Bike Sharing Analysis

Group 15 [Int Elligence]

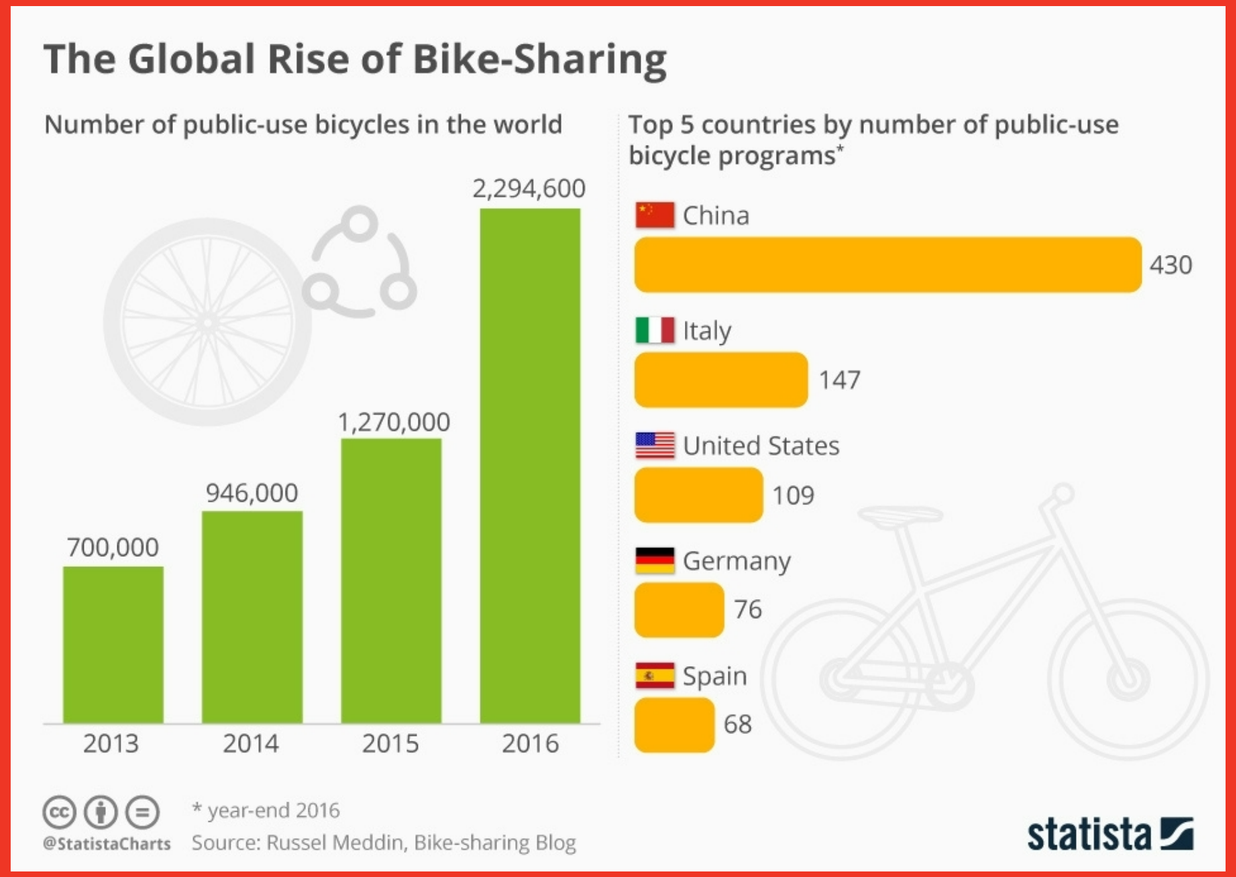
Kai-Hsiang Lin, Shubhangkar Girish Jain,

Marco Sesay, Samrat Leuva, Shivakumar Vastrad, Yanjie Chen,

Bradley R. Mascarenhas

Introduction

SHARED MOBILTY



- Exponential increase in its adoption
- Bike sharing has increased during the 2014-2017 period.
- Sustainable mode of transportation

