DBA3702 Group Project

Bike Sharing Services in San Francisco

SEC[A2]: Bing Hongjian, Bryon Liew, Gavin Lim, Zhou Kai Jing

NUS National University of Singapore | NUS BUSINESS SCHOOL

1

What problem do we want to tackle?

An introduction to bike-sharing services

NUS National University of Singapore | NUS BUSINESS SCHOOL

Bike-sharing services

Bike-sharing services

A service where bicycles are made available for shared use to individuals on a short-term basis

Business model

Docking Station: Bicycles must be returned at designated docking stations

Geo-fencing: Enables bike hires to be ended only within a virtual fence

Free-floated: Allows bike drop-off at any location within a city's boundaries

The Problem: Bike-sharing in Singapore

Dominated by dockless bikes

Local bike sharing business began in early 2017 with Obike.
The wave of bike-sharing services continued with Ofo and Mobike who were established players in the bike-sharing market with a huge presence in China.

The Problem in Singapore

The uneven spread of shared bicycles across different the stations in Singapore
Surplus and shortage of bicycles in certain stations

Our Aim: Creating an application that aids in problem solving and analysis

2

What are some existing market practices?

Technology adopted and bike management

QR Code System

What are the functions of these QR codes?

To unlock the bicycles, QR codes must be scanned on the company's app.
To lock the bicycle, it must be returned to a stipulated parking location.
To end your trip, a QR code at the parking location must be scanned.

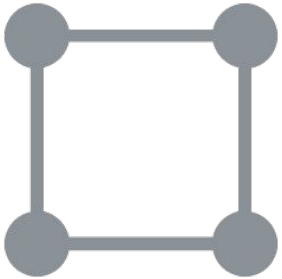
Geo-Fencing System

Only allows bikers to park and lock their bikes when they are within the virtual perimeter that is assigned by the biking company.

Station organization & Bike management

Station zones

Stations are classified into 3 groups: Small, Medium and Large zones
Each zone accommodates 5, 8 and 10 bicycles respectively
Large stations are located in areas with high traffic: MRT & Bus stations



Problem: Surplus and shortage of bikes

Some Small and Medium stations having high rental rates,
Some Large stations having low rental rates,
Tackling the problem of surplus and shortages at the various bicycle stations.

3

How are you going to tackle the problem?

What data to use and where to get them?

What data to use?

- Initially...
  - Local bicycle sharing companies such as OFO, Mobike and SGBike
- Limitations
  - Availability of dataset
  - Planned exit of companies from SG

What data to use?

- Final dataset for this project
  - Overseas bicycle sharing data from San Francisco Bay Area
- Source of Data
  - https://www.kaggle.com/benhamner/sf-bay-area-bike-share

Structure of Dataset

- Components of dataset
  - Station, Status, Weather and Trip data
- Data Cleaning
  - Aggregation of hourly docks and bicycles available
  - Determination of Usage Rate
  - Cleaning of NA values for variables used

Initial Data Analysis (ANOVA)

- Variables used: Mean Temperature, Mean Humidity, Cloud Cover
- Correlation Matrix ⟶ No multicollinearity

	mean_temperature_f	mean_humidity	cloud_cover
mean_temperature_f	1.0000000	-0.2254128	-0.1827273
mean_humidity	-0.2254128	1.0000000	0.5756338
cloud_cover	-0.1827273	0.5756338	1.0000000

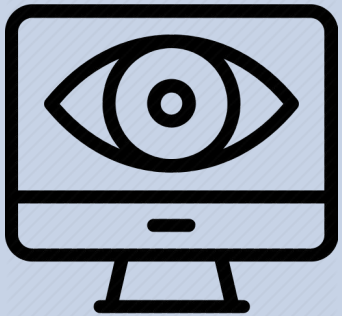
Initial Data Analysis (ANOVA)

