Proposal for
A Demand Prediction Model and
A Dynamic Pricing Model
for Share Bike Business
Utilizing Machine Learning

weiwei1988

1. **Basic Understandings for the Share Bike Business**

2. **Analysis on the Influence Factors on demands**

3. **The Demand Prediction Model Utilizing Machine Learning**

4. **The Dynamic Pricing Model Based on the Demand Prediction Model**

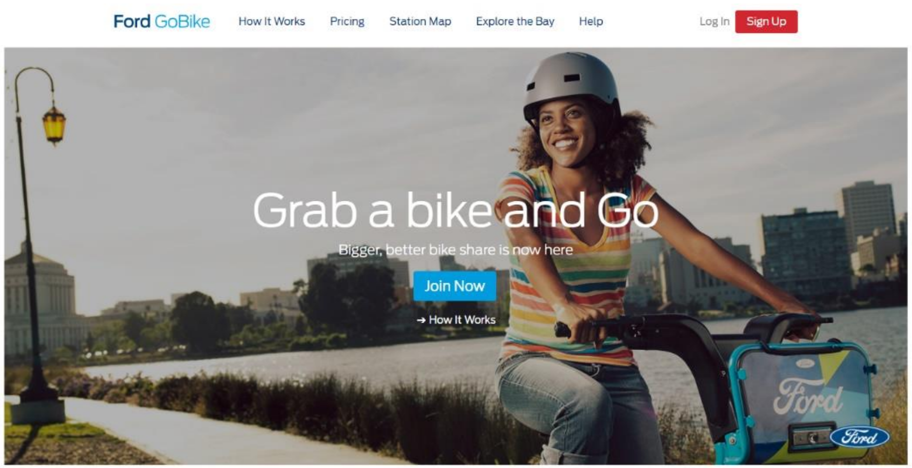
1. Basic Understandings for the Share Bike Business

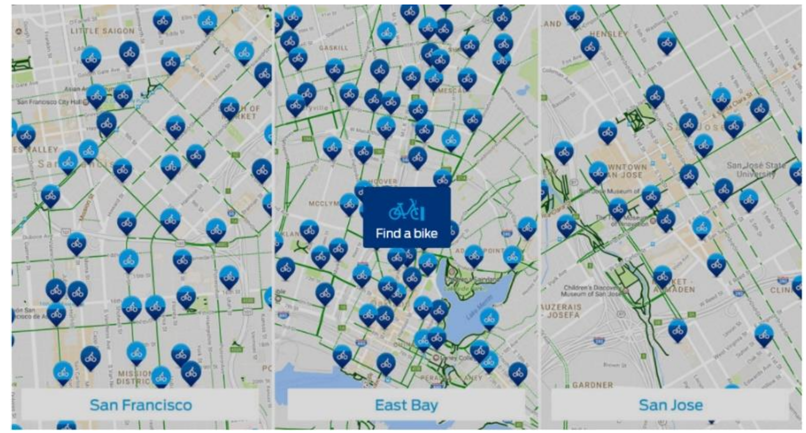
2. Analysis on the Influence Factors on demands

3. The Demand Prediction Model Utilizing Machine Learning

4. The Dynamic Pricing Model Based on the Demand Prediction Model

The "Gobike" is a share bike business provided by Ford Motor in San Francisco bay area since 2013.