Data Cleaning

DATA CLEANING

```
[4]: trip_id                  0
     usertype                 0
     gender              836434
     starttime                0
     stoptime                 0
     tripduration             0
     from_station_id          0
     from_station_name        0
     latitude_start           0
     longitude_start          0
     dpcapacity_start         0
     to_station_id            0
     to_station_name          0
     latitude_end             0
     longitude_end            0
     dpcapacity_end           0
     temperature              0
     windchill                0
     dewpoint                 0
     humidity                 0
     pressure                 0
     visibility               0
     wind_speed               0
     precipitation\t          0
     rain                     0
     conditions               0
     year                     0
     dtype: int64
```

- Extract data: 2017
- Gender: customers (non-subscribers)
- 75% males & 25% females
- Assign missing values with valid based on the percentages
- A good reflection of the recorded values.

DATA CLEANING

- Identify pattern in the variable 'starttime'
- Extract Month, Date, Hours, Minutes and Seconds
- Create variable Week Number
- Convert all created variable into integer for further analyses

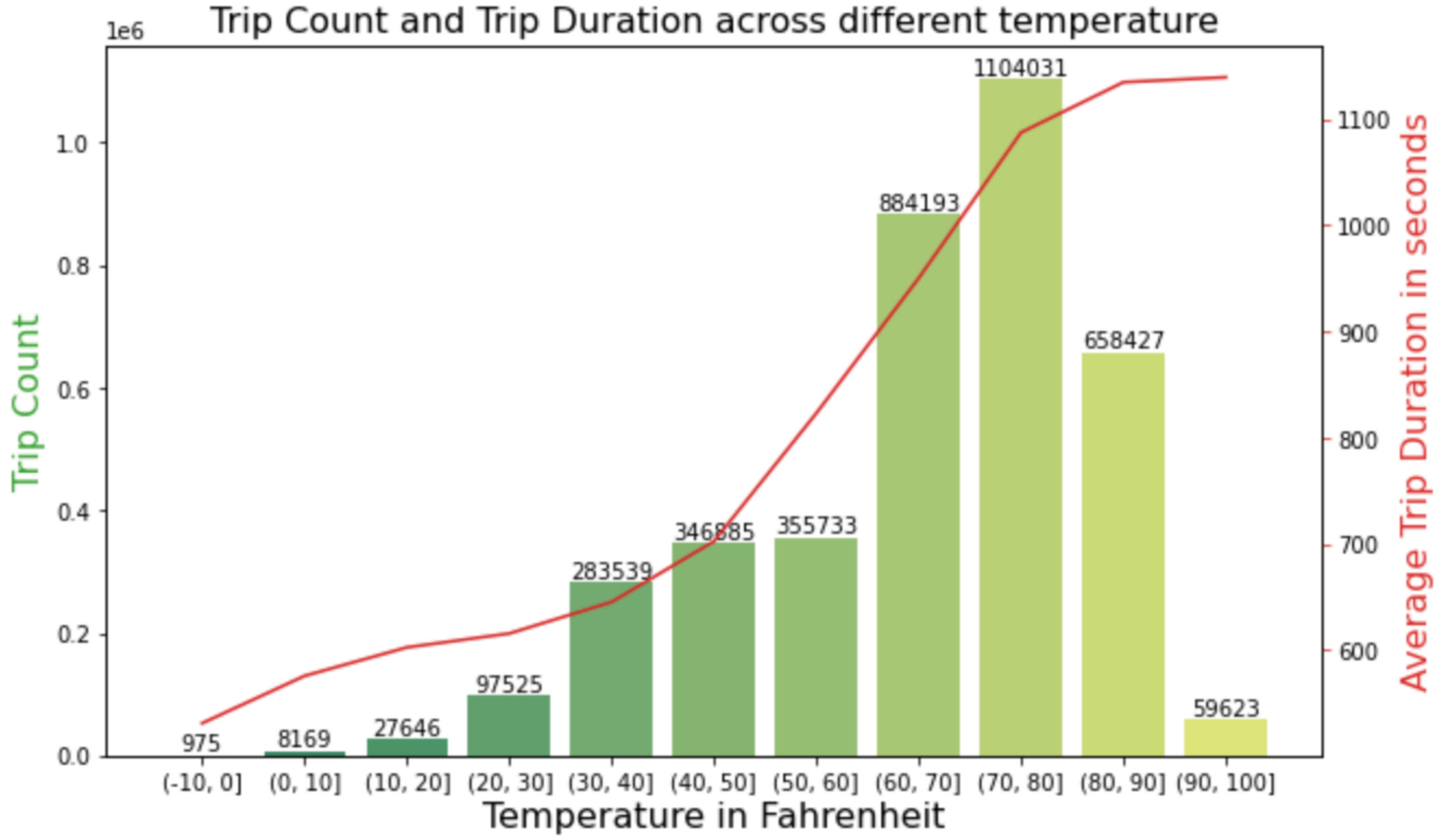
DATA CLEANING

- Identify minimum and maximum temperature: create bins accordingly
- Create variable to store weekday or weekend values
- Create an array of dates that represent the national holidays in 2017

Data Analysis

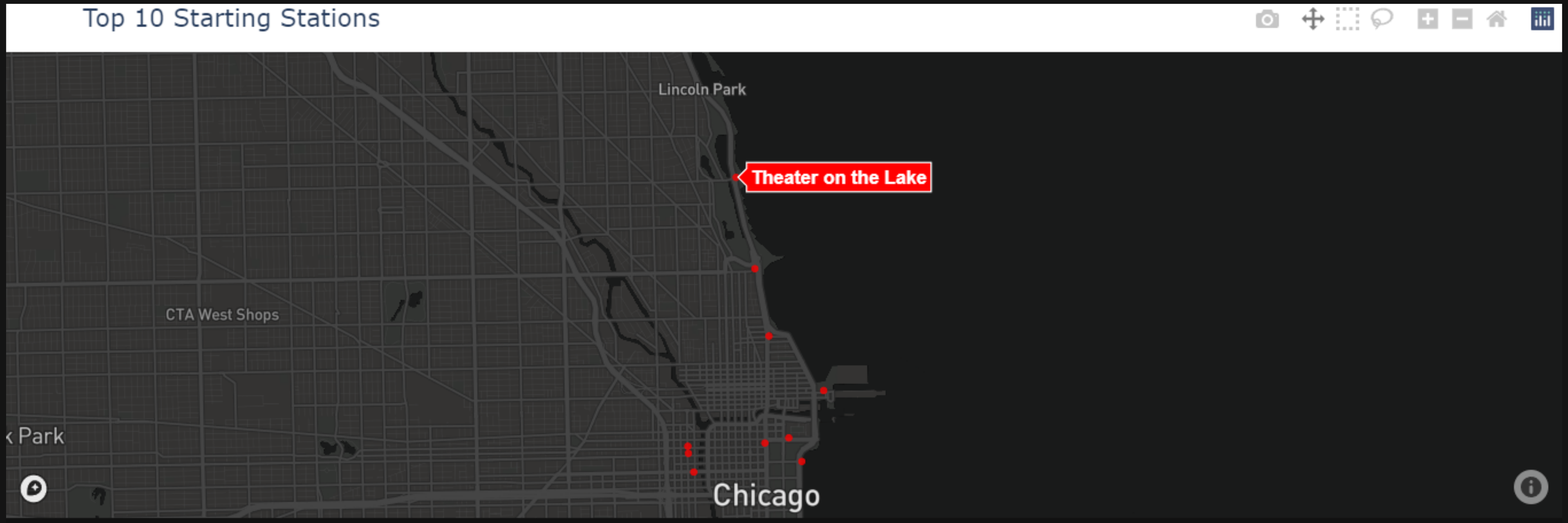
How does temperature influence the number of trips and average trip duration?