- ANOVA Analysis Results

```
                    Df Sum Sq Mean Sq F value   Pr(>F)
mean_temperature_f    1     63   63.39 1872.88  < 2e-16 ***
mean_humidity         1      1    1.25   36.94 1.22e-09 ***
cloud_cover           1      8    7.91  233.70  < 2e-16 ***
Residuals       2945130  99687    0.03
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

→ All 3 variables are good predictors of Usage Rate

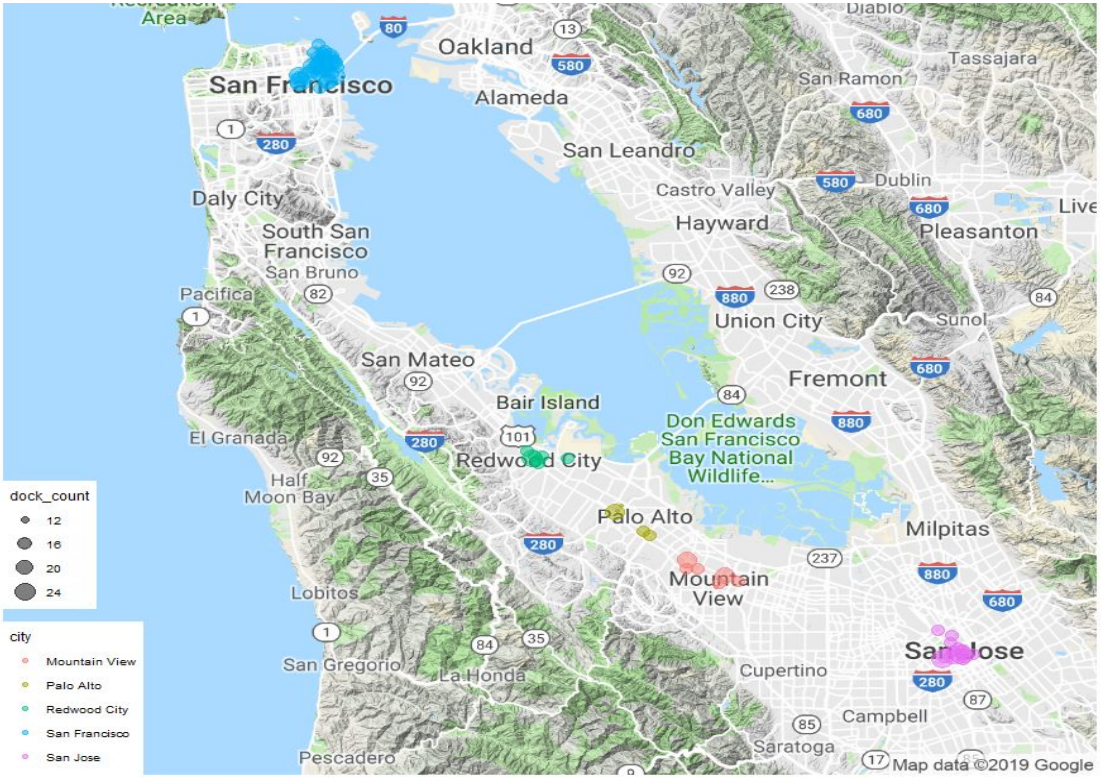
4



Geospatial Visualisation

What can we visualize from the data?

Overview of bike stations at the Bay Area



city

- Mountain View
- Palo Alto
- Redwood City
- San Francisco
- San Jose

