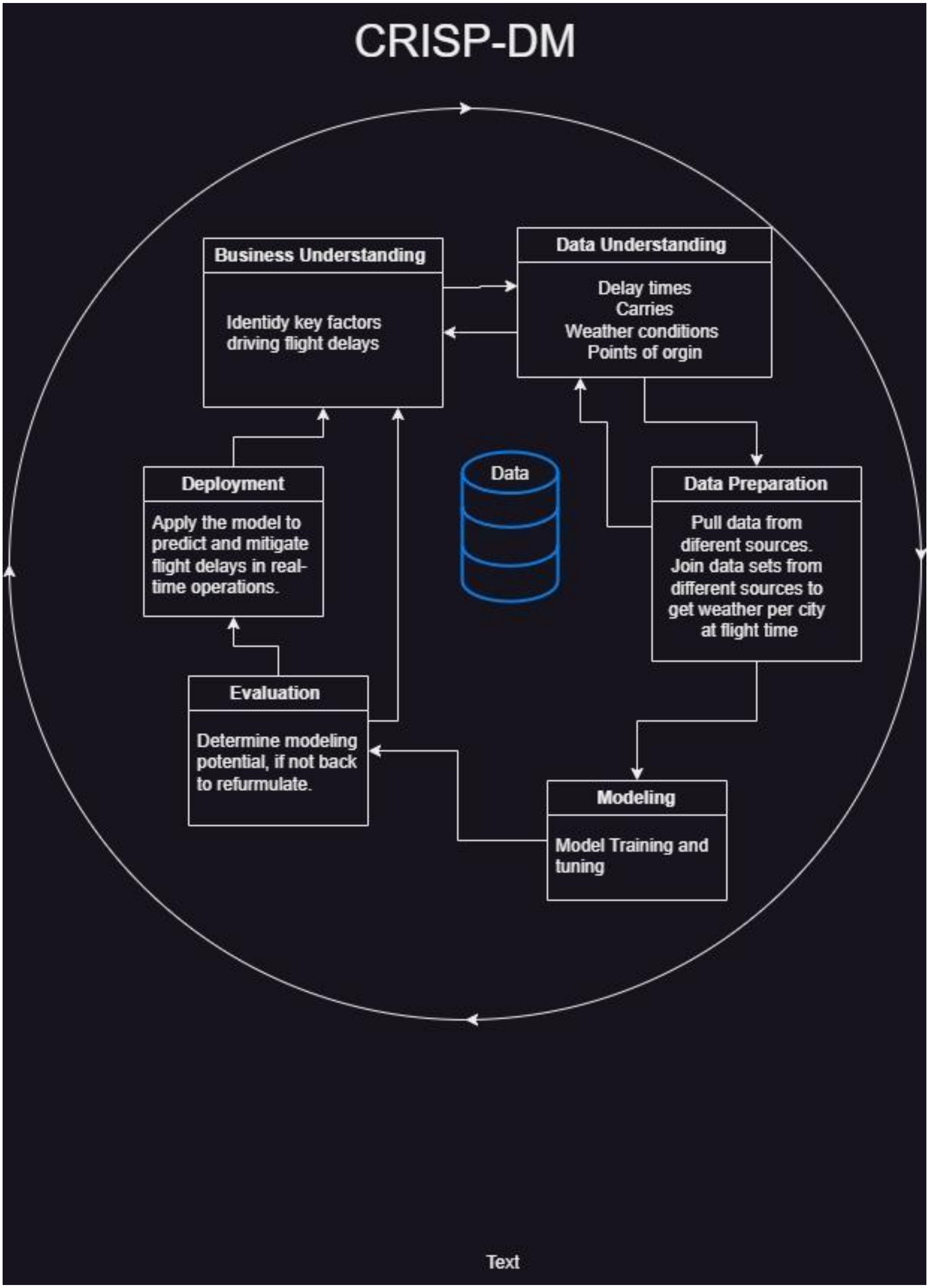
Problem to solve:

We are trying to identify the key driving factors behind flight delays. There are several variables that can potentially impact delays, including carrier, weather, origin, and route distance. The purpose is to determine which of these factors is most likely to drive flight delays. Understanding these factors is crucial for improving operational efficiency, enhancing customer satisfaction, and reducing costs associated with delays.