ANALYSIS ONE


Trip Count and Trip Duration across different temperature
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Trip Count
1100
1000
900
800
700
600
Average Trip Duration in seconds
975
8169
27646
97525
283539
346885
355733
884193
1104031
658427
59623
(-10, 0]
(0, 10]
(10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
Temperature in Fahrenheit

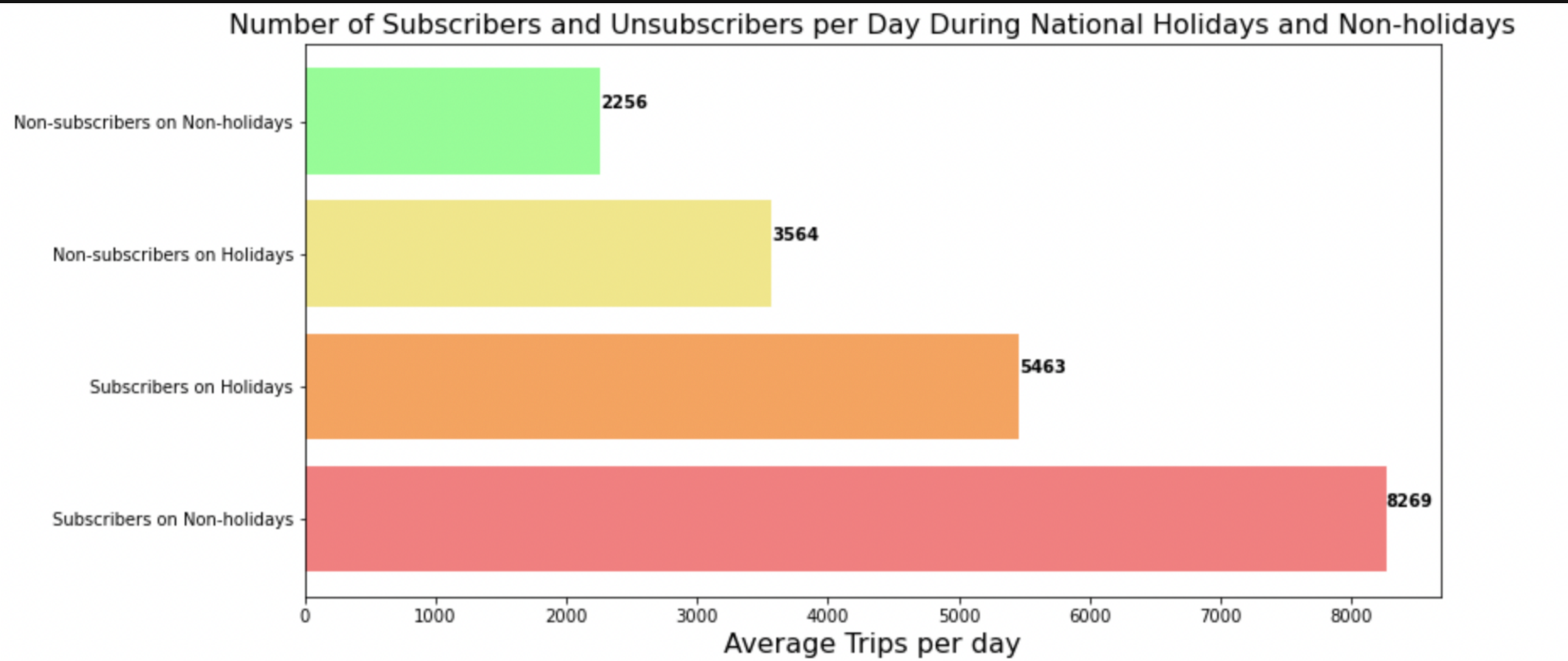
Identify the top 10 most popular start stations.

ANALYSIS TWO


Top 10 Starting Stations
Lincoln Park
Theater on the Lake
CTA West Shops
Park
Chicago

Analyze the riding behaviors of subscribed and unsubscribed users during national holidays and non holidays

ANALYSIS THREE


Number of Subscribers and Unsubscribers per Day During National Holidays and Non-holidays
Non-subscribers on Non-holidays
2256
Non-subscribers on Holidays
3564
Subscribers on Holidays
5463
Subscribers on Non-holidays
8269
0
1000
2000
3000
4000
5000
6000
7000
8000
Average Trips per day

CONCLUSION

1. The ideal temperature to ride is around 60-90 F and higher the temperature, longer the duration per trip.

2. Most of the busiest stations were on the coastline indicating great business opportunities there.

3. Rides taken by users on holidays and non-holidays was disparate.

4. On weekends, the rides seem to be normally distributed around 12:00 but on weekdays, it was dispersed.