Overview of stations and docks at the Bay Area

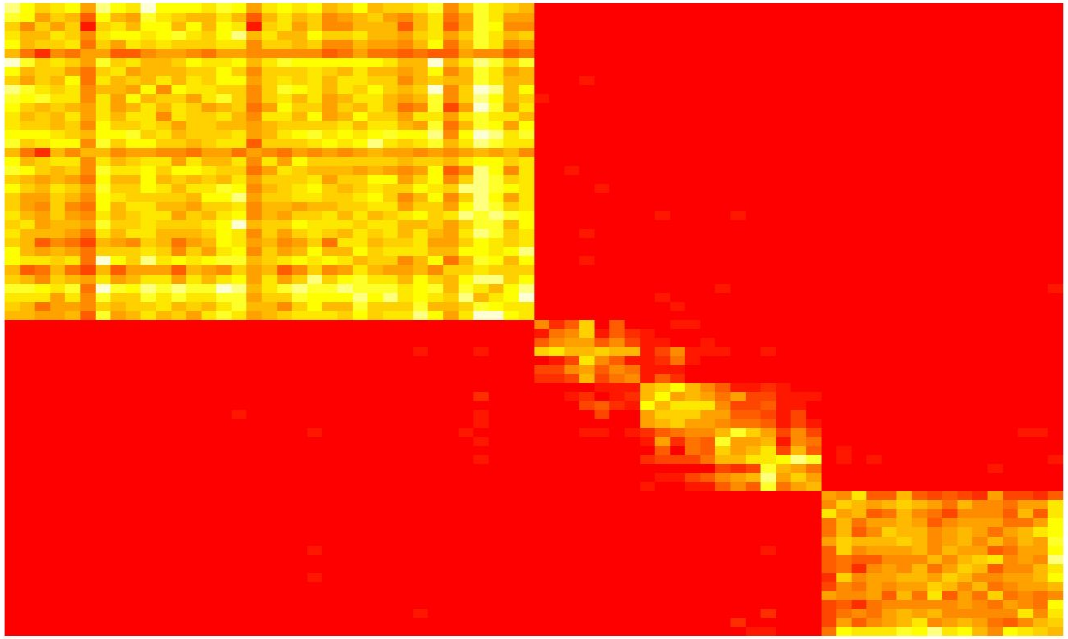
Travel Pattern

	Embarcadero at Sansome	Embarcadero at Vallejo	Broadway St at Battery St	G A C A
Embarcadero at Sansome	2084	625	318	
Embarcadero at Vallejo	1178	617	59	
Broadway St at Battery St	302	49	237	

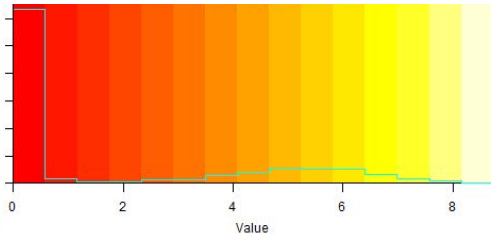
-sort the stations based on their latitude so that the stations are more close to each other in geography

-conduct network analysis to view the trip numbers(connection) between each 2 stations

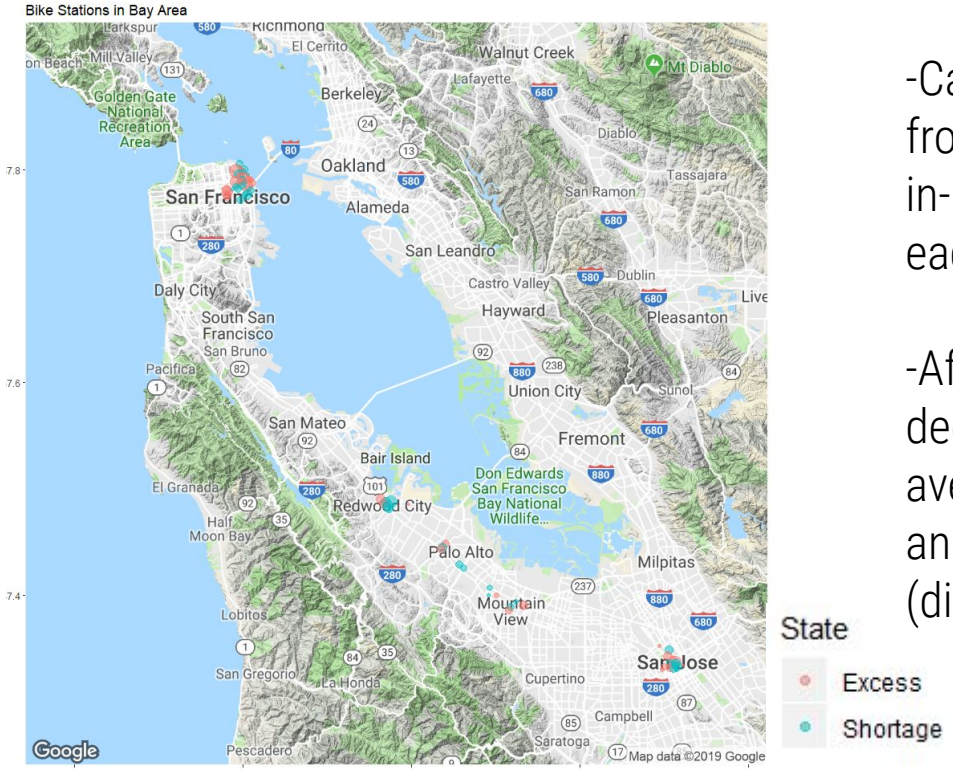
Travel Pattern



- Trips tend to be intra-city instead of inter-city