	Setup	Actions	Outcomes	Results
Current Implementation	• Collect initial data on flight delays from various airlines. • Identify key metrics: carrier, distance, origin, time of day.	• Analyze collected data to determine initial trends. • Assess carrier delays, route distances, and points of origin.	• Gain preliminary understanding of delay patterns. • Identify which carriers have higher delays.	• Basic insights into flight delays.
Future Implementation	• Collect comprehensive historical data on delays, including weather impact.	• Develop predictive models to calculate expected delays. • Continuously monitor and update the model.	• Detailed understanding of delay causes. • Identification of key operational factors driving delays.	• Reduction in flight delays through targeted operational improvements.

Data sources:

Various data sources were used to pull data and combine the data in order to conduct the analysis. In order to get the final dataframe that was used for analysis, we pulled BTS flight data that had various features like the date of the flight, the carrier, the departure time, the delay, etc. That data focuses on flights where the AUS airport was the destination and flights were not canceled or diverted. We also pulled data from Iowa State University's Iowa Environment Mesonet (IEM) to get the weather data in different cities on the dates in the date range that we looked at flights for. This data contains sky cover and visibility data and let us calculate the flight conditions for particular flights. Flight Aware is where we were able to pull upcoming flights into AUS and make predictions for these flights and their potential delays, based on the historical flights' data.


CRISP-DM
Business Understanding
Identidy key factors driving flight delays
Data Understanding
Delay times
Carries
Weather conditions
Points of orgin
Data
Deployment
Apply the model to predict and mitigate flight delays in real-time operations.
Data Preparation
Pull data from diferent sources. Join data sets from different sources to get weather per city at flight time
Evaluation
Determine modeling potential, if not back to refurmulate.
Modeling
Model Training and tuning
Text

Data analyzed:

There were several data points we assessed including average delay by the carrier, distance, origin, time of date, and deviation between expected flight delay for weather and actual flight delay.

The first data group that was assessed, was what we should have expected as far as carrier delay prediction. As highlighted in Figure 1.1 below, it appears that there are certain airlines that we would expect to have longer delays than others. While this does show that certain carriers should have longer delays it doesn't answer the question of whether the weather or the carrier operations are the drivers behind this delay.