No. of Stations before Aug. 2015

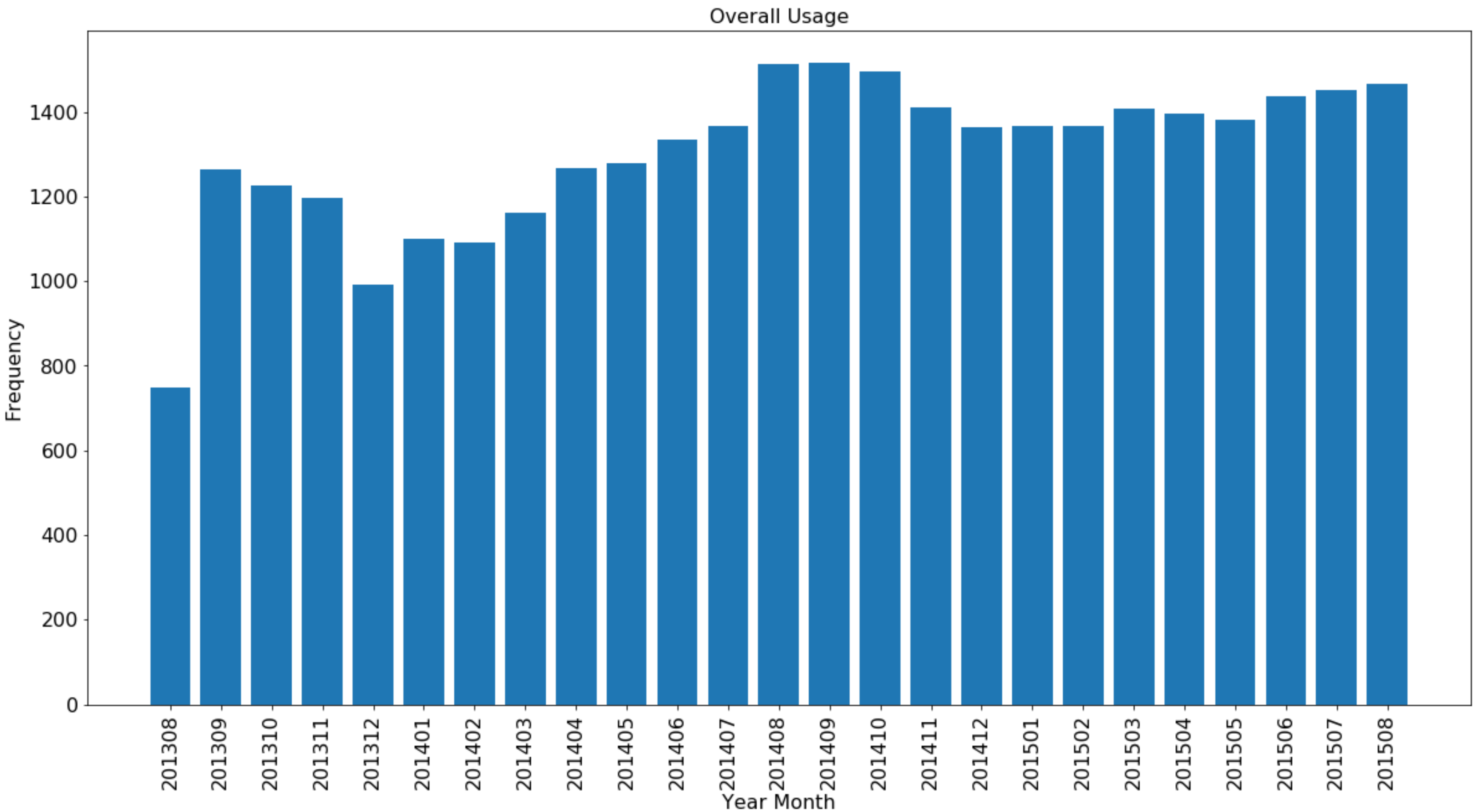
City	Stations
Mountain View	7
Palo Alto	5
Redwood City	7
San Francisco	35
San Jose	16

Source: https://www.fordgobike.com/

Issue for the share bike business:

The growth of demand has stagnated since Oct. 2014.

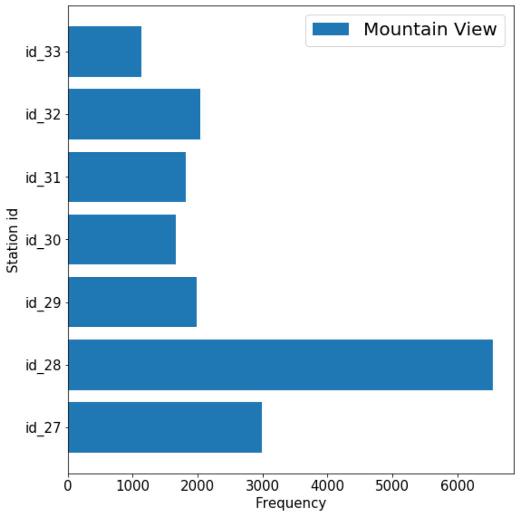
No. of Demands by Month [Trips]

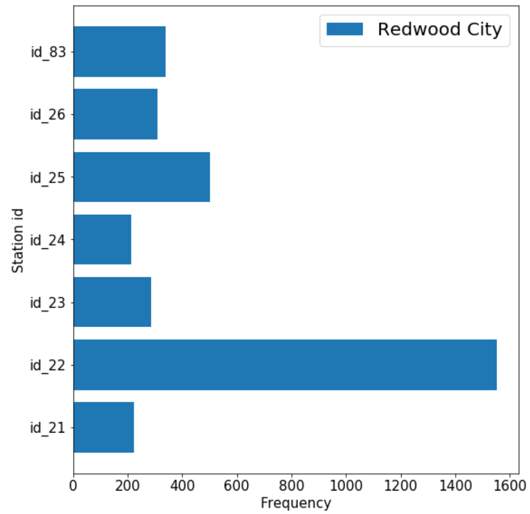


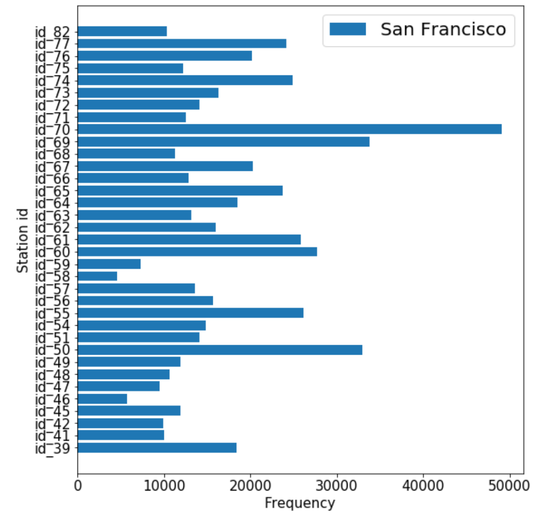
The cause of growth stagnation:

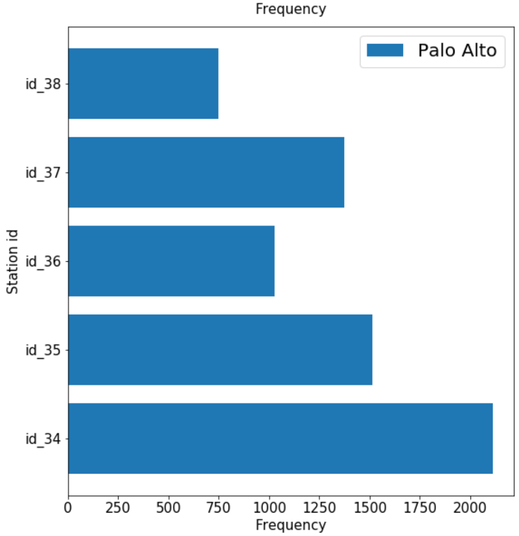
The No. of trips varies by the station in different cities.

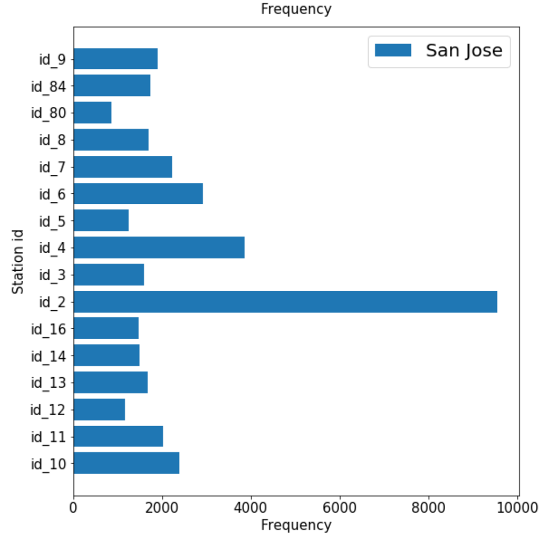
For ex: Demands in downtown area like San Francisco has maximum No. of 50 thousand trips, Palo Alto, on the other hand, had only 2 thousands trips in total.











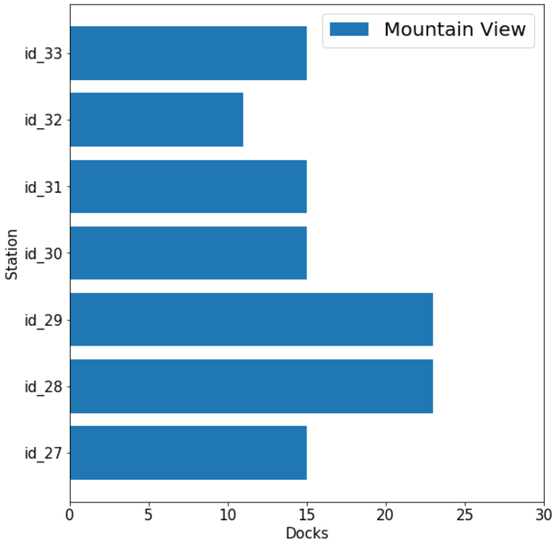
Other factors such as Month, Date, Weather also influence the No. of demands