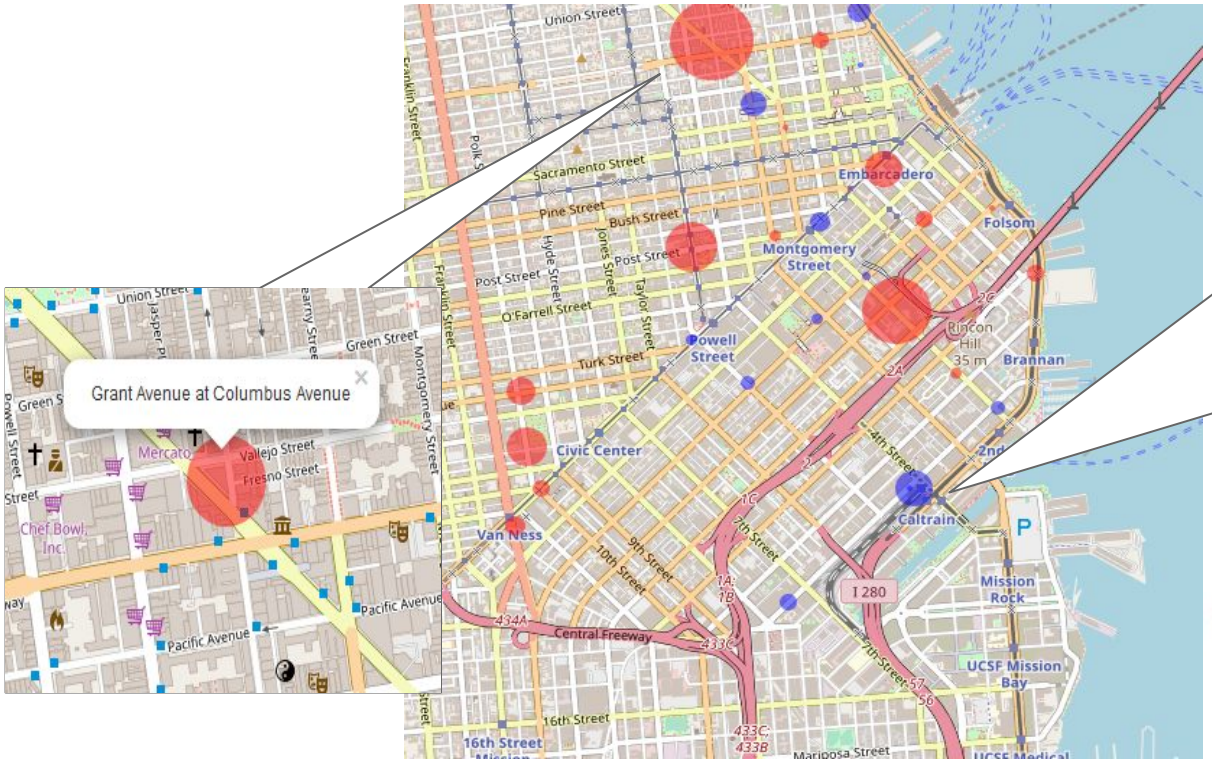
Shortage and Excess at each station



-Calculate trips count starting and ending from each station, to calculate the in-degree and out-degree centrality of each node (station) in this network graph.

-After that, the difference of out degree in degree is calculated. By dividing the average of the two degree, we could get an index the severity of bike excess (diff<0) or bike shortage (diff>0)

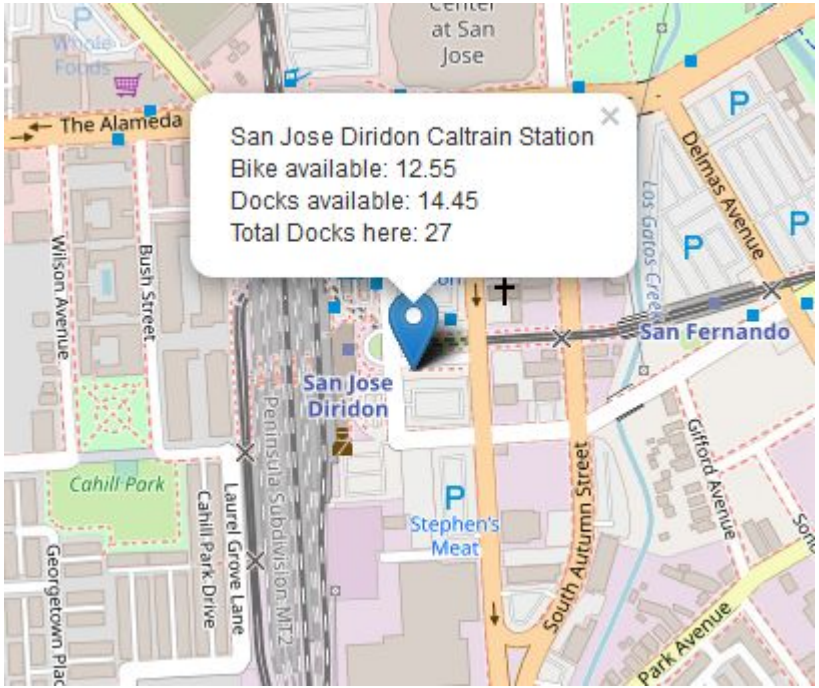
Shortage and Excess at each station



State

- Excess
- Shortage

Finding bikes at a station



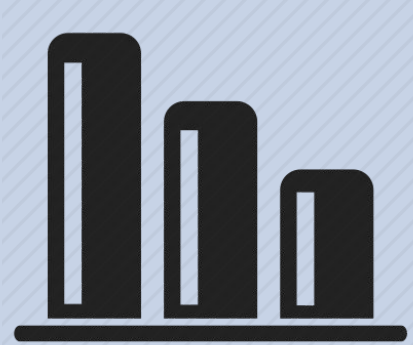
DATE: 2014-05-01

TIME: 12PM

CITY: San Jose

