Mighty Pandas

Lorenz Mateo
Noe Torres
Justin Wong
Christian Chu

Motivation & Summary

- Hypothesis: Quantity of Uber and Lyft requests is inversely correlated with temperature.
- It should be the strongest correlation compared to other weather factors.
- How does temperature affect the amount of rides requested?
- An analysis on transportation during less ideal weather conditions may be beneficial for marketing and recruitment during higher demands in specific season

Questions & Data

Are there any seasonal patterns or are the rates reflective only of the weather for the day?

What is the relationship between number of rides requested and Precipitation?

How are uber & lyft ride requests impacted during the warmer / colder seasons?

Are there patterns of exponential growth with Uber and Lyft rides through years and months?

Data Exploration

Originally planned to leverage UBER and Lyft API's but discovered they reserve data on number of ride requests. API specifically designed for application integration

Discovered a database, NYC open data. Adjusted the scope of our research to New York City but provided free data on UBER and Lyft ride requests filtering by "Base Name"

- Did not anticipate the limitations around Lyft data which only began tracking in late 2015
- Data is limited by month. Not much granularity

openweathermap.org provided no historical data. We signed a free trial with WeatherSource and pulled data from the last 5 years

- Data can only be pulled by day and hour
- Weather Source imposed data request limits per minute

Data Cleanup

Open NYC data was filtered by company name “UBER and Lyft”. Which generated an aggregate report of rideshare requests by month

Weather Source provided a select list of conditions

- Temperature (F), Humidity (%), Snowfall Depth (in), Wind Speed (MPH), Precipitation (in), Max Temperature (F), & Max Wind Speed (MPH)

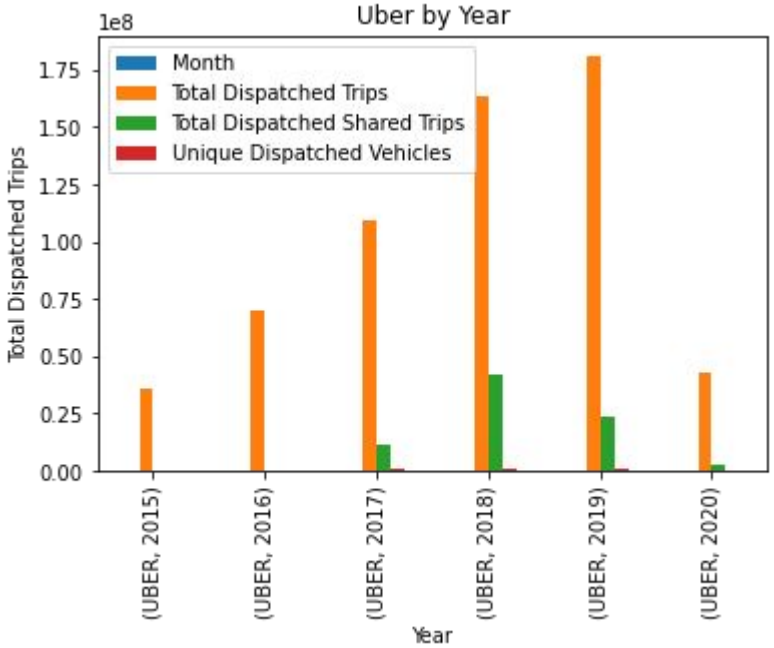
Cleaning the data

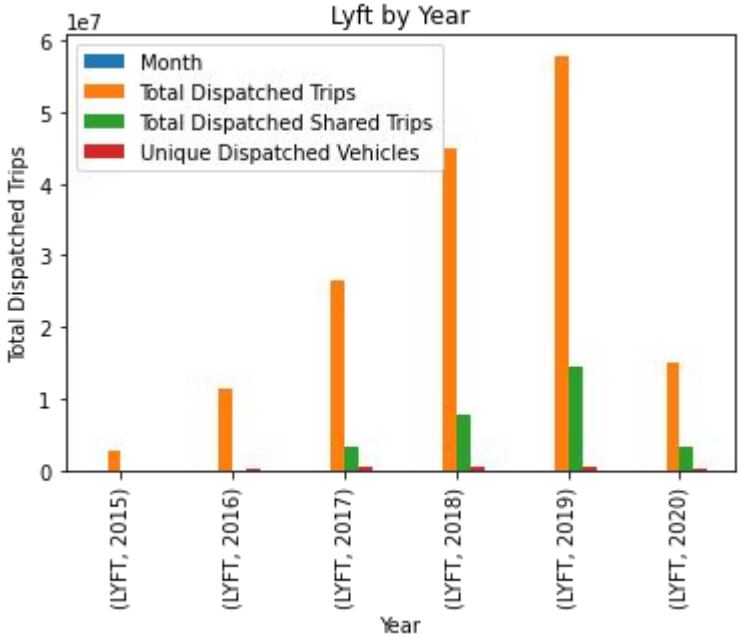
- Convert ISO 8601 timestamp to datetime
- Multiple API calls for specific date ranges
- Merged data for past 5 years
- Merged weather data with UBER / Lyft

