# **Project Title**

Impact of weather-related changes on ride share services in New York City

# **Team Members**

Mark Hirschenberger

Tomorrow Rose

Zach Herington

# **Project Description/Outline**

Using census and weather, how are different ride share services impacted.

# **Research Questions to Answer**

1. What areas of the map are most affected by weather changes?
2. Ride cancellations by month?
3. What is $/mi across all ride share services?

# **Datasets to Be Used**

## Ride Share Sources

* 1. https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city
  2. https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices
  3. https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq

## Weather Data

* 1. <https://openweathermap.org/api>
  2. <https://www.ncdc.noaa.gov/cdo-web/datatools/records>
  3. <https://dev.meteostat.net/api/point/daily>

## Census Data

* 1. https://www.census.gov/data/developers.html

## Alternative Sources

* 1. https://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance/

# **Rough Breakdown of Tasks**

## Scope

1. What boroughs?
   1. Manhattan
2. What ride share services?
   1. Yellow Cab
3. What year?
   1. 2019

## Data Handling

1. Import Taxi data (complete, still need Feb-Jun 2019)
2. Import weather data (complete!!)
3. Import Uber data (open, still need source)
4. Merge data sets by date (not started)
5. …

## Visualizations

1. What time frame during the day do we care about looking at?
2. What public events could skew the anticipated decline in ride share volume?
3. Do we focus our comparison on geographical areas or by ride share service?
4. Weather Events (refer to alt source for deciding these weather events)
   1. Snowfall: 4+ inches
   2. Rain: 0.5+ inches
5. Create bins for “weather event” and “non-weather event”
6. Avg ride share volume by borough on “non-weather event” days
   1. Bar chart of each volume (total NYC rides)
   2. Pie chart of relative volume (% of total NYC Rides)
7. Avg ride share volume by borough on “weather event” days
   1. Bar chart of each volume (total NYC rides)
   2. Pie chart of relative volume (% of total NYC Rides)
8. Show percent decline of rides for each borough given a “weather event”
   1. Bar chart
   2. Heat Map

## Weekend tasks

Orange text = dataframe

1. Weather Data bar graphs (Tomorrow)
   1. Create a new dataframe for
      1. Top 10 Rain days as rain\_df
      2. Top 10 Snow days as snow\_df
      3. Top 10 Wind days as wind\_df
      4. Top 10 Max Temp days as tmax\_df
      5. Top 10 Min Temp days as tmin\_df
      6. All Top 10s removed as normalweather\_df
2. Taxi data visuals (Mark)
   1. Import lats and longs for pick zones
   2. Stat test for each month’s sample to make sure p-value is < 0.05.
      1. Might need to increase sample size to make this happen
   3. Ride volume by day
   4. Ride volume by zone
   5. Any other metrics you want
3. How to merge weather and taxi data (Zach)
   1. Merge Options:
      1. Merge on Lat, Lng (pros vs cons)
      2. Merge on Date (pros vs cons)
   2. What analysis will answer our hypothesis?
      1. Ride volume for each Top 10 weather categories
      2. New df that shows the change in ride volume for Top 10s vs normal weather days
      3. Line graph showing these changes
4. Ideas to consider:
   1. Natural disasters

## Weekend tasks

1. **Intro**
   1. ..
2. **Motivation and summary slide**
   1. Why we asked these questions?
   2. Were able to answer?
3. **Datasets (Taxi’s, Other Rideshare Services, Weather Data)**
   1. What kinds of data were needed?
   2. Where did we get the data?
4. **Data Cleanup & Exploration**
   1. …
5. **Data Analysis**
   1. …
6. **Discussion**
   1. …
7. **Post Mortem**