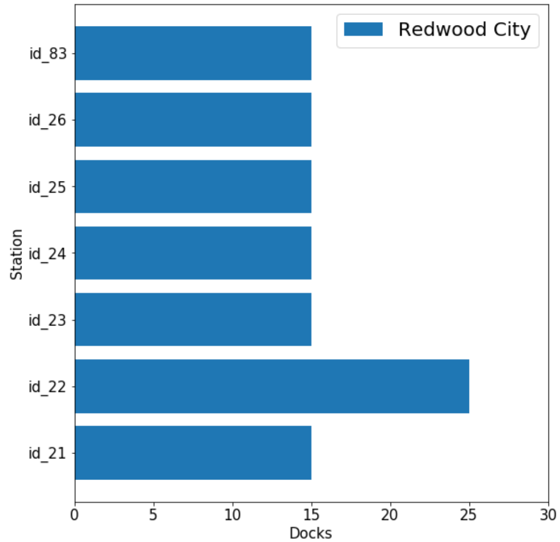
Period for analysis：

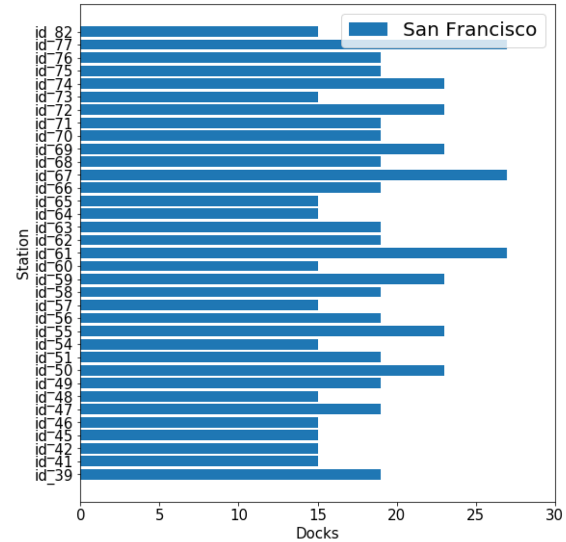
Aug. 2013/8~Aug. 2015

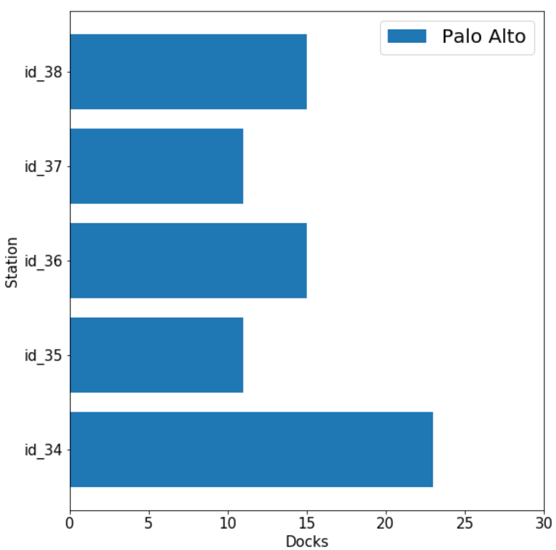
The cause of growth stagnation:

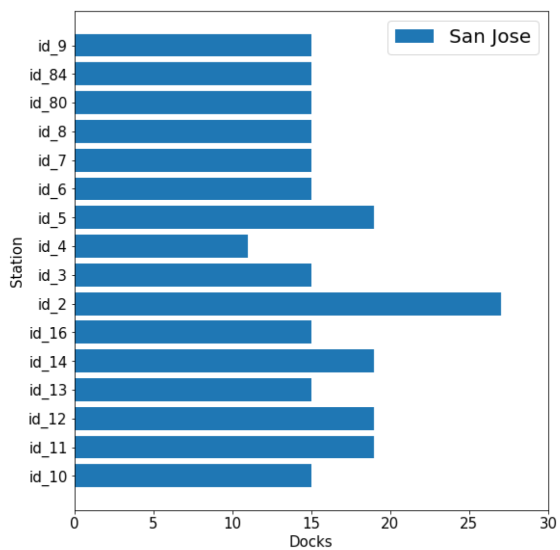
However, No. of docks in each station did not match the diversified No. of trips in different stations.









