

MINING FOR RIDESHARE PATTERNS

Ronald Thompson
Jonathan Prindle
Saloni Sharma

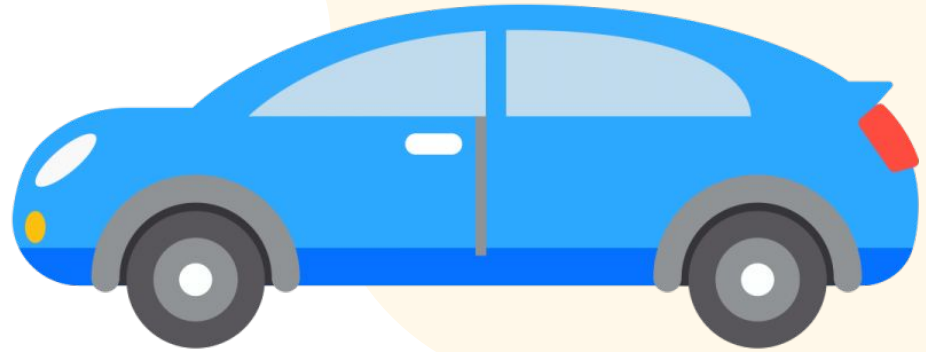




TABLE OF CONTENTS

01

QUESTIONS

02

DATA PREPARATION

03

TOOLS

04

NEW YORK CITY ANALYSIS

05

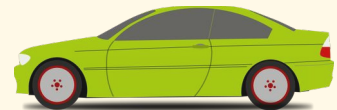
CHICAGO ANALYSIS

06

APPLICATIONS

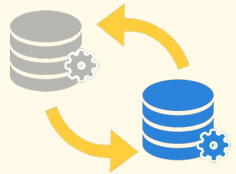
Questions

- **Overall goal:**
 - Reveal rideshare patterns that can improve customer experiences and optimize profits
- **Understand the patterns that rideshares take**
 - Are there recurring patterns that show when peak traffic is hit?
 - What are the average duration of rideshares at different times?
 - When are the better and worse times to take rideshares?
 - When do drivers hit diminishing returns?
- **How is the data affected by why rideshare is used?**
 - What are the most common areas in a city that rideshares start from and end at?
 - Do the trips demonstrate people going to work or to do other activities?



Data Preparation

- Data processing
 - Cleaning:
 - Need to clean the various cities datasets and standardize the fields
 - Remove any unnecessary or outlier values
 - Transformation & Reduction:
 - Transformed latitude and longitude data to plot maps
 - Reduced Uber NYC data set to include only April - June, to more accurately compare the same months from both years 2014 and 2015
 - Transformed time data for analysis of peak times
 - Integration:
 - NYC data set 1: Combine 2014 Uber files into one for NYC, and compare with 2015 Uber data
 - NYC data set 2: Combine monthly data to get all of 2015 NYC data



Tools Used

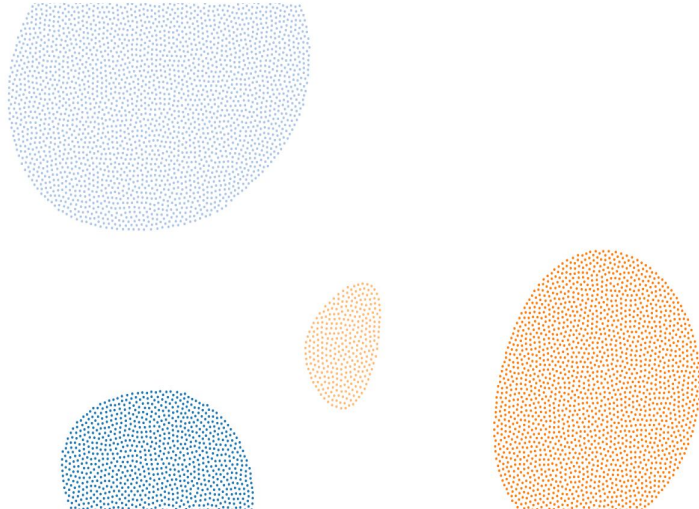
- **Programming languages/Applications:**
 - Python (via Jupyter Notebook)
 - Pandas, Matplotlib, NumPy
 - GeoPandas, GeoPlot, MapClasssify
 - Sklearn
 - Spark via Jupyter Notebook
 - RapidMiner
 - Qlik
- **Data:** three datasets in .csv files
 - All downloaded to each student's computer
- **Code:** storing and sharing code through GitHub