Data Analysis

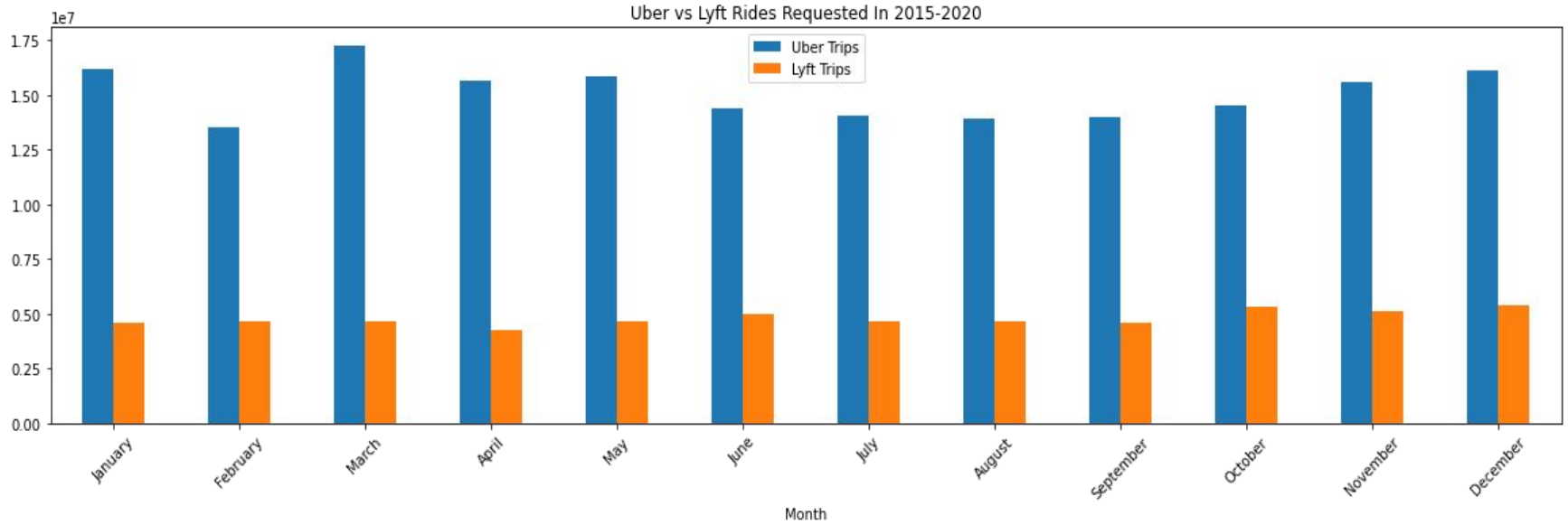
- NYC Uber/Lyft Data:
  - Year to year:
    - Both have experienced steady increase in rides since their starts
      - With Uber reporting much higher numbers.
  - Month to month:
    - For Uber there was a slight dip towards the end of winter (February).
      - Other than that it's been a steady increase.
      - No data for the back half of 2020 leading to the appearance of a dip in the ride requests for Uber.
    - Lyft has had little growth with various fluctuations.
- Weather Source Data:
  - Many outliers for snowfall and precipitation given that the majority of the data for both was 0 in.
  - Humidity was the most consistent across seasons.
  - Wind speed was mostly consistent across all seasons with the exception of a slight dip in the summer.
  - Temperature had the most change across seasons.
- Rides vs. Weather:
  - All had slightly positive correlations with the exception of temperature, which had a slightly negative correlation.
  - Most of the data was not statistically significant with p-values well above 0.05.