STATION: San Jose Diridon Caltrain Station

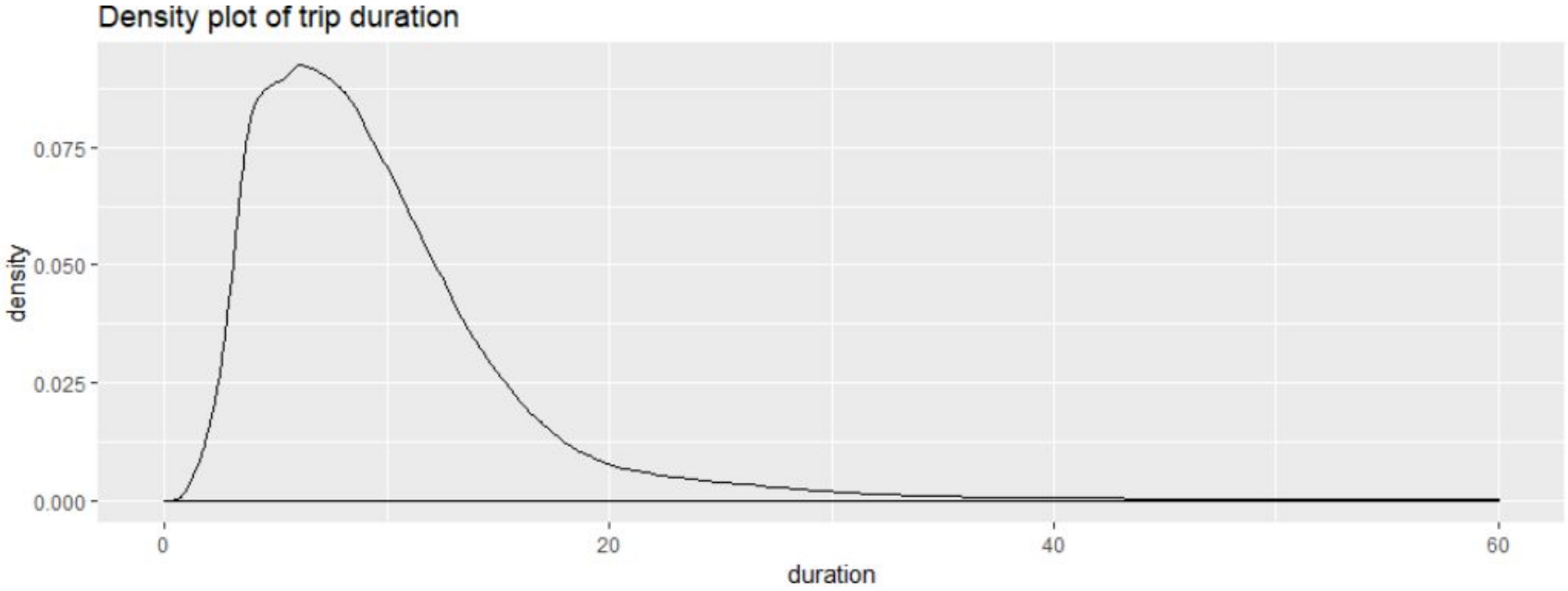
5



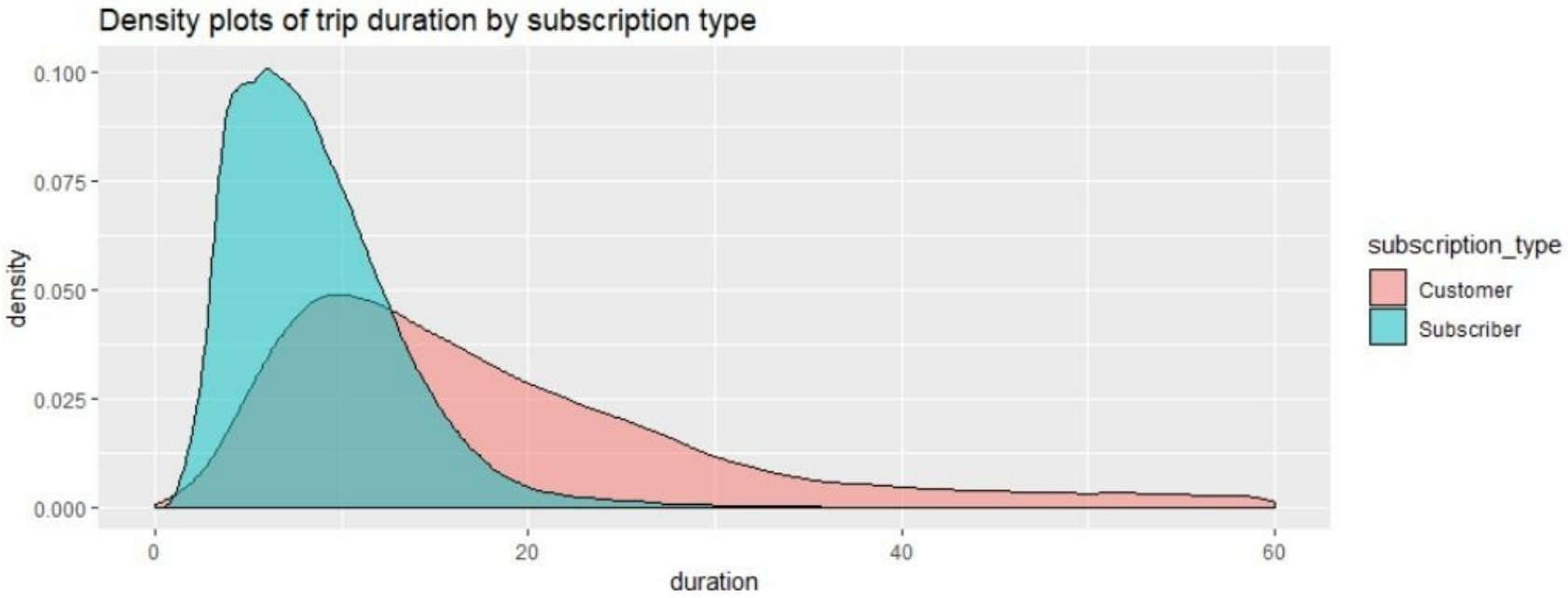
Graphical Visualisation

What graphs can be plotted?

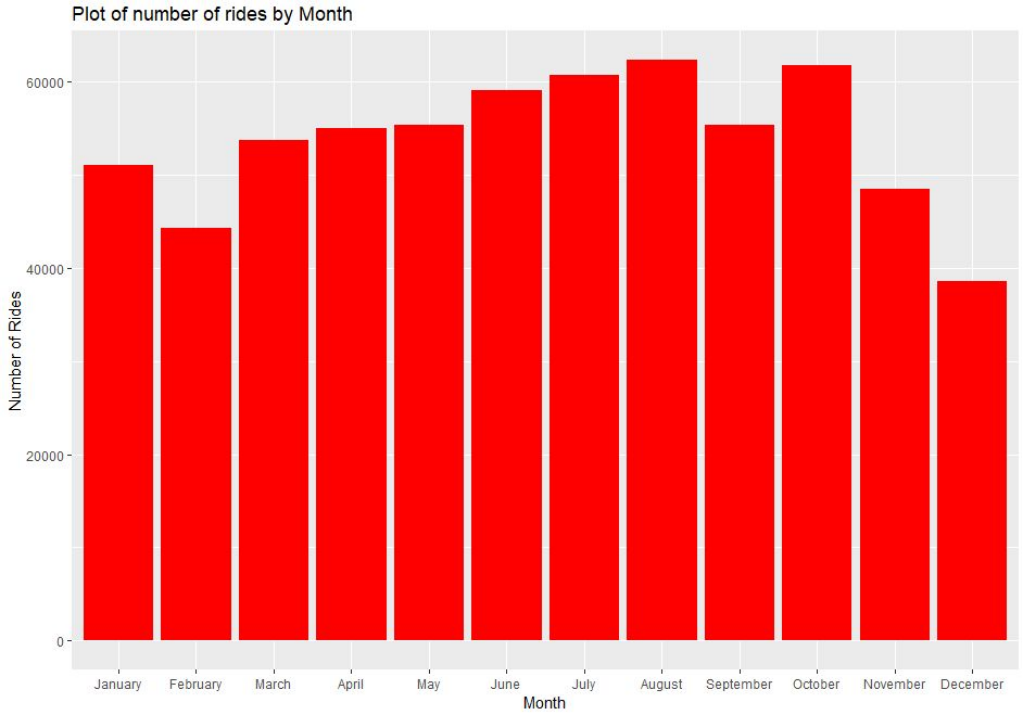
Density plot of trip duration



Density plot of trip duration by subscription type



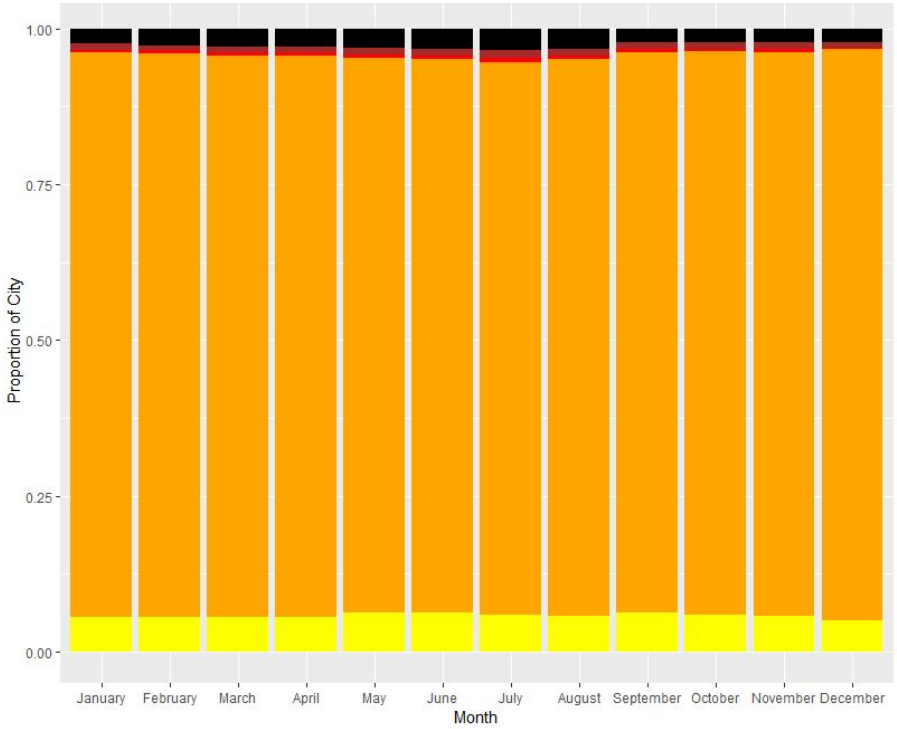
Distribution of trips by Month



-Demand for bikes will generally increase from March to October, except a fall in September

-November, December, January and February are less peak

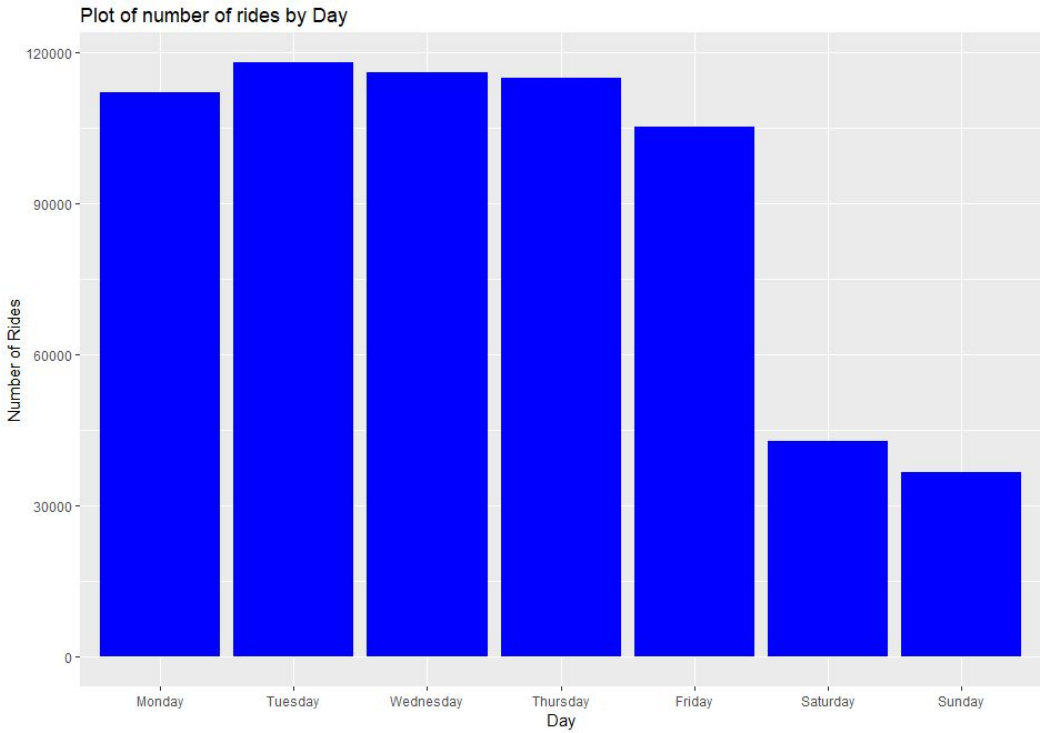
Proportion of trip numbers by City and Month