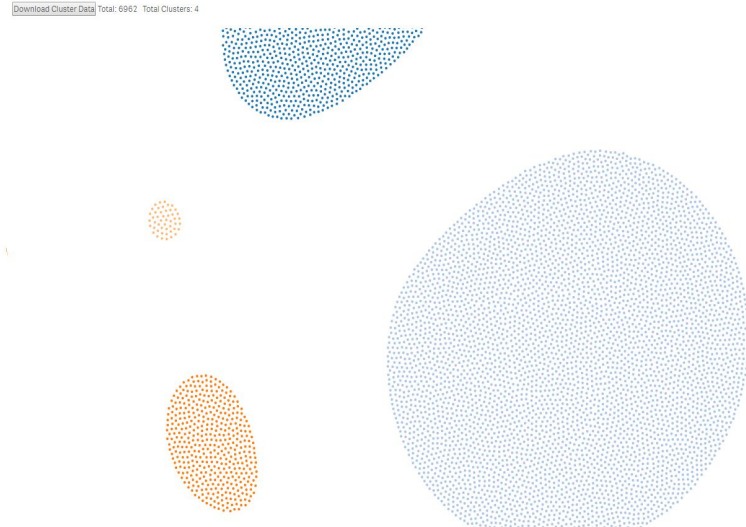
Maximizing the Experience in New York City

- Tips are clustered more on the lower end (light blue).
- Fare is more evenly distributed amongst the clusters.
- Fare in relation to distance was typical, larger distance travels means larger fare.
- Tip in relation to distance showed discrepancies, larger distance traveled didn't consistently mean a higher tip.

K-means Fare Cluster



K-means Tip Cluster



Ideal Target Range for Drivers in New York City

The Good

7 passengers: \$7.008/per mile.
Tip: \$.666/per mile.

8 passengers: \$6.629/per mile.
Tip: \$.633/per mile.

9 passengers: \$6.39/per mile.
Tip: \$.580/per mile

The Bad

1 passenger: \$5.44/per mile.
Tip: \$.586/per mile.

5 passengers: \$5.22/per mile.
Tip: \$.548/per mile.

6 passengers: \$5.27/per mile.
Tip: \$.566/per mile.

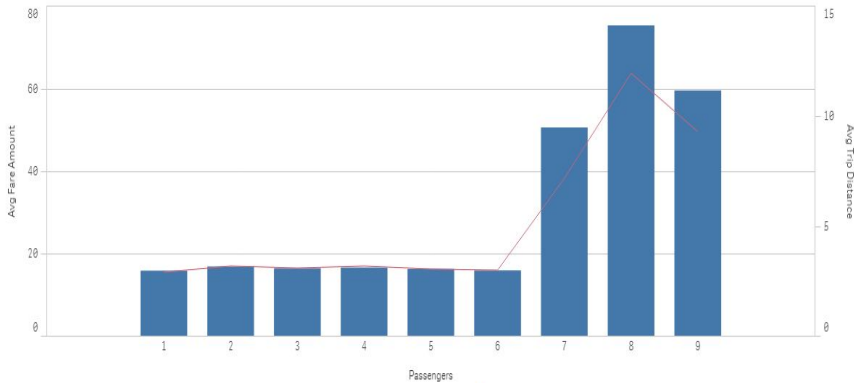
The Ugly

2 passengers: \$.5.24/per mile.
Tip: \$.531/per mile

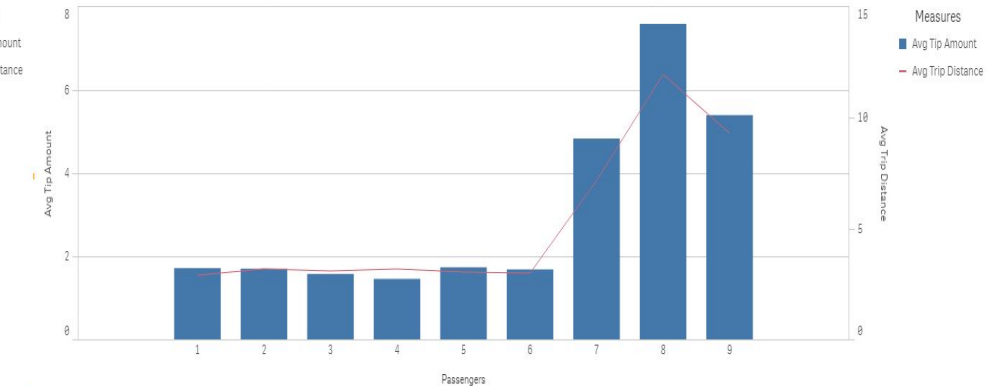
3 passengers: \$.5.26/per mile.
Tip: \$.516/per mile