Uber and Lyft Demand 2015 - 2020

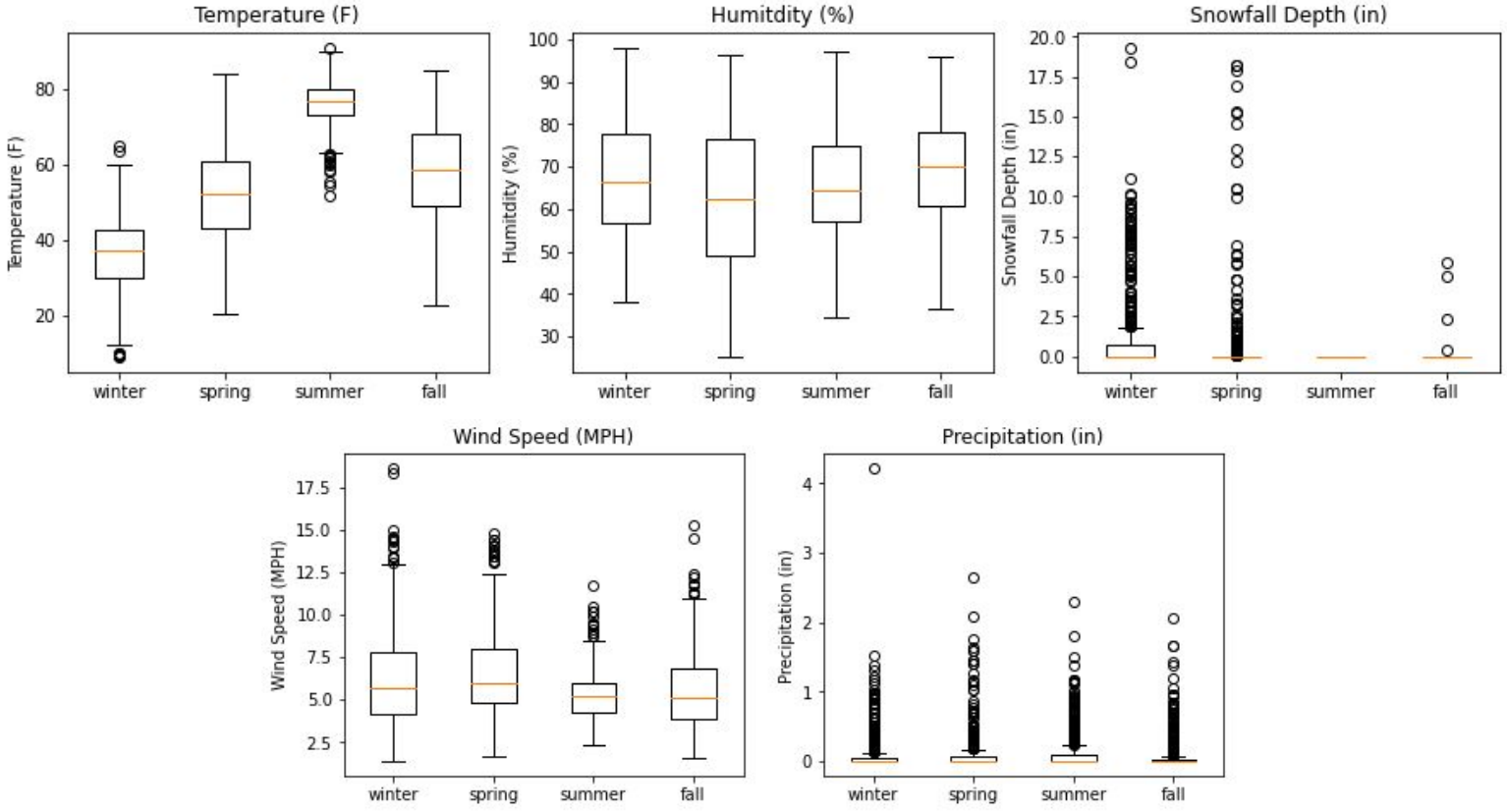




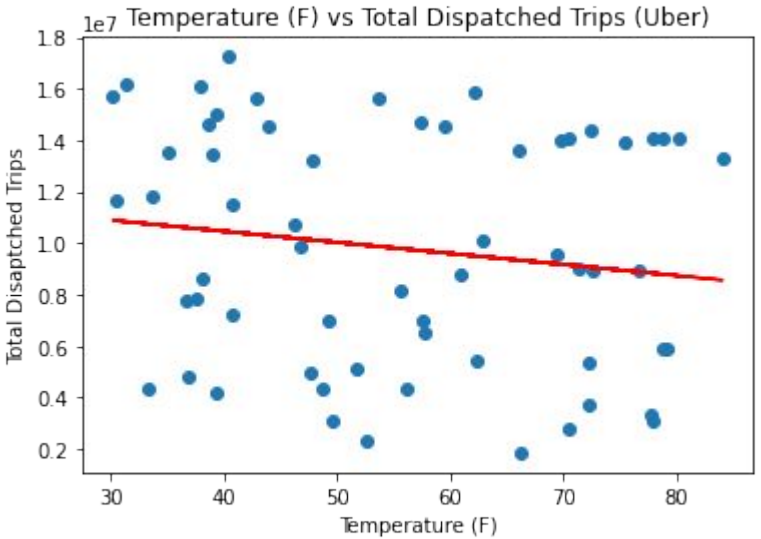