-Percentage of bikes from each city is relatively constant

-Although the number of trips changes, the proportion of bikers are roughly the same

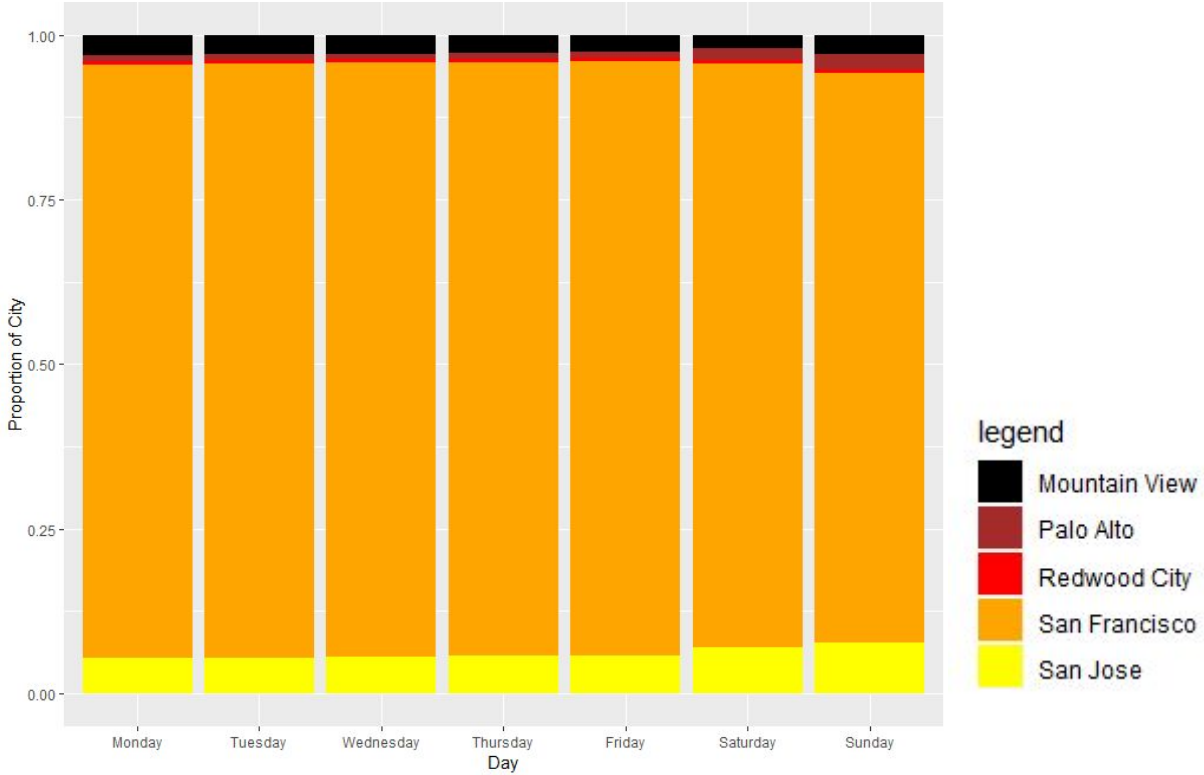
Distribution of trips by Day



-Overall, there are less rides on weekends (Saturday & Sunday)

-This could indicate that a majority of bike-share users use these bicycles to commute to work or school on weekdays

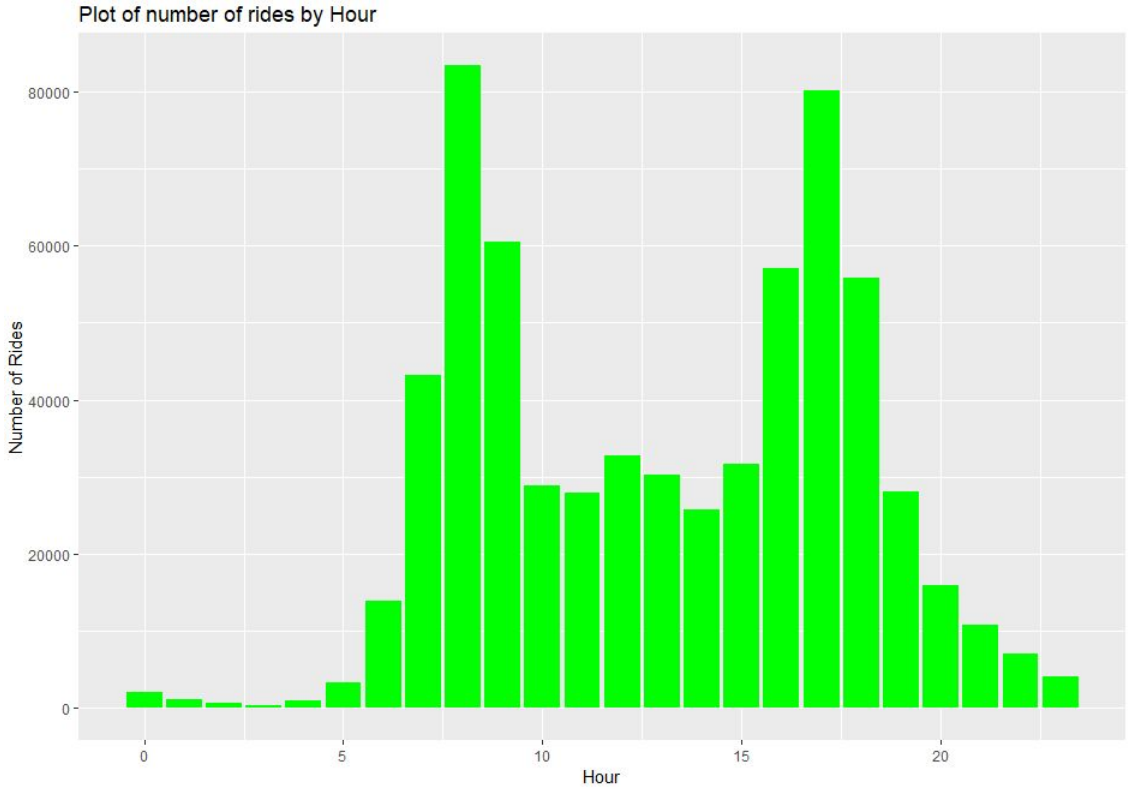
Proportion of trip numbers by City and Day



-San Francisco sees a decrease in bike trips on weekends

-Palo Alto and San Jose experience higher proportion bike rides on weekends

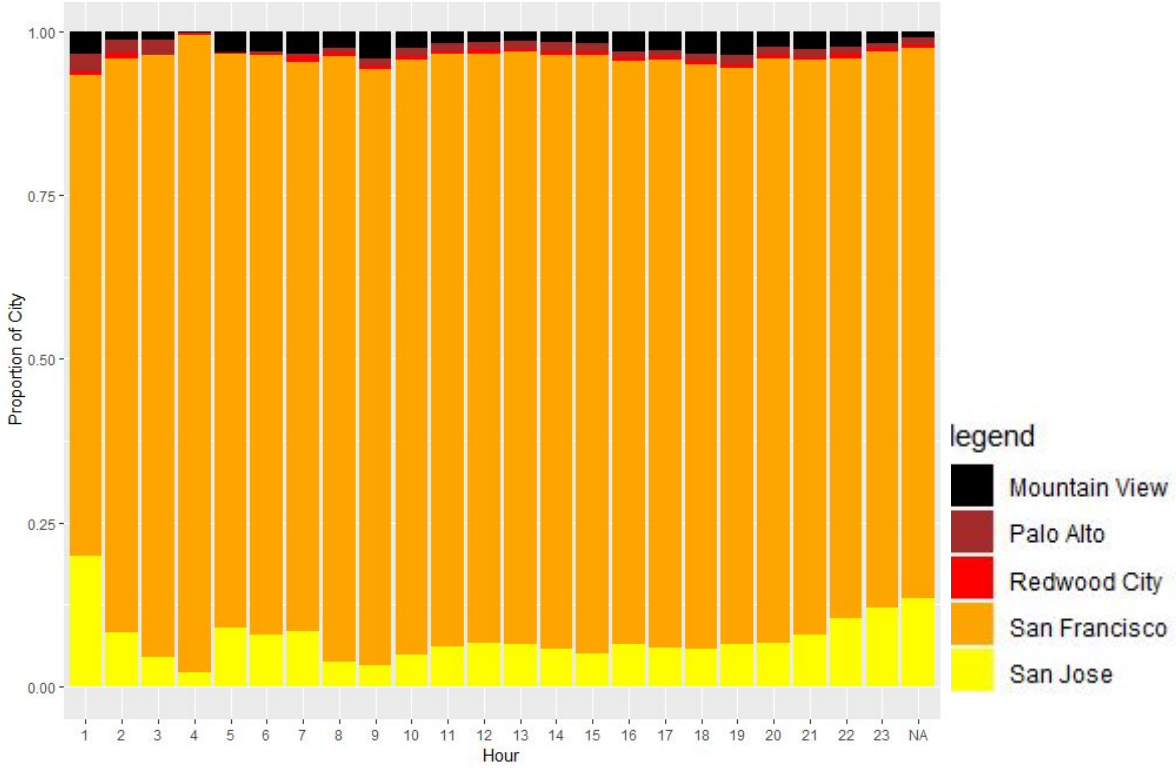