4 passengers: \$.5.15/per mile.
Tip: \$.468/per mile

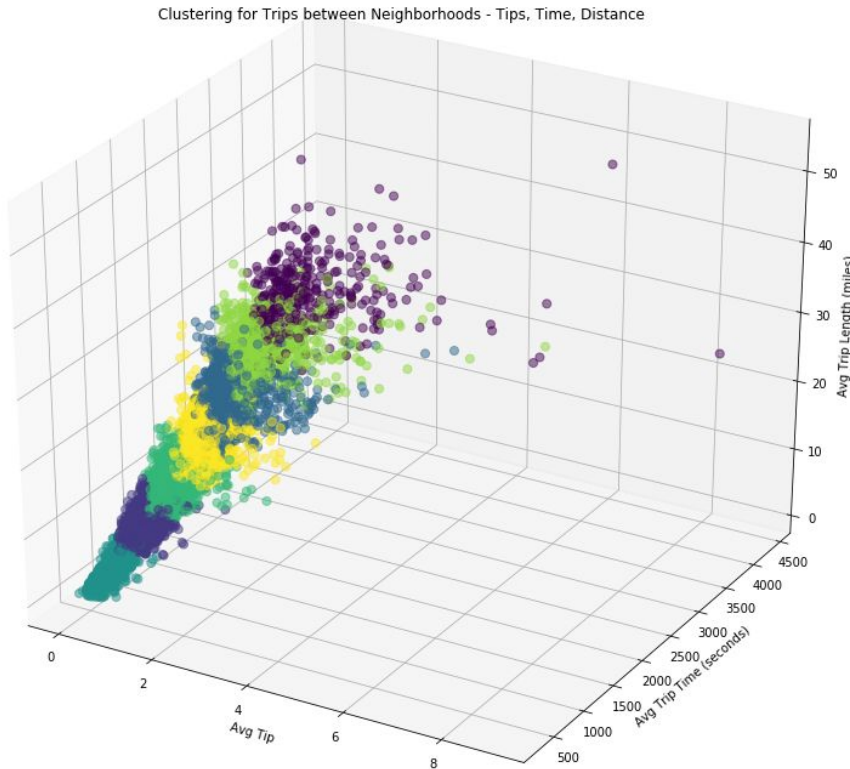
Passenger Averages for Fare/Distance



Passenger Averages for Tip/Distance



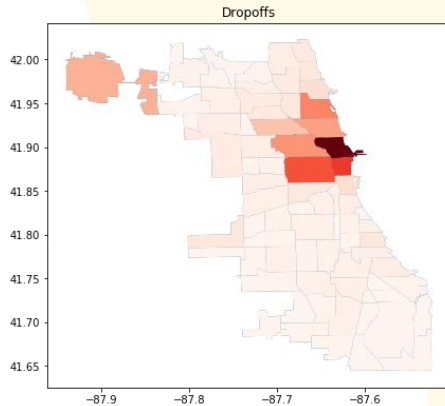
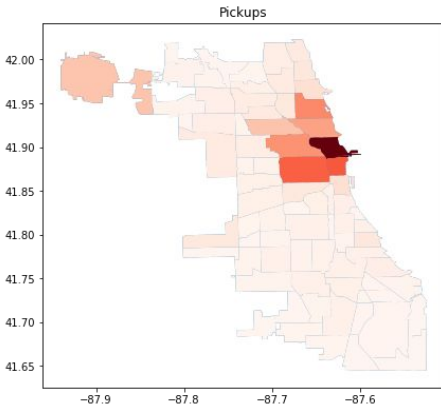
Common Characteristics of High Tip Trips



