Owner	Phase	Name of Notebook	Description of process	Link to Notebook?				
Dave	Data_acquisition	A1_Aircraft_Extract_Join.ipynb	Scrapes data from aerobasegroup.com of all tail numbers; data exploration and filtering of specific mfrs	https://drive.google.com/file/d	/12ZCxpWS9ku6Rm_	UxOxhtcdg9	Tf0bQSoO/view?usp=drive_link	
Dave	Data_acquisition	A2_FAA_Aircraft.ipynb	Attempts to use Selenium to extract aircraft details					
Kuan	Data_aquisition	BTS_scrape						
Kuan	Data_aquisition	Combine_BTS_OnTime_data						
Dave	Data_Preprocessing	Data_Transformation_Pipeline.ipynb	(Recieved after Stacey's combined Weather with Flight Data) Transforms data into labelencoded and one hot encoded features. This should be run on both dev to fit_transform and then run transform on the test set	https://drive.google.com/file/d/1Q8ecSZkV5IJCFSJ81_QT1fG_3Qnyzdwg/view?usp=drive_link				
Stacey	Training Alternate Data Creation	Create_DownsampledUpweighted_Sample.ipy	rr Creating Alternative Training Data Input: *Training_set, parquet Output: *Train DownSample_UpWeight, parquet *Dev_DownSample_UpWeight, parquet *Dev_DownSample_UpWeight, parquet	https://colab.research.google.com/drive/1ifrEHHqrPm4e4RF4tIDthJ5fjimuF7pA?authuser=1				
Stacey	Data_Preprocessing	W1_Make_OnTime_DFW_joining_keys.ipynb	Create the keys for the OnTime data to join with the Weather	W1_Make_OnTime_DFW_join	ning_keys.ipynb			
Stacey	Data_Preprocessing	W2_DetermineUSvsNon.ipynb	Determine if a weather file is for a US location or not.	W2_DetermineUSvsNon.ipynb				
Stacey	Data_Preprocessing	W3_RenameFileToAirportCode.ipynb	Rename Files to Correspond to Airport and Create Link in Weather data to be able to join to OnTime data. This notebook sifts through all the domestic weather stations matching them to the airport stations and adding a column in each file indicating the airport code in order to determine which records correspond to which airport after joining the files.	W3_RenameFileToAirportCode.ipynb				
Stacey	Data_Preprocessing	W4_MakeYearlyWeatherForStations.ipynb	This notebook joins all the airport weather stations for each year into yearly files.	W4_MakeYearlyWeatherForS	Stations.ipynb			
Stacey	Data_Preprocessing	W5_YearlyWeather_ColumnWork.ipynb	Work with Weather Columns - expanding columns, filtering columns, coding missing values	W5_YearlyWeather_ColumnWork.ipynb				
Stacey	Data_Preprocessing	W6_DFW_YrlyWeather_NoDups.ipynb	The OnTime keys I was using were narrowed down to the DFW related airports only and duplicated keys were dropped. I believe a previous step was revised after this and the narrowing down happened previously but this didn't hurt to keep it and double check things were done.	W6_DFW_YrlyWeather_NoDups.ipynb				
Stacey	Data_Preprocessing	W7_Join_and_CreateFinalWeatherData.ipynb	The weather data is joined to the OnTime key data here. This is where a Predicted time is calculated at 45 minutes prior to departure and then the weather is pulled in for each column at the 45 minute mark or prior, up to one hour earlier. It looks for the most recent non missing value in this time frame. Looking at each individual column, one at a time.	W7_Join_and_CreateFinalWe	eatherData.ipynb			
Stacey	Data_Preprocessing	ReviewAllWeatherCols.ipynb	Creates frequencies for every weather column outputting each columns information to individual tabs of an Excel workbook. Along with the frequencies it provides the percentage the value appears out of all Non-Delay rows and the percentage the value appears in all Delay rows. This was used to determine which columns to keep and which to drop.	"/content/drive/Shareddrives/STUDENT-Capstone SS23/Weather_Processing_Notebooks/ReviewAllWeatherCols.jpynb"				
Stacey	Data_Preprocessing	FlightWeatherOrig_2019_Freqs.xlsx	Not a notebook but wanted to list this somewhere. This is the output from ReviewAliWeatherCols.ipynb. This was used to review the data in each weather column to help narrown down which weather columns we should keep since we felf there were too many.	"/content/drive/Shareddrives/STUDENT-Capstone SS23/Weather_Processing_Notebooks/FlightWeatherOrig_2019_Freqs.xisx"				
Stacey	Joining	Join_Weather_To_DFW_OnTime_TRAIN.ipynt	Joining the Weather Data to the OnTime Training data. The Aircraft data was joined prior.  Reads in: OnTime DPM Imputed.csv OnTime Weather_DFW_ORIG_10_19.parquet OnTime_Weather_DFW_DEST_10_19.parquet Outputs: OnTime_w_Weather_10_19.parquet OnTime_w_Weather_10_19.parquet	https://colab.research.google.com/drive/1KK1urcBqC3Fp7xplkKMeXNKq0EqZdcdc?authuser=1				
Stacey	Joining	Join_Weather_To_DFW_OnTime_TEST.ipynb	Joining the Weather Data to the OnTime Test data. The Aircraft data was joined prior.  Reads in ONT_DF_2023_DFW_Filtered_Imputed.csv OnTime_Weather_DFW_ORIG_23.parquet OnTime_Weather_DFW_DEST_23.parquet Outputs: OnTime_Weather_23_mg_cots_parquet	https://colab.research.google.com/drive/1si-qXUmfpnYcEl0y91jl8nS_hnwzNsE8?authuser=1				
Dave	Models	Boudia_NN_Model.ipynb	Creates the Neural Network Model; Fits through GridSearch, Review Performance, Save off Model	https://drive.google.com/file/d/1goU-IAihQZzeZP5KL_BVyr311PrYmlxp/view?usp=drive_link				
Dave	Models	Load_NN_Model.ipynb	This gives details about how to load the saved pipeline and h5 file (there are a couple of file dependencies)	https://colab.research.google.com/drive/16BoSxrBirGRYDqpBqNqhJXu1sJ-6aWdi?usp=drive_link				
Stacey	Models	LR_Classifier.ipynb	Creates the Logistic Regression Model	LR_Classifier.ipynb				
Stacey	Models	RF_Classifier.ipynb	Creates the Random Forest Classifier Model	RF_Classifier.ipynb				
Dave & Stacey	Models	Test_Models.ipynb	Model Evaluation of all Machine Learning Classifiers	Test_Models.ipynb				
Kuan	Models	NetworkAnalysisVis_FlightLevel						
Kuan	Models	Network_Flight_SIR_Rates						