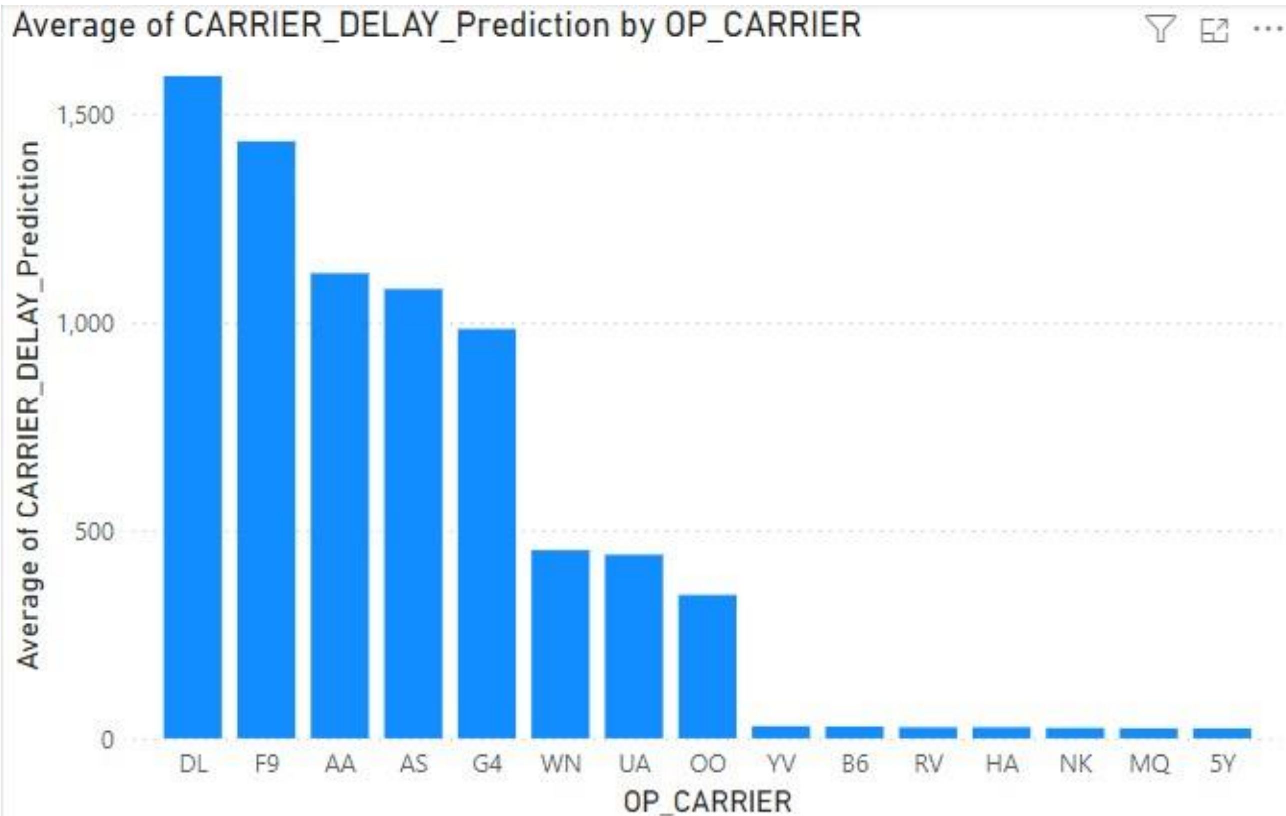


Figure 1.1

The next area assessed in the data set was to see if there is any correlation between average delay prediction and the route distance. This would allow us to determine if longer routes should expect longer delays. As Figure 1.2 illustrates, longer routes do not necessarily mean that there are longer delays. Therefore, there is no strong correlation between distance and expected delay. This implies that route length is not a driving factor behind delays.

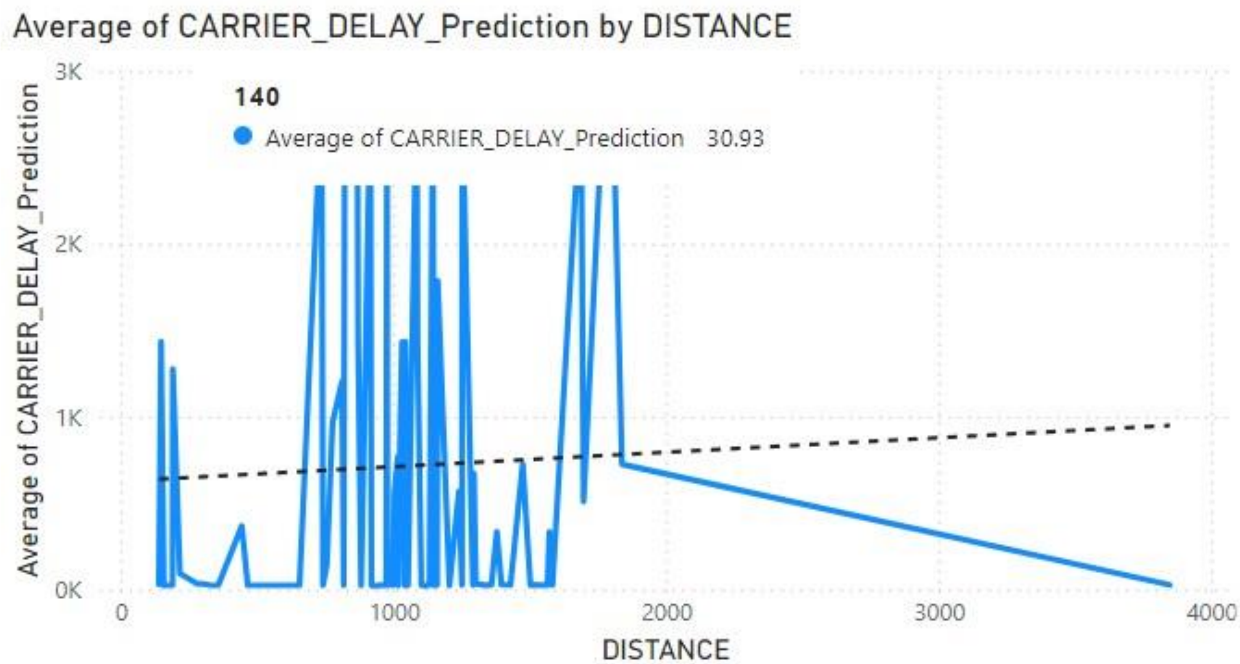


Figure 1.2

The following data analysis that was evaluated was identifying if there was a correlation between expected carrier delay and point of origin. As seen in Figure 1.3, there are several points of origin with similar standard deviations and no clear outlier. This implies that there is not a specific point or origin that drives an expected carrier delay.