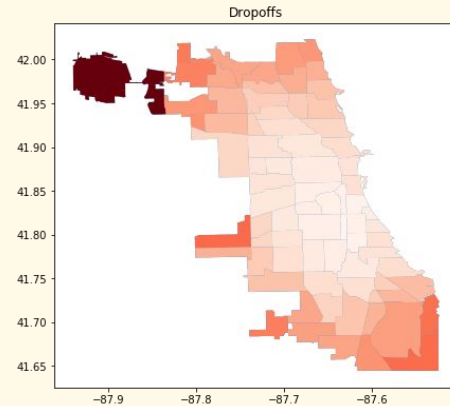
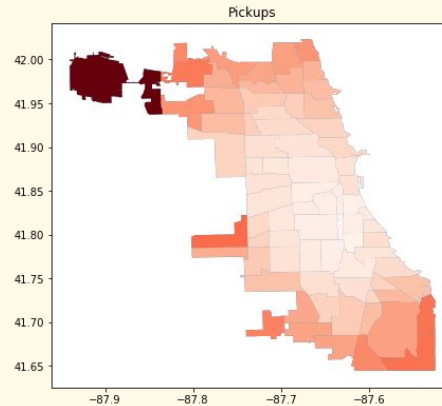
- Tips are generally determined by length of trip (time and distance)
- Distance is a bigger determining factor than time
- Tips on average don't exceed \$8, which means that there's a diminishing return for a driver looking to get tips
- Short trips (those that are under 10 miles and under 10 minutes) usually don't get tips
- Trips that are over 20 miles lead to the larger tips on average, potentially due to passengers feeling obligated based on distance.
- Trips that are going to the airport are more likely to have tips (and are generally longer).

