**IDS 561 – Analytics for Big Data**

**Final Report – Divvy Data Analysis**

**Problem Setting and Background:**

Divvy is Chicagoland’s bike share system, with 6,000 bikes available at 570+ stations across Chicago and Evanston. Divvy provides residents and visitors with a convenient, fun and affordable transportation option for getting around and exploring Chicago.

Divvy, like other bike share systems, consists of a fleet of specially designed, sturdy and durable bikes that are locked into a network of docking stations throughout the region. The bikes can be unlocked from one station and returned to any other station in the system. People use bike share to explore Chicago, commute to work or school, run errands, get to appointments or social engagements, and more.

Divvy is available for use 24 hours/day, 7 days/week, 365 days/year, and riders have access to all bikes and stations across the system.

In this Project, we have conducted a series of exploratory data analysis and Modelling on Chicago Divvy bicycle sharing data. The goal of this project is to:

* Visualize the bicycle sharing data
* Try to find some interesting patterns and
* relations behind the data
* Business Optimization Suggestions

**Data Description:**

We are using the Divvy 2018 bike data that we retrieved from <https://www.divvybikes.com/system-data> The original data includes:

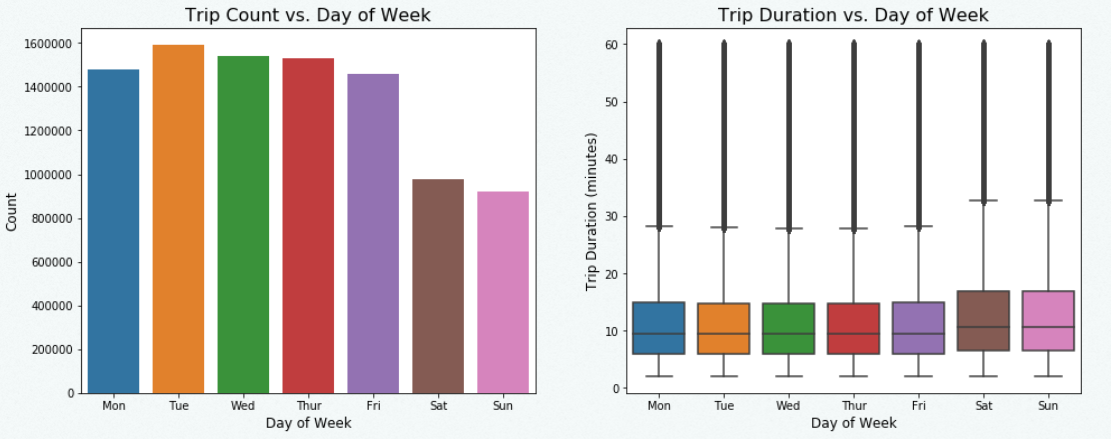
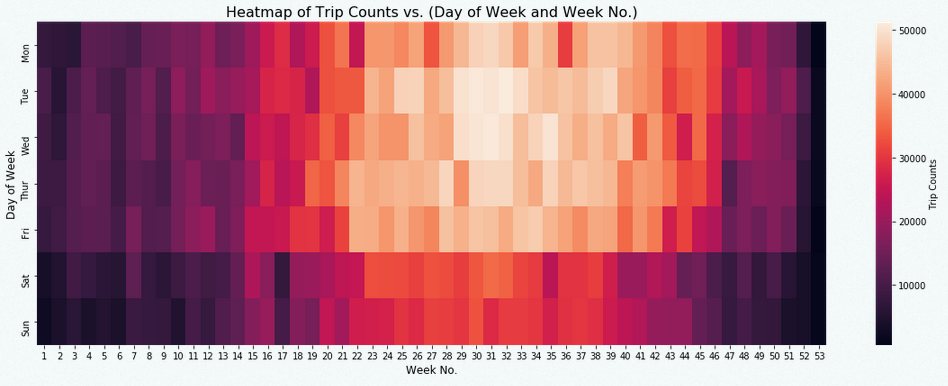
* Trip start day and time
* Trip end day and time
* Trip start station
* Trip end station
* Rider type (Member, Single Ride, and Explore Pass)
* If a Member trip, it will also include Member’s gender and year of birth

The weather information from wunderground.com. It contains weather information for the given area.