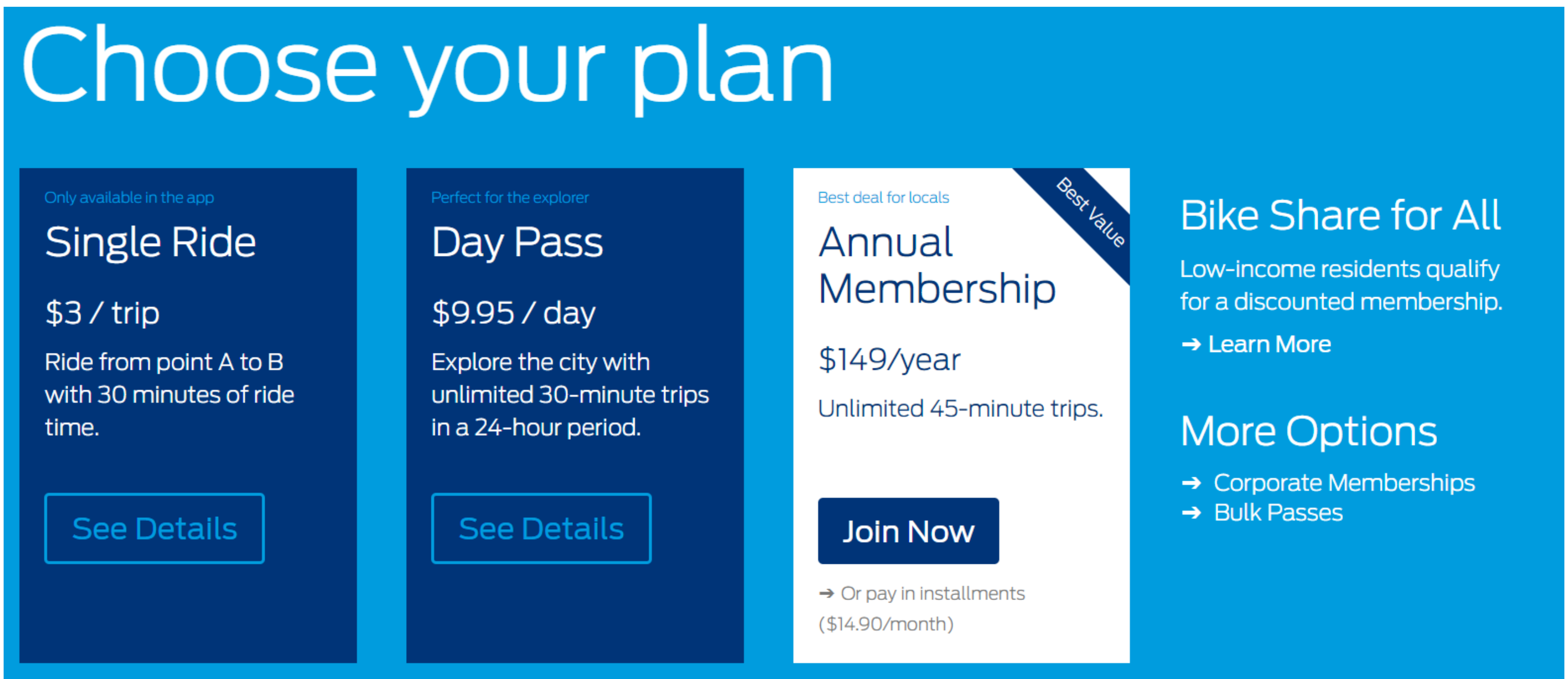


Period for analysis：
Aug. 2013/8~Aug. 2015

The cause of growth stagnation:

Inflexible price plan was another big problem.

Current Price Plan:
$3 / Trip (Up to 30min）
$9.95 / Day (Up to 30min）
$149/Year（up to 45min for subscribers）



Source: https://www.fordgobike.com/

It is necessary to understand the precise demands of trips to avoid supply and demands un-matching situation.

No. of demands varies by station, city, and some other factors.

However, No. of docks and price plan are inflexible.

Un-matching problems occurs between supply and demand, which lead to the stagnation of the business growth.

It is necessary to understand the precise demand of trips to adjust the no. of docks and price plan.

A Proposal for demand prediction utilizing the machine learning technique.

1. Basic Understandings for the Share Bike Business

2. **Analysis on the Influence Factors on demands**

3. The Demand Prediction Model Utilizing Machine Learning

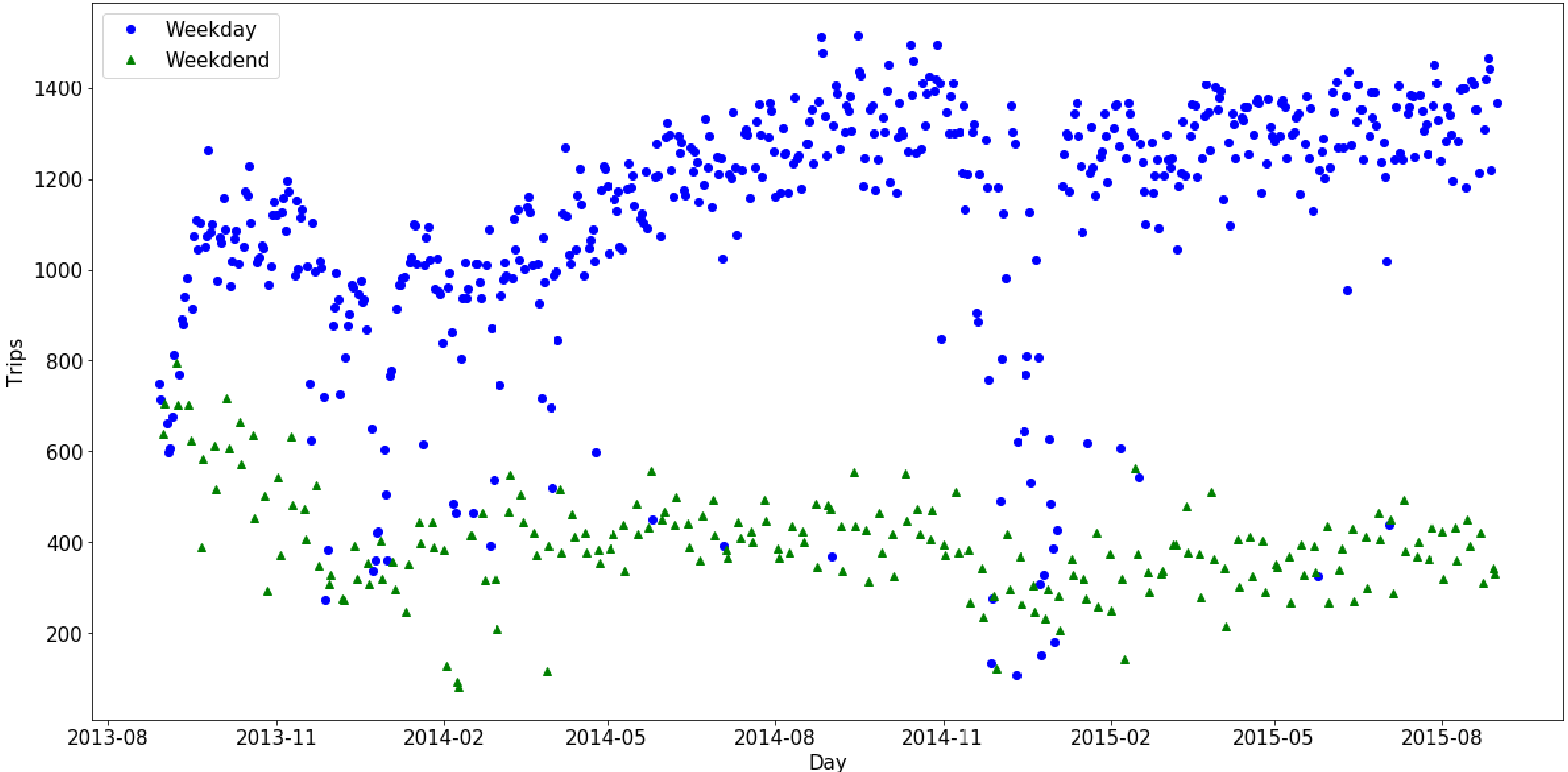
4. The Dynamic Pricing Model Based on the Demand Prediction Model