Distribution of trips by Hour



-Largest concentration of rides occur in the morning from 7-9am and in the evening from 4-7pm

-Reinforces our team's observation that commuters use these bikes to get to-and-from work

Proportion of Trips by City and Hour



-Proportion of number of trips by city remains even

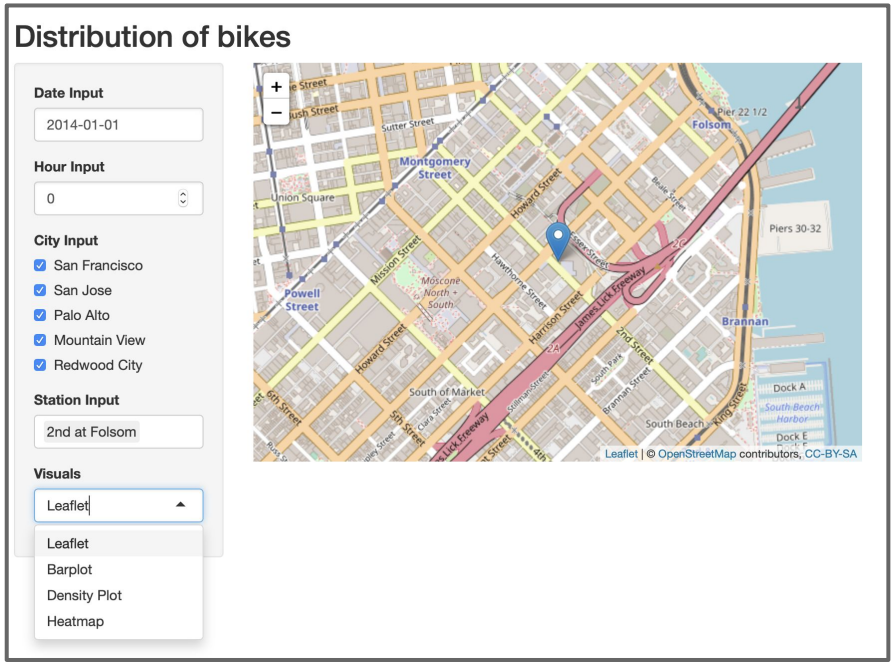
-Some variations in trips from 2200hrs to 0500hrs, but the number of total trips at these timings are much lower thus there will be larger variations in proportion

6

Shiny App

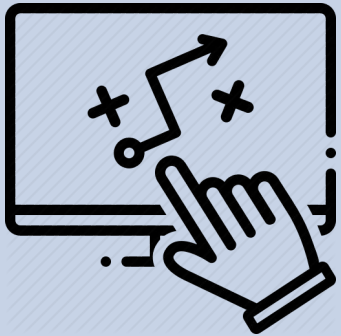
How does Shiny App help us in better understanding our data?

Purpose of Shiny App