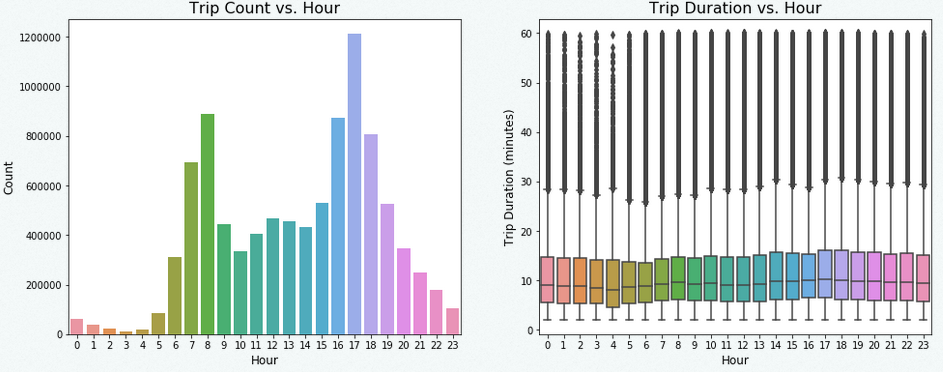
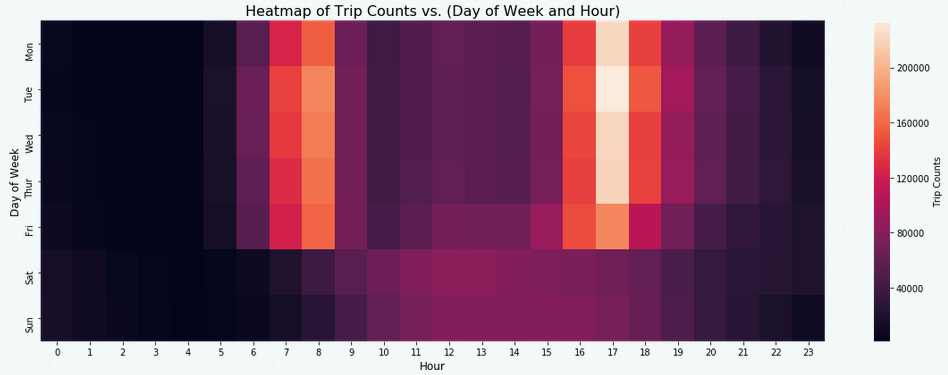
**Techniques:**

After initial exploration, we deleted the data with missing values and only used the data that have trip duration from 2 minutes to 1 hour. For the weather information, we only kept the temperature and events.

After cleaning the data, we conducted a series of data visualization.

### Trip Duration Vs Days of Week Heat map of Trip Counts vs. Day of Week and Week No.

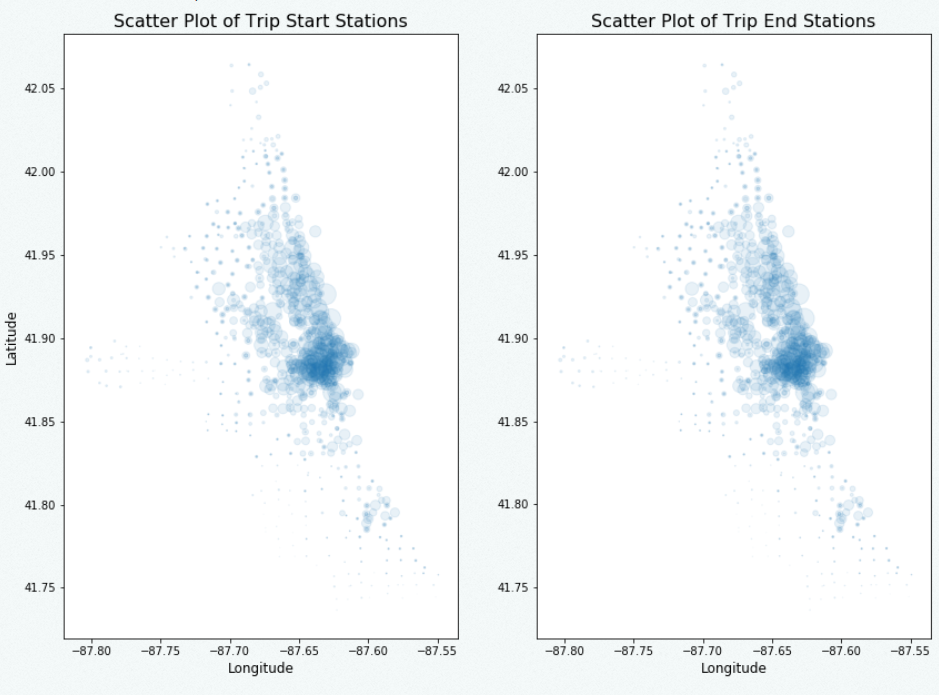
### Trip Duration Vs Hour Heat map of Trip Counts vs. Day of Week and Hour

### 

### Trip Duration vs Customer Category Trip Duration vs Gender

### 

**Trip Duration vs Weather Trip Distribution Duration**



**Scatter Plot of Trip Start Stations and End Stations**

**Results:**

Through above visualization figures, we have noticed several interesting phenomena.

* In, there are increasing demand for sharing bicycles. Since the population of Chicago is relative constant, we can assume that people tend to live healthier as time goes on.
* People use sharing bicycles more in summer than winter (more frequently usage and longer trip duration)
* People use sharing bicycle more frequently in weekdays than weekends, but the average trip duration is longer in weekends than weekdays.
* During rush hours (~8:00 am and ~5:00 pm), there are increasing demand for sharing bicycles during weekdays, but on weekends, there is no such clear trend.
* Most users purchased the Annual Membership. But their trip durations are relatively shorter compared with ordinary customers and dependent. Those who purchased 24-hour pass tend to have longest trip.
* Male users use sharing bicycle more often than female users, but female users have longer trips.
* Most trips are within 15 minutes.
* Most usage is in downtown Chicago. There are some stations rarely used.

**Role of Team Members:**

* Data Exploration & Cleaning: Sharadind Peddiraju
* Coding: Ajay Srivats & Utkarsh Pratap Singh
* Presentation Slides: Ajay Srivats
* Report: Sharadind Peddiraju & Utkarsh Pratap Singh