CODE 8 Report

The objective of this code and report is to perform descriptive and predictive analytics on U.S. flight data from 2018, focusing on relays in arrivals (ARR\_DELAY). The dataset contains over 280,000 observations.

The purpose of this descriptive analytics is to identify patterns and relationships between flight delay variables and to select the most relevant predictors for arrival delays.A screenshot of a computer screen

AI-generated content may be incorrect.

With the heatmap above, we can see that DEP\_DELAY has a strong correlation with ARR\_DELAY (0.96); flights that depart late usually arrive late. CARRIER\_DELAY has the next strongest correlation with ARR\_DELAY (0.57), delays caused by the airline are closely tied to arrival delays. The next strongest correlation is ATE\_AIRCRAFT\_DELAY, it has a 0.52 correlation with ARR\_DELAY, delays form previous flights using the same aircraft carry over. Weather issues also have a significant effect on arrival delays; WEATHER\_DELAY has a correlation of 0.39 with ARR\_DELAY. The last one is NAS\_DELAY, which has a correlation of 0.27 with ARR\_DELAY, delays from air traffic or system-related issues.

TAXI\_IN, TAXI\_OUT, SECURITY\_DELAY, DISTANCE, and AIR\_TIME have low or near-zero correlation with ARR\_DELAY, so they likely don’t help much with predicting arrival delay.

This is the code used to produce this heatmap.

A computer code on a black background

AI-generated content may be incorrect.

To create predictive analytics a Linear Regression Model (OLS) was built using statsmodels.api. The predictors used were DEP\_DELAY, CARRIER\_DELAY, LATE\_AIRCRAFT\_DELAY, WEATHER\_DELAY, and NAS\_DELAY.

A screenshot of a computer

AI-generated content may be incorrect.

ARR\_DELAY is the value that we are trying to predict. The R-squared, which is 0.999, means the model explains 99.9% of the variation in arrival delays. The F-statistic (0.00) indicates the model is statistically significant. In the coefficients Table, firstly, we have DEP\_DELAY, which has a coef of 0.0265, meaning each additional minute of departure delay increases arrival delay by 0.027 mins. The CARRIER\_DELAY has a coef of 0.9730, meaning almost all of this delay passes through to arrival delay (97%). Similar to CARRIER\_DAELAY, LATE\_AIRCRAFT\_DELAY 97% of this delay also affects arrival time. WEATHER\_DELAY directly contributes to arrival delays. NAS\_DELAY has the highest impact among all, nearly 98% of NAS delays pass through.

The code used to produce the OLS regression is below.

A black screen with green and orange text

AI-generated content may be incorrect.