To provide an Interactive Interface for our users to view the different data visualizations available and also how the different parameters (Date, Time, City, Station) affect the visualization

7



Strategies Employed

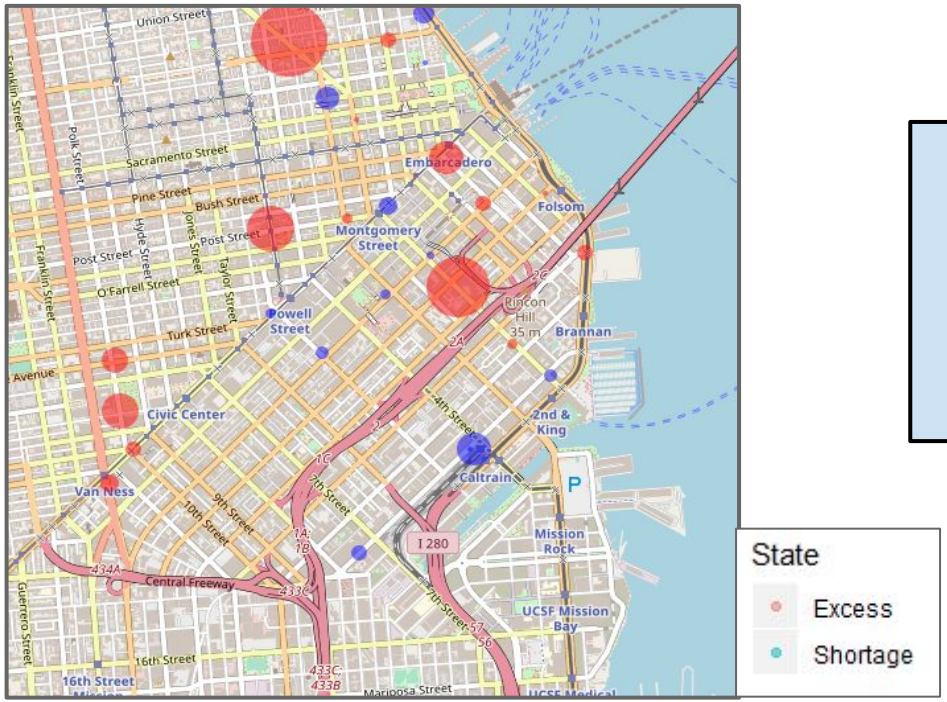
What strategies can we adopt to improve bike-sharing services?

Dynamic Pricing

Example from Uber

- Introduce Dynamic Pricing
- Pricing to depend on demand and availability of bikes
- Helps to improve shortage issues

Re-distribute supply of bikes



- Re-distributing supply of bikes
- Areas with considerable excess in bikes can be redistributed to areas with shortage in bikes

THANK YOU!