Ride Requests by month: 2015-2020



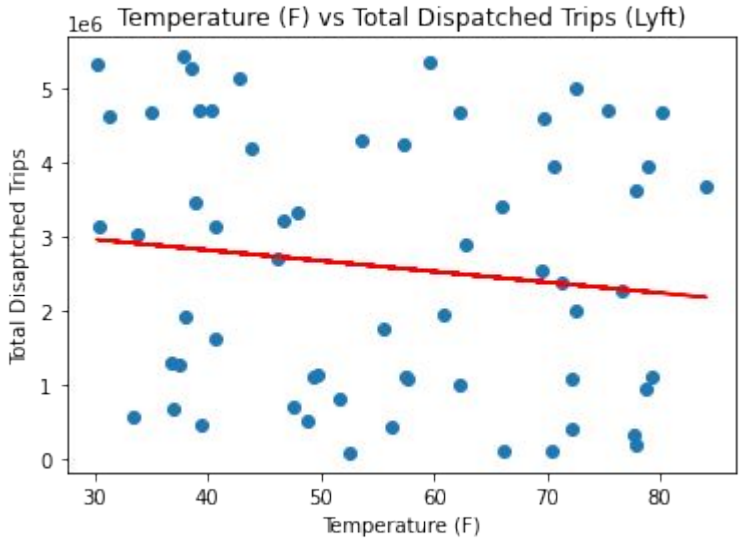
Boxplots of Weather Attributes 2015-2020



