Uber vs Lyft Surge Prices based on temperature and rain

* The Data
  + The datasets we used were 2 CSV files. The first file provided the number of cab rides over about a 20 day time period with the cab-service type, surge multiplier, destination and a time-stamp. The second file provided the temperature, rain, wind and time-stamp.
* The Transformation
  + The first area we had to transform after importing the files into pandas as a DataFrame was the time-stamp columns in both CSV files. The time-stamp was provided in an epoch format. We used a date\_time function to convert the time-stamp into date and time. The DateTime information was put into a new column in the data frame.
  + Our next transformation task was to bin the time-stamp column into 2 separate columns by date and a specified time range. We used the below code to bin the time by typical rush hour timeframes, 2 mid-day timeframes and an overnight timeframe and extracted only the date from the DateTime column into a new date column.

def bin\_f(x):

if x.time() > dt.time(6) and x.time() < dt.time(10):

return "6AM - 10AM"

elif x.time() > dt.time(10) and x.time() < dt.time(15):

return "10AM - 3PM"

elif x.time() > dt.time(15) and x.time() < dt.time(19):

return "3PM - 7PM"

elif x.time() > dt.time(19) and x.time()< dt.time(22):

return "7PM - 10PM"

else:

return "10PM - 6AM"

cab\_rides\_df["time\_of\_ride"] = cab\_rides\_df["DateTime"].apply(bin\_f)

cab\_rides\_df['date'] = cab\_rides\_df.DateTime.dt.date

* + The surge price multiplier was not separated by provider so we had to create 2 new dataframes based on the cab service provider. We filtered the original dataframe by Lyft and Uber to extract the surge multiplier for each provider. We grouped the data by date, time and location to create our new dataframes and took the mean of the surge prices for those data points. We merged the new ride dataframes by the date, time and location to create 1 dataframe. We did the same process for the original weather dataframe, and took the mean of the rain, temp and wind.
  + We merged the new ride and new weather dataframes into one dateframe for further analysis. We merged on the date, time and location columns.
* The database
  + A new database and table were created in postgres for the final dataframe to be imported.
  + We made a connection to the newly created database and confirmed the connection by pulling the table name.
  + We pushed the dataframe to sql and imported the information to the summary table.
  + Confirmed the connection by reading the table.