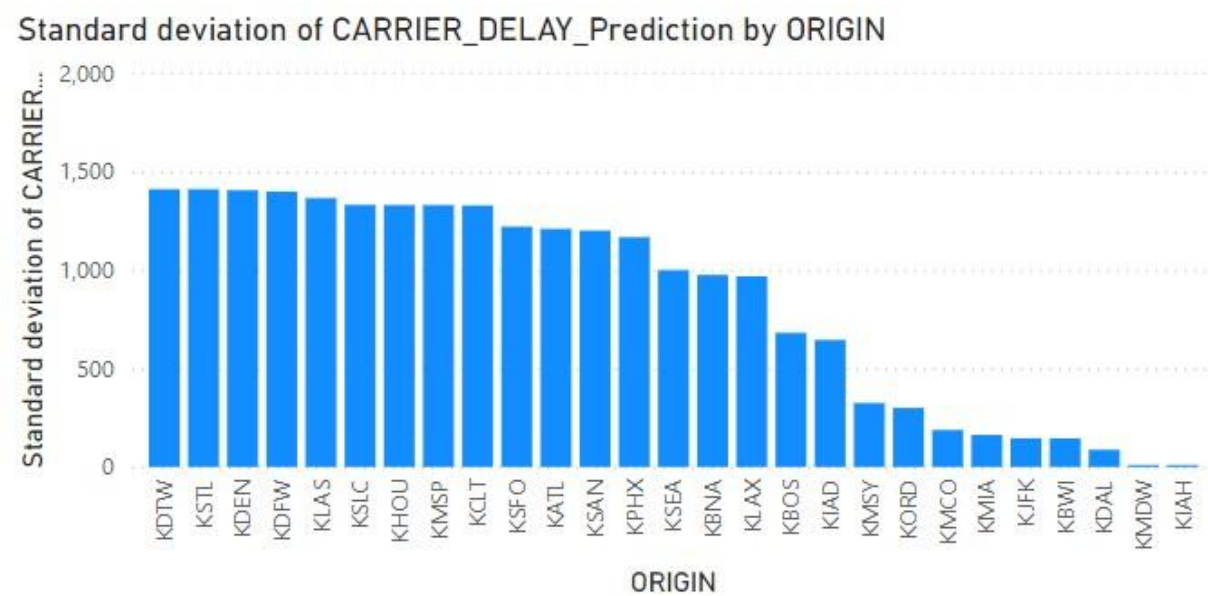


Figure 1.3

Moving on to our next set of data points, we wanted to identify if the time of day would be a predictor of the expected delay. As indicated in Figure 1.4, it does not appear that there is a clear difference in expected delays in PM and AM departures.