Influence Factors on demands:

The No. of trips in weekday and weekend have substantial difference. Weekdays have 2-3 times larger trips.

Weekday: Mon/Tue/Wen/Thu/Fri
Weekend: Sat/Sun

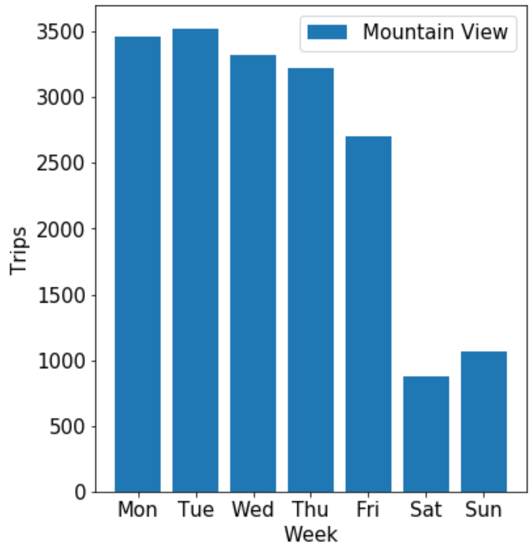
No. of trips by day

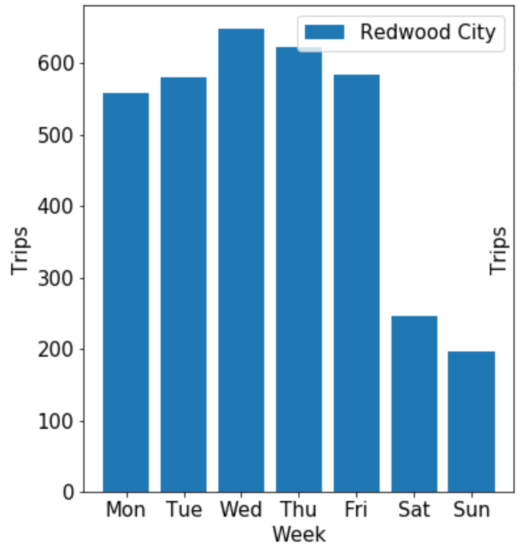


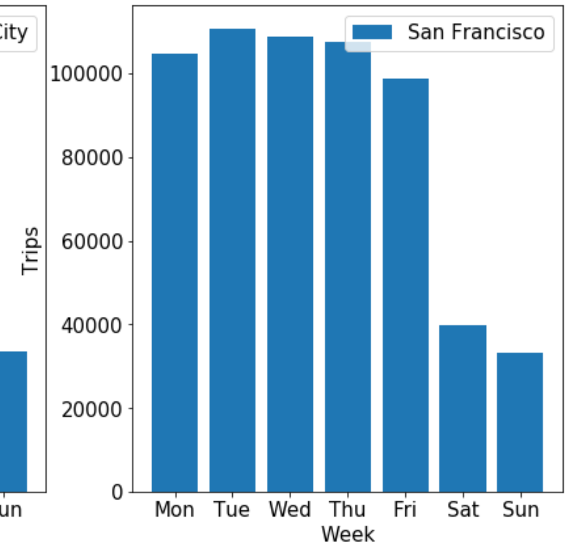
Period for analysis：
Aug. 2013/8~Aug. 2015

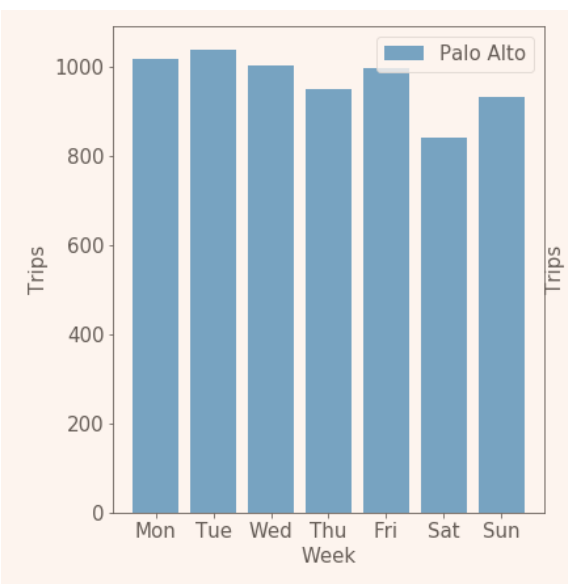
Influence Factors on demands:

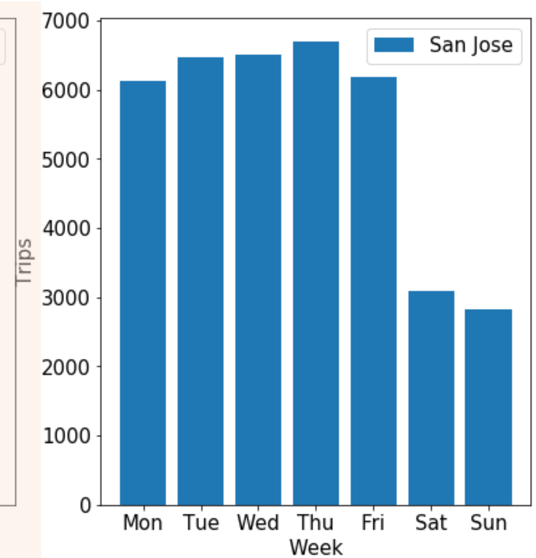
Influence of weekends also varies by cities:
Weekend trips in downtown area drops substantially.







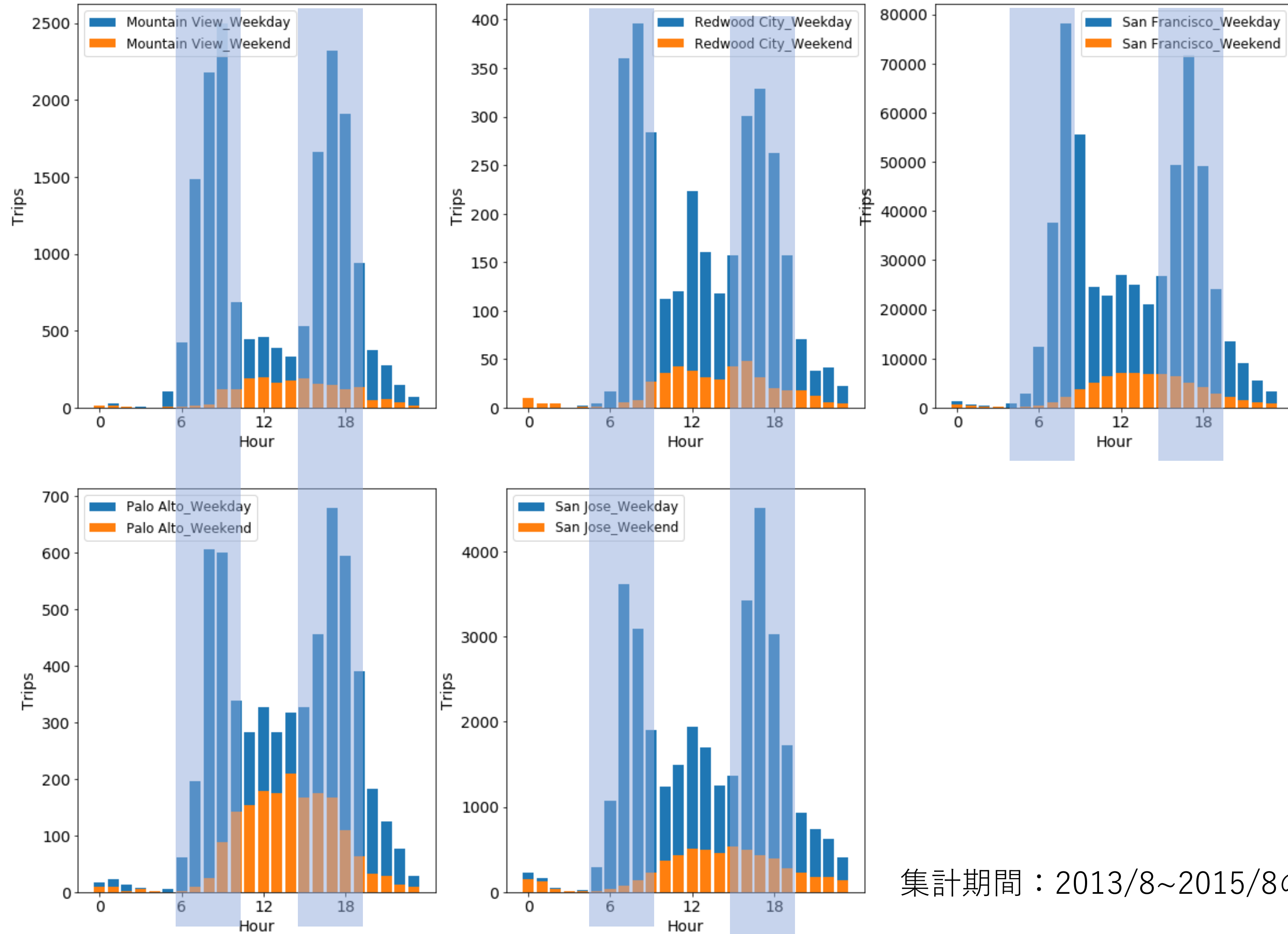




Period for analysis：
Aug. 2013/8~Aug. 2015

Influence Factors on demands:

