**Casa 08 – Airline Delays Analysis**

Reily Valenzuela

University of Texas at El Paso

CIS 3330: Analytic Programming Tools

Dallin Fairbanks

April 5, 2025

**INTRODUCTION**

In this report, you will be introduced to the Descriptive Analytics and Predictive analysis between ARR\_DELAYS and DEP\_DELAY. This analysis will incorporate visuals, including OLS Plots, OLS Results, Boxplots, and a Correlation matrix.

Regarding ‘ARR\_DELAYS’ or, more accurately, Delays in Arrivals, I decided to test a couple of options. I wanted to base my decision on a plot that best caught my attention. I tested the following:

* CRS\_DEP\_TIME

A graph of purple dots

AI-generated content may be incorrect.

* DEP\_DELAY (Departure Delay)

A graph of a graph showing a curve

AI-generated content may be incorrect.

* WEATHER\_DELAY

A graph of a graph

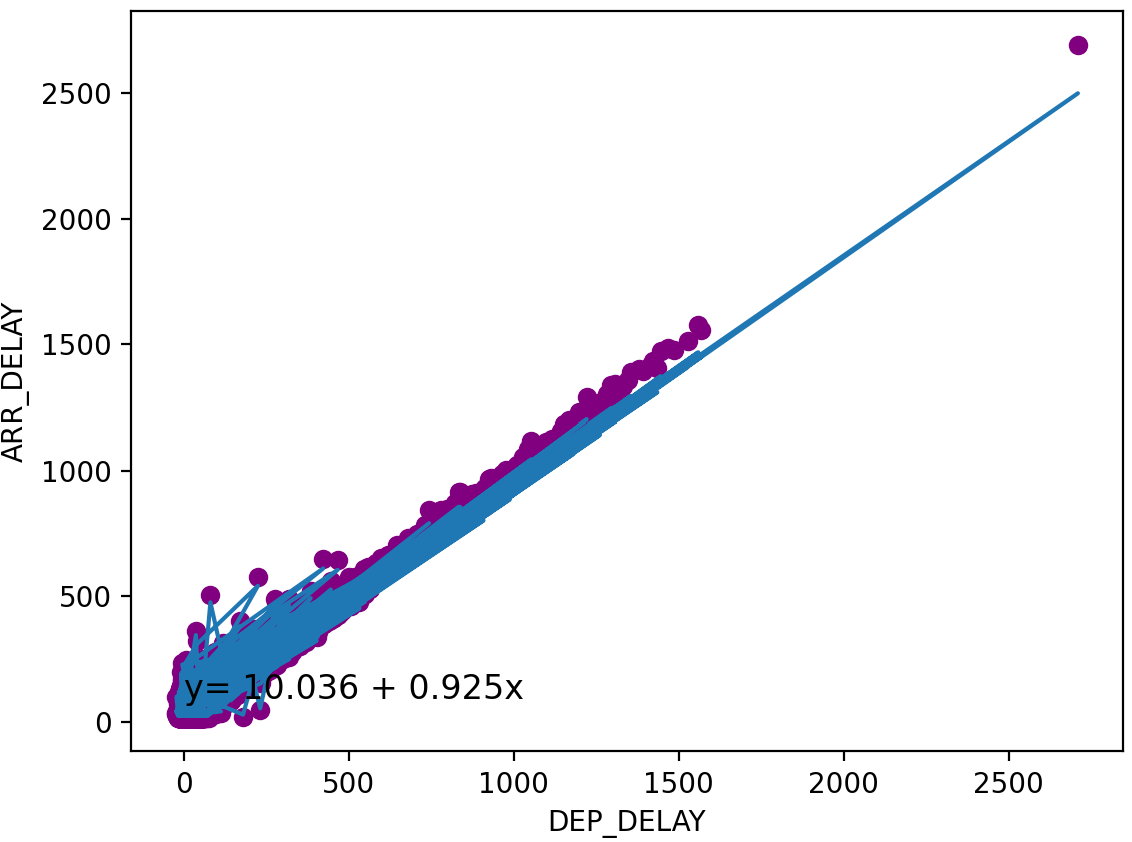
AI-generated content may be incorrect.

