Delta Team 1 - MSBA 2024 Practicum
Data Dictionary

In a single folder should be the modeling code (MealPrediction.ipynb), a folder with input data (‘data’) which will contain ‘Fleet Key_v1.xlsx’ and ‘airport-codes.csv’. The output will be stored in a new folder within the root folder after the code is executed.
The most up to date version of the modeling and feature engineering code is MealPrediction.ipynb.

MealPrediction.ipynb

This is a Jupyter notebook file which contains all of our feature engineering pipeline steps, model building, and output code. Within the notebook you can find descriptions for each pipeline function, as well as instructions on how to use the various features of the code. As it stands, the code expects to find a meals per flight historic dataset, passenger load data historic dataset, the fleet key and airport codes dataset we have provided, as well as an input file provided by Delta, all in a ‘data’ folder.

For the input file, each row should reflect a singular flight which you’d wish to predict the meal demand for. Each row requires the following features:
'DepartTime', 'FlightDate', 'FlightOrigin', 'FlightDestination', 'FlightNumber', 'AircraftType'

PipelineAgg EDA.ipynb

This is a Jupyter notebook file which contains all of our feature engineering pipeline steps, along with code to build and evaluate various models. Like with the final code, this contains descriptions for each pipeline function, however it doesn’t contain any specific instructions for how to ‘use’ it. This file is useful for exploring the different ways error occurs, can be analyzed, and can be minimized for various models.

FleetKey.xlsx

This is an Excel spreadsheet file containing the capacity of each cabin for each Aircraft Model and Subfleet. The data found in this file was based on information provided to us by Delta. In case there are any issues with the data contained in this file, we have stored it as an Excel spreadsheet so values can be easily added or changed.

airport-codes.csv

This is a .csv file that stores the IATA code for all airports, as well as the country where the airport is located and the coordinates of the airport (among many other features we did not use). We use this within the code to include the latitude and longitude of the origin and destination airports as features to our input datasets, as well as the origin and destination countries. We use

this to compute the flight distance and easily identify whether the flight is international or domestic.