Milestone 1: Delays and Declines
Happy Herold and Syed Mehran

The authors intend to investigate how airline operational performance can predict stock prices. The data sets for this project are stock market data obtained from Yahoo Finance and domestic flight data from the Bureau of Transportation Statistics TranStats Marketing Carrier On-Time Performance table. To investigate, the authors have compiled data from the last 6 months of available data (01/01/2023 to 05/31/2023) on airline financials and performance. With the exception of the aviation index the 9 major airlines from the stock market data correspond to the market carriers in the flight data .

Since the aim of this project is to predict the stock prices based on airline performance, the stock market data provides actual values of the stock prices. One of the questions that the project aims to answer is whether or not the performance of a stock can be predicted based on the performance of an airline, and does the performance of multiple airlines impact the stock price of a single airline. As part of the ETL process the daily returns for each stock were calculated and used in creating a correlation matrix for daily returns. Time series data of the trading volume, open price, and closing price were also plotted.

To measure airline performance in this project flight data from the Marketing Carrier On-Time Performance table. The flight data contains flight characteristics including date, market carrier, operating carrier, tail number, origin and destination, scheduled departure time, actual departure time, scheduled arrival time, actual arrival time, canceled and diverted flights as well as other key features of the flight. The flight data does not specifically include airline performance metrics. One of the questions that the project aims to answer is if an airline's flight data can be used to create a metric that can quantify the performance of an airline and can these metrics be used to compare different airlines and be able to predict the stock price. During the ETL process daily number of flights, daily number of cancellations, and proportion of flights that are operated by the market carrier have been calculated using the flight data and plotted.

Collaboration Plan:

A private project git repository has been created to store the data and the Colab Jupyter notebook. Authors intend to meet once a week to discuss weekly project goals, raise questions, and develop the final product. Meetings will take place by Zoom or in-person and the pair will communicate using text messaging. The weekly goals and project to-do list are shared in a OneNote document that is shared between the pair. Up until Milestone 2 the authors will continue to calculate various airline performance metrics

from the flight data and begin predicting the stock prices based on these metrics with different scikit-learn regression models. After completing Milestone 2, the authors plan on gathering data visualizations and the model to develop a final website and presentation that clearly explains the aim of the project, the steps and tools utilized in the project, and the findings from the analysis and model created.

Data Set Websites:
https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGK
https://finance.yahoo.com/