Best Locations for Trips in Chicago

Number of trips



Total fare

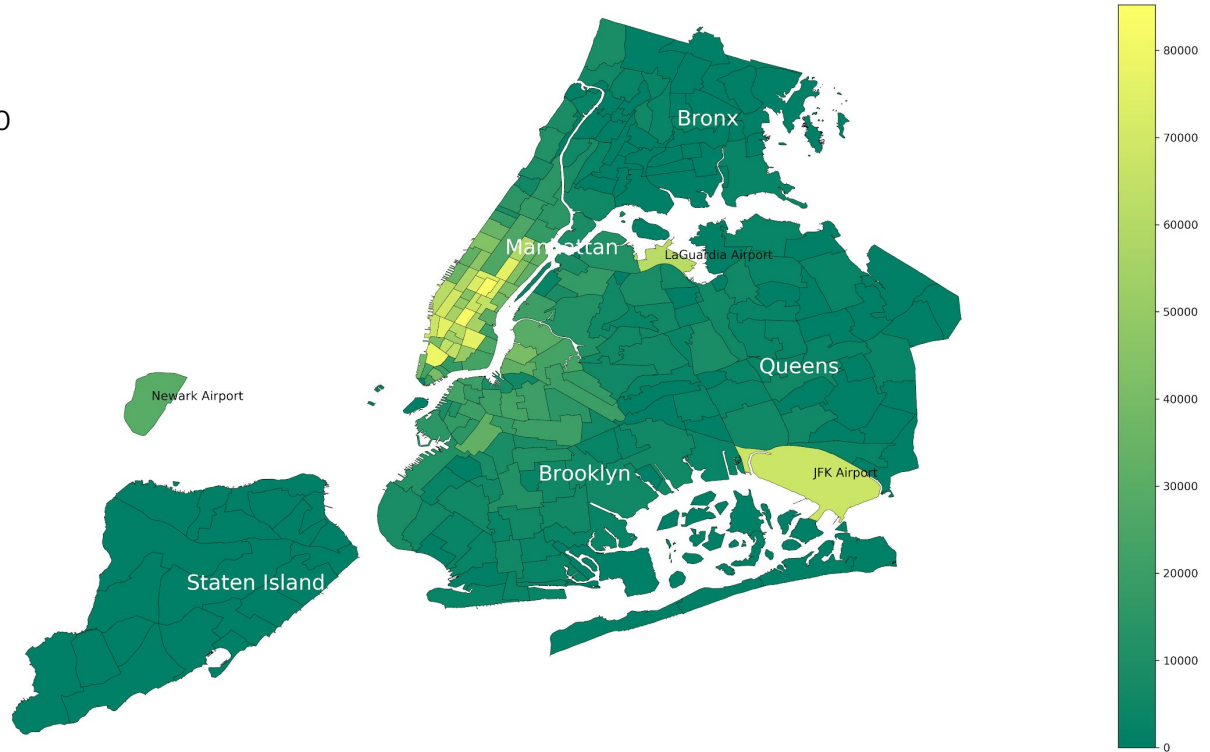


- Trips to and from the airport generate the largest fares followed by the outer suburbs
- The most pickups occur on the “Central Side” of Chicago and the West Side areas that are close to the center of Chicago. Airport pickups and dropoffs are common as well
- Optimal location for drivers is to remain in Downtown Chicago (Near Northside and Loop). This maximizes the number of trips and the area that is most likely to lead to an airport trip
- Additionally because trips are so short for Downtown Chicago these have some of the higher fare per miles other than the outer suburbs

Best Locations in New York City

- Largest amount of rides start in Manhattan and near airports
- Places where people are likely to go without their own vehicles and need transportation
 - Tourist attractions
 - Downtown - Manhattan
 - Airports

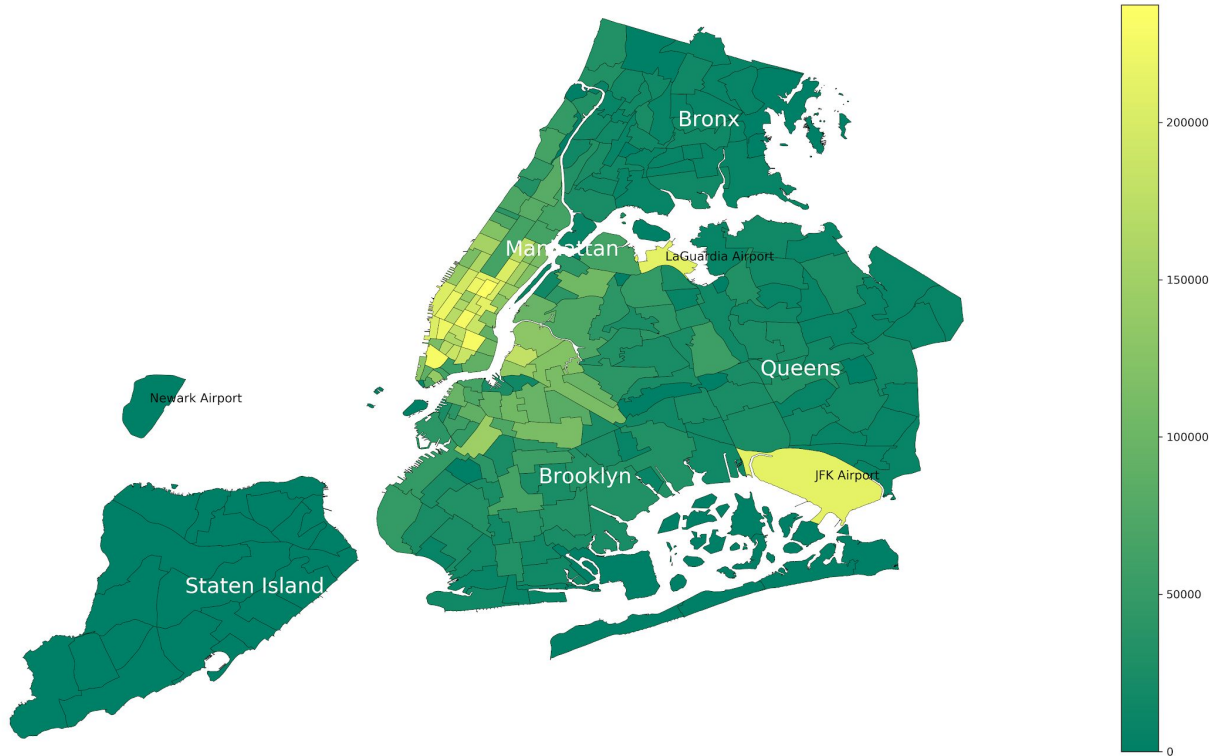
NYC Uber 2014: Pickups per Neighbourhood