As you can see, CRS\_DEP\_TIME and WEATHER\_DELAY had very interesting plots, and although it would’ve been fun trying to describe those predictors, I decided to play it safe and proceed with DEP\_DELAY as it seems to be more consistent and transparent.

**TERMINAL – DATA OVERVIEW: OLS**

The following image contains the OLS Regression Results & Visual:

A screenshot of a computer

AI-generated content may be incorrect.

As displayed on the results image, the R-squared value is one that immediately captures one’s attention. With a value of 0.927 (92.7%), the R-square establishes a strong linear relationship between arrival and departure delays, proving that departure delay is a big predictor of arrival delay. Additionally, when it comes to P-values, we strive for values under 0.05. Within our results, we can see that our P-value demonstrates a value of 0.00, indicating a strong statistically significant association. Moreover, focusing on the constant, as displayed on both results and visual, *y = 10.036 + 0.925x* displays a highly positive linear relationship. The intercept 10.036 indicates that when the Department Delay is at 0, there is still an expected arrival delay of approximately 10.04 minutes. Regarding the coefficient for Department delay, it indicates that for every 1-unit/minute increase in DEP\_DELAY, there is a predicted increase in Arrival Delay of 0.92 (0.9245) minutes. Lastly, as seen in the plot, there appears to be a single, significantly distant outlier.

A graph showing a line of blue and red dots

AI-generated content may be incorrect.

From my understanding, the plot displayed above represents the predicted and actual observed arrival delays. The blue symbols represent the actual observed arrival delay, the raw data. The red symbols plotted represent the “fitted” or the predicted arrival delays, the estimates. The fact that the blue and red symbols positively co-align with each other in the trend indicates that the linear regression model fits the data significantly well.

**TERMINAL – DATA OVERVIEW: BOXPLOT**

**A graph of a graph showing a line graph

AI-generated content may be incorrect.**

The visual above represents the box plot of the analysis. It appears that the median is below the center of the box, indicating a positive skewness with a right-skewed distribution. The plot also seems to reveal numerous outliers, suggesting that most flights go through short or concentrated arrival delays. Although most outliers seem to be under the 1500 mark, there is a significantly distant outlier, expressing the fact that there are instances where delays can be longer than predicted. This outlier was also mentioned on the OLS visual displayed earlier in the analysis.

**TERMINAL – DATA OVERVIEW: CORRELATION MATRIX**

A screenshot of a computer game

AI-generated content may be incorrect.

The correlation matrix displayed above reveals a significant positive correlation between ARR\_DELAY and DEP\_DELAY with a value of 0.96. This is exceptionally high, and just like every other visual that has been demonstrated, it expresses a powerful linear relationship. The relationship suggests that departure delays are a strong predictor of arrival delays and implies that for flights that experience departure delays, regardless of whether they’re short or long, it is expected to affect the arrival will be delayed just as much.

Given the correlation matrix, I also squared the WEATHER\_DELAY and the CRS\_DEP\_TIME to compare the values displayed to DEP\_DELAY. Firstly, with CRS\_DEP\_TIME, you can see that it is significantly low at 0.05. Going forward, WEATHER\_DELAY also shows a low correlation with a value of 0.39, proving that these options wouldn’t have been a powerful predictor compared to DEP\_DELAY.

**CONCLUSION**

In comparison with other predictor options within the CSV file, DEP\_DELAYS has proven to be a strong predictor and a positive one at that. Given the results and all visuals displayed above, DEP\_DELAYS has demonstrated a powerful correlation and relation with ARR\_DELAY.

**CODE**

**A screenshot of a computer

AI-generated content may be incorrect.**

**AI DOCUMENTATION**

Line 50

A screenshot of a computer screen

AI-generated content may be incorrect.

A screenshot of a phone

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a chat

AI-generated content may be incorrect.

Line 56

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a chat

AI-generated content may be incorrect.A screenshot of a chat

AI-generated content may be incorrect.A screenshot of a chat

AI-generated content may be incorrect.

**REFERENCES**

GEMINI (2.0 Flash). Date of query (4/04/25). Generated using OpenAI GEMINI. https://gemini.google.com/app