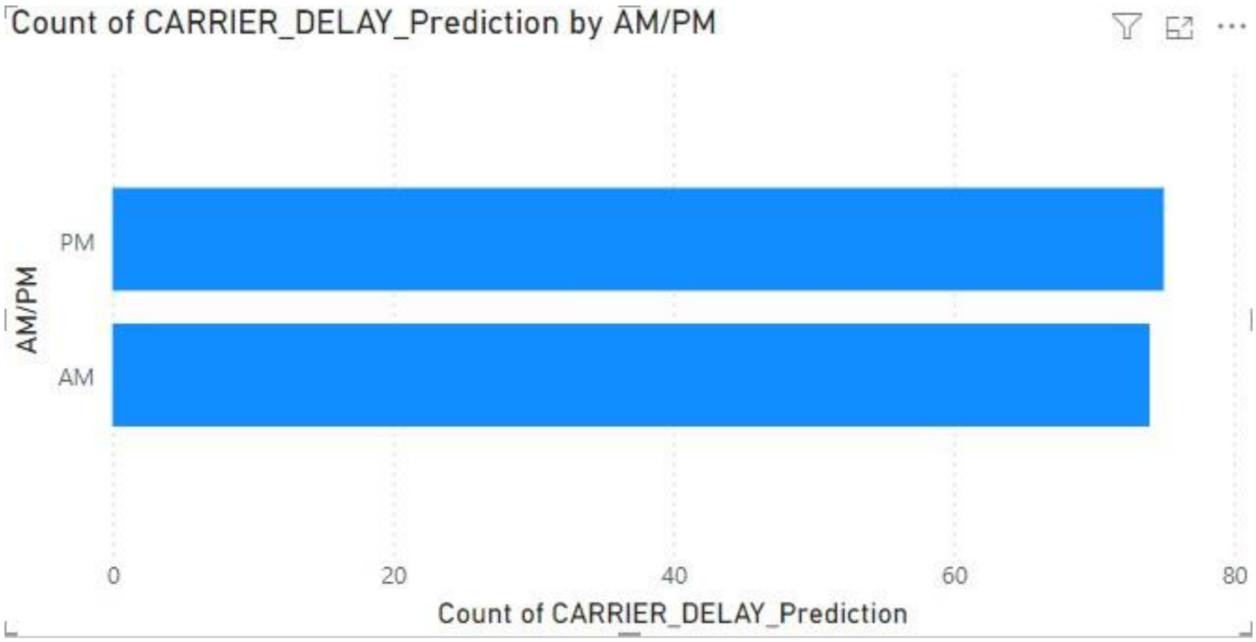


Figure 1.4

Lastly, we looked to see if there was a difference between the expected delay and the actual delay. This was done to see if normalized for weather there was a significant difference between expected delay and actual delay. The outcome of this should demonstrate if the weather or the carrier operations drove the expected delay. As indicated in Figure 1.5, it appears that carrier and not the weather is the driving factor behind delays with several airlines deviating significantly from the expected value, with DL being the greatest offender. Additionally, the r2 of the regression analysis shows .996 indicating a strong correlation between airline operations and expected delay.

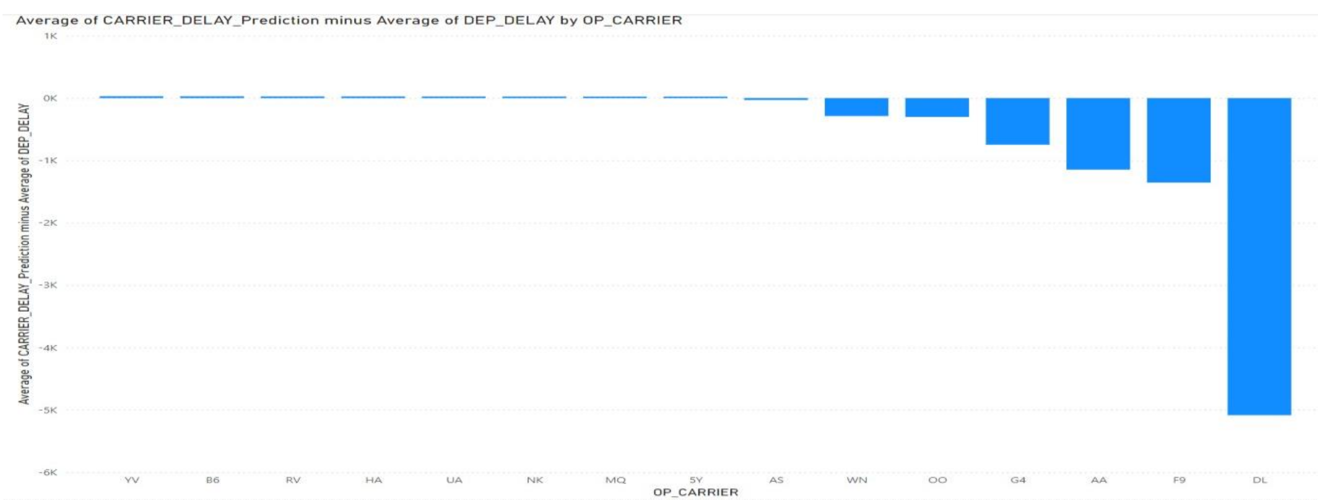


Figure 1.5

Conclusion:

After analyzing multiple factors that might impact delay by carrier, including distance, origin, time of day, and the difference between expected and actual delays for weather, it is concluded that the biggest driving factor of expected delay is carrier operations. This finding highlights the importance of focusing on improving operational efficiency within airlines in order to minimize delays. Addressing these internal factors can lead to more reliable flight schedules and better overall performance in the aviation industry.