Best Locations in New York City

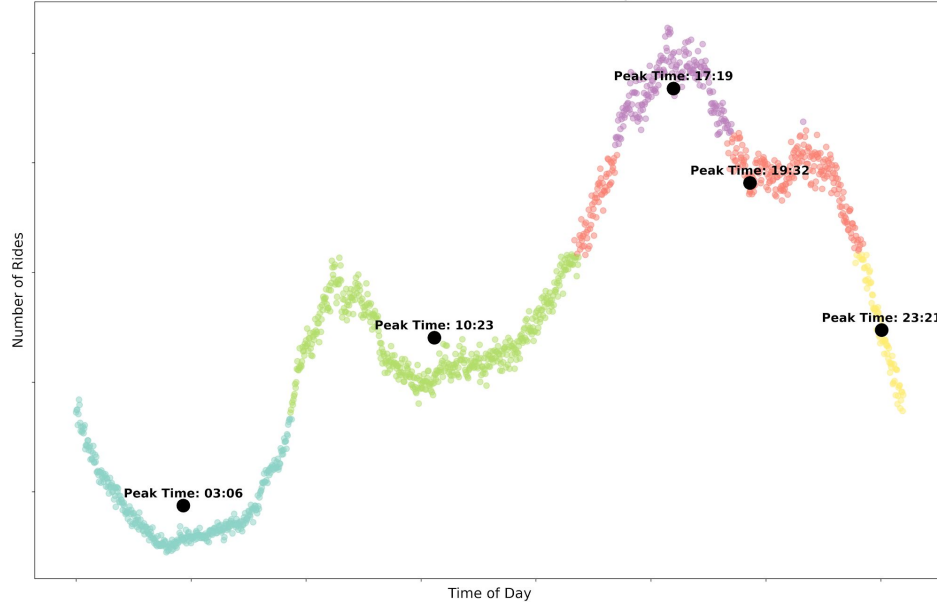
- Similar results from both data sets (2014 & 2015)
- Locations with most rides:
 - Tourist areas
 - Downtown - Manhattan
 - Airports