Rides and Temperature Uber/Lyft 2015-2020

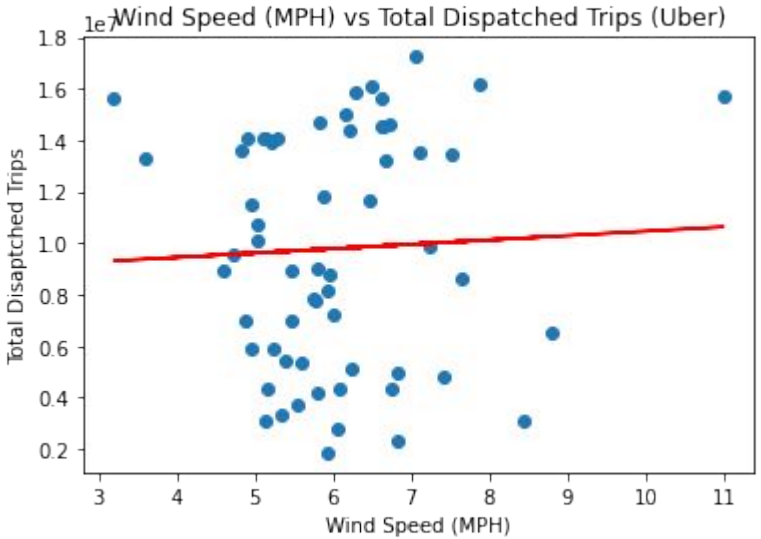


Correlation = -0.15 P-Value = 0.24

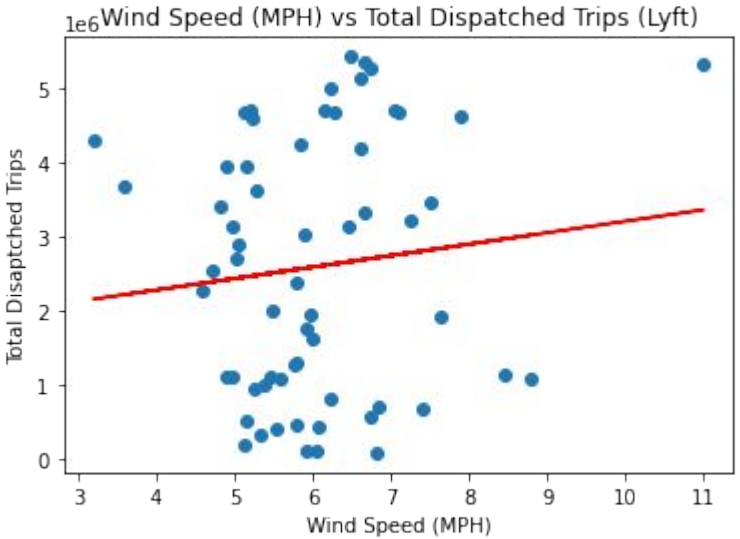


Correlation = -0.13 P-Value = 0.31

Rides and Wind Speed Uber/Lyft 2015-2020

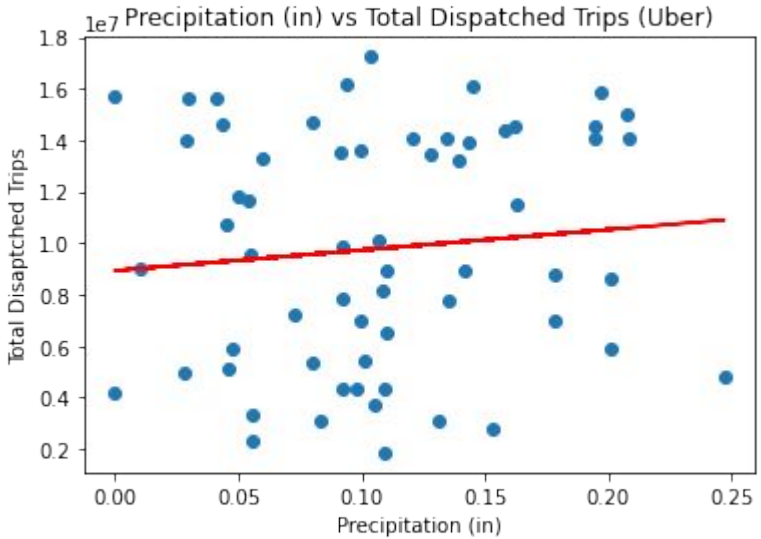


Correlation = 0.05 P Value = 0.73

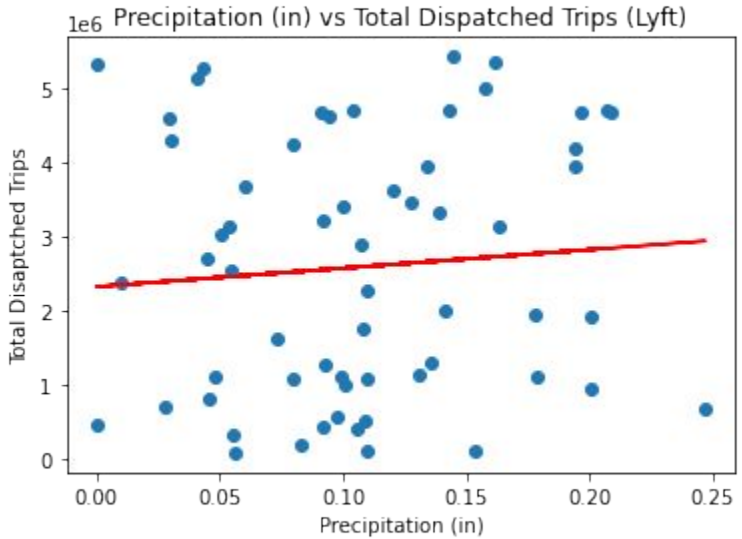


Correlation = 0.11 P Value = 0.4

Rides and Rainfall Uber/Lyft 2015-2020

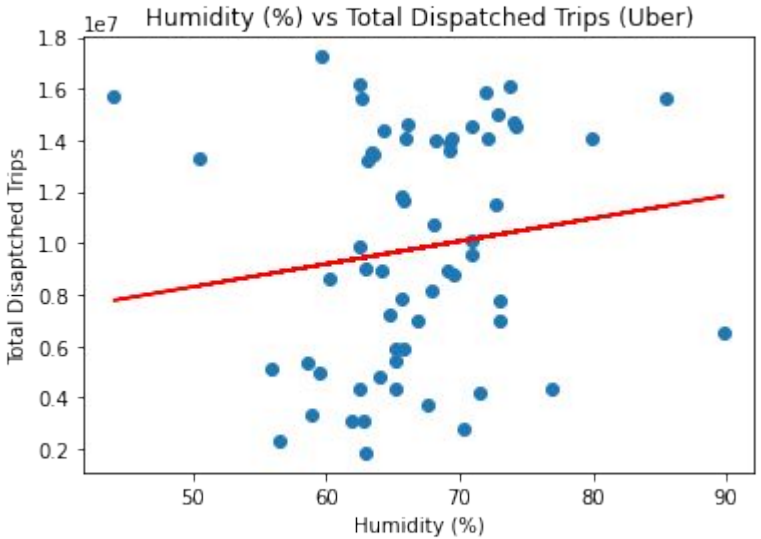


Correlation = 0.1 P Value = 0.44

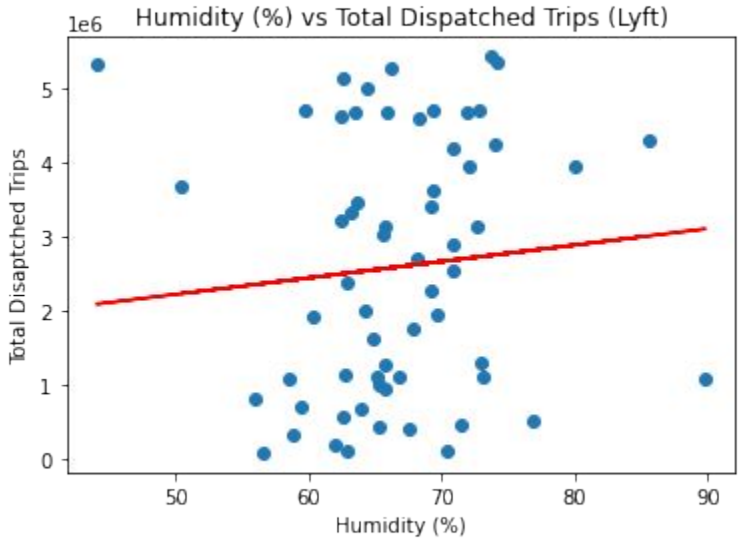


Correlation = 0.08 P Value = 0.53

Rides and Humidity Uber/Lyft 2015-2020



Correlation = 0.14 P Value = 0.28



Correlation = 0.09 P Value = 0.48

Discussion

- Our original hypotheses were proven likely false, as our data showed no significant correlation between weather factors and ride requests.
- Possible reason for slight trends: during “good weather” months, people could be less opposed to walk/be outside in nice weather.
- Other non weather factors likely more significantly correlated: surge pricing, seasonal discounts, etc.
- Dispatch explosive upward trend: Uber and Lyft foothold in the ride dispatch market.
- Possible correlation between high wind speed and dispatch demand - not enough data.
- 2020 ride request trends - greater than that of 2015-2016 in January and February.

Post Mortem

- Discuss any difficulties that arose, and how you dealt with them
  - Weather API troubles and Uber/Lyft data set limits
- Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?
  - Limitations of the data: specific location data.
  - Researching alternative modes of transportation and how they are correlated with weather.
  - Alternate transport: Uber and Lyft bike/scooter sharing in 2019.
  - Generate a heat map with number of rides requested in a given area.
  - Other weather factors, and combination of factors

Questions!

Ask Us Anything!