NYC Uber 2015: Pickups per Neighbourhood

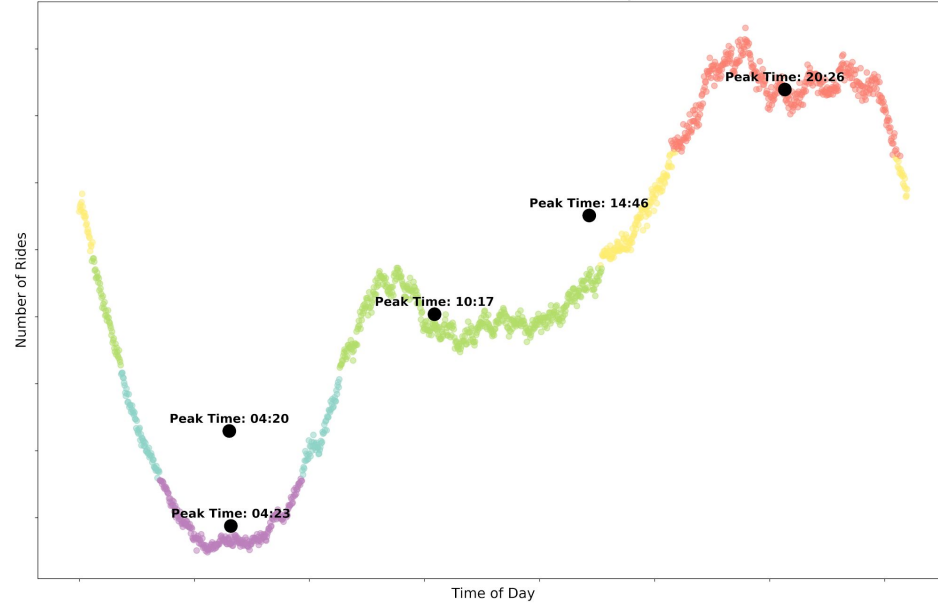


Best Times of Day for Trips in New York City - Overview

NYC Uber 2014: Peak Times of Day



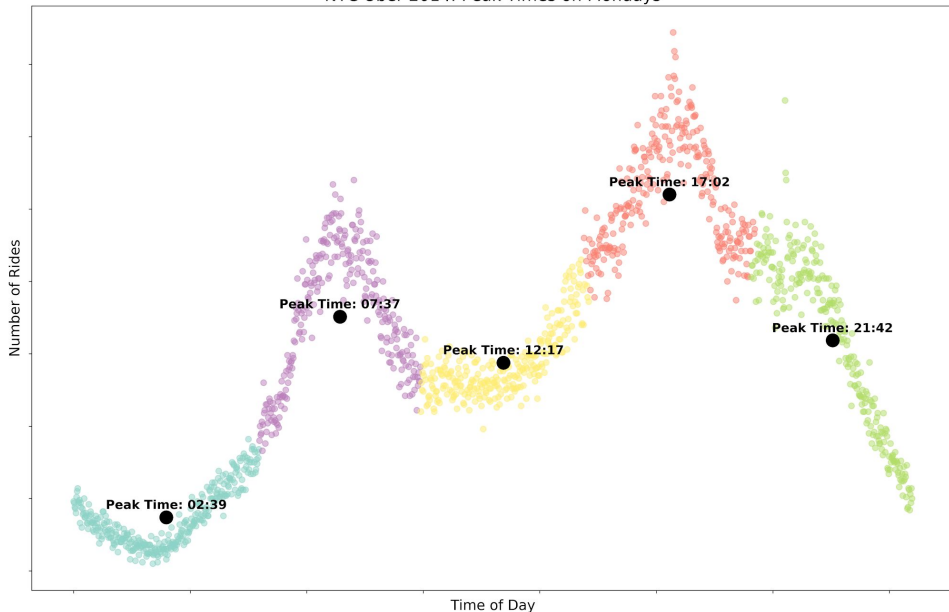
NYC Uber 2015: Peak Times of Day



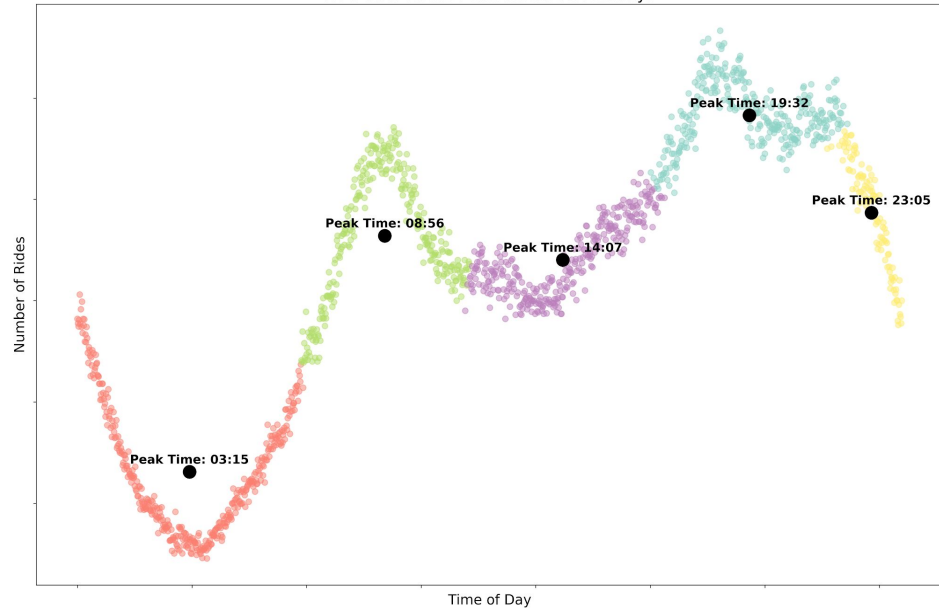
- K-means clusters of all time data (April - June) and the number of rides at that time
- Peak times are skewed by outliers
- Different peak times on different days of the week

Best Times of Day for Trips in New York City - Mondays

NYC Uber 2014: Peak Times on Mondays



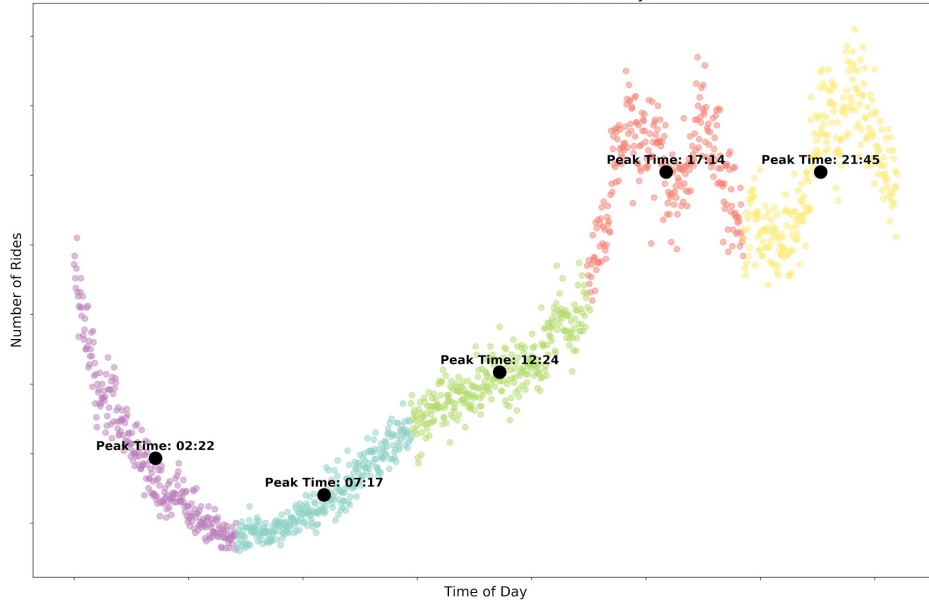
NYC Uber 2015: Peak Times on Mondays



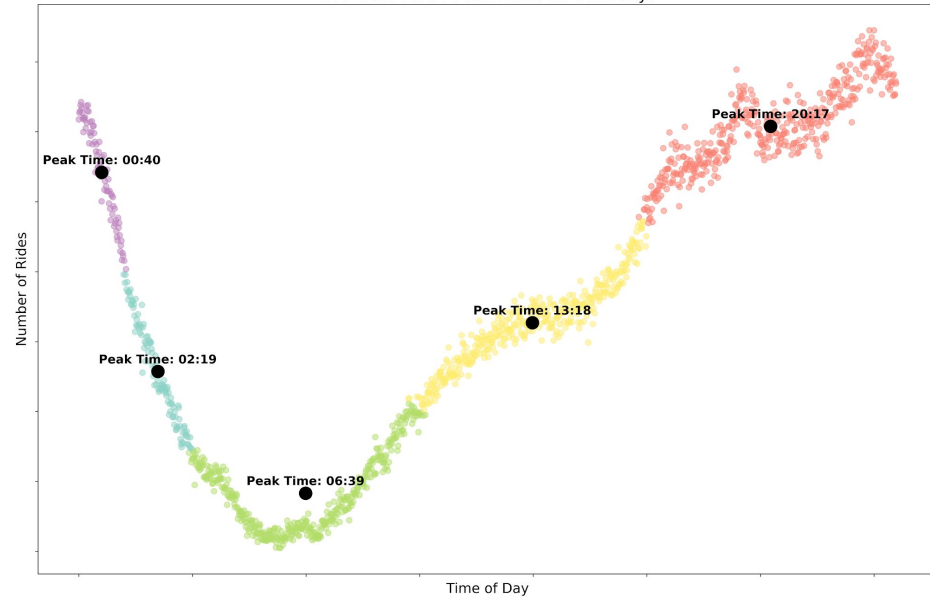
- K-means clusters of times on Mondays
- Somewhat similar to the clusters for entire data, but fewer outliers and more accurate peak times
- Peak times occur when people travel to and from work on Mondays

Best Times of Day for Trips in New York City - Saturdays

NYC Uber 2014: Peak Times on Saturdays



NYC Uber 2015: Peak Times on Saturdays



- K-means clusters of times on Saturdays
- Considerably different results than Mondays and the overall clusters
- Peak times at night when people are socializing on weekends

Applications of Our Analysis

- Combining peak times and most popular locations:
 - Our analysis can help rideshare driver's find the peak times on different days of the week and aid in determining the best times of day to expect customers
 - Drivers can also use the location analysis as a way to establish where they are most likely to find customers
- Fares and tips

Overall there was a serious lack of tipping in both NYC and Chicago. This is consistent with the anecdotal evidence that we collected. During our analysis we were able to identify areas and circumstances where drivers could maximize their fares and tips.