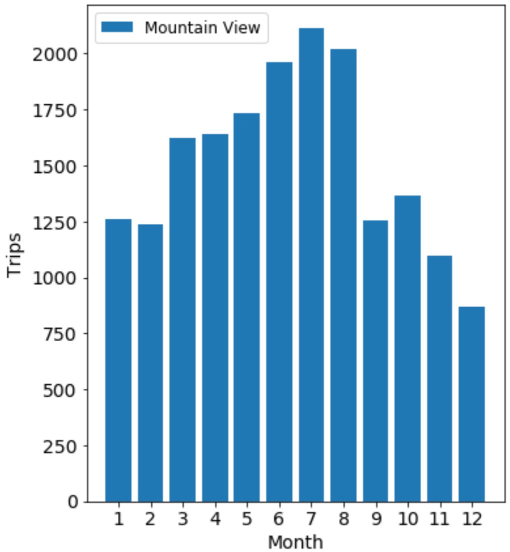
The No. of trips also depend on the time zone. Moring rush & evening rush indicate sharp increase on demand.

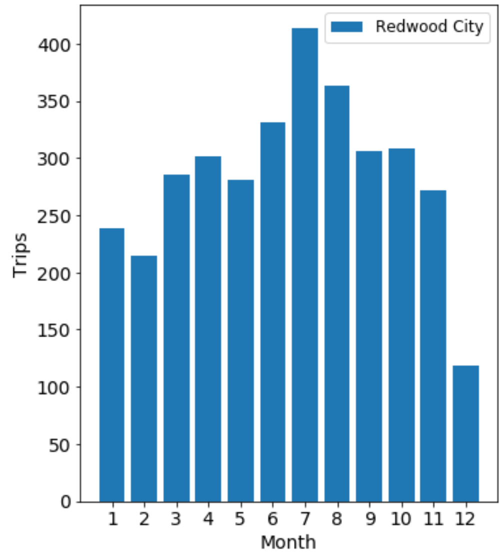


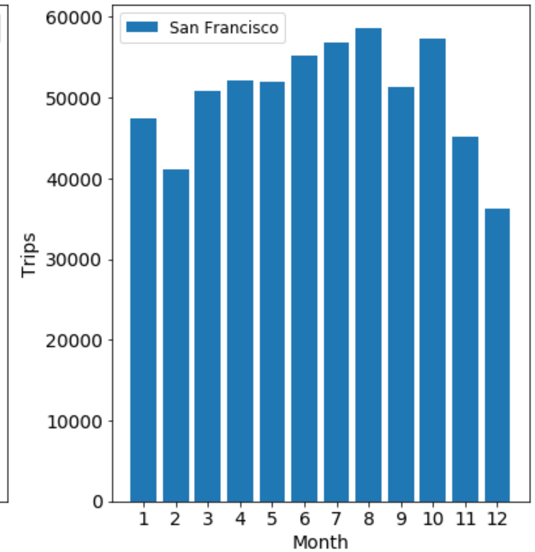
集計期間：2013/8~2015/8の３年間

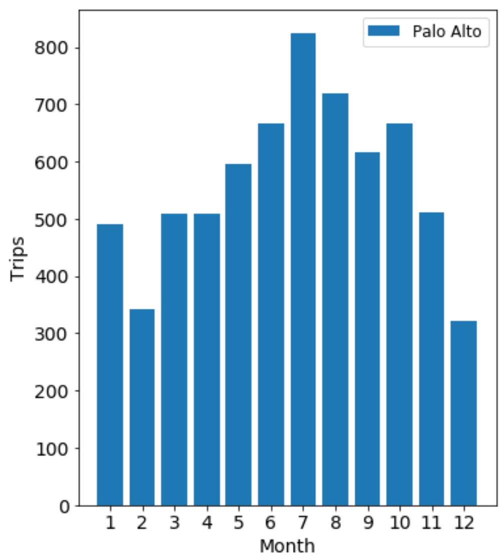
Influence factors on demands:

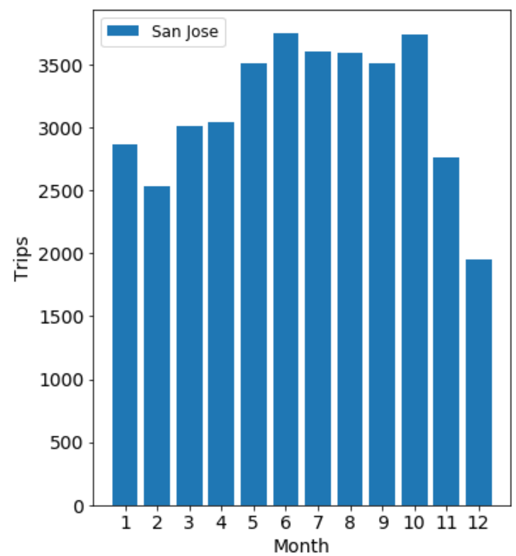
No. of trips tend to increase in summer, but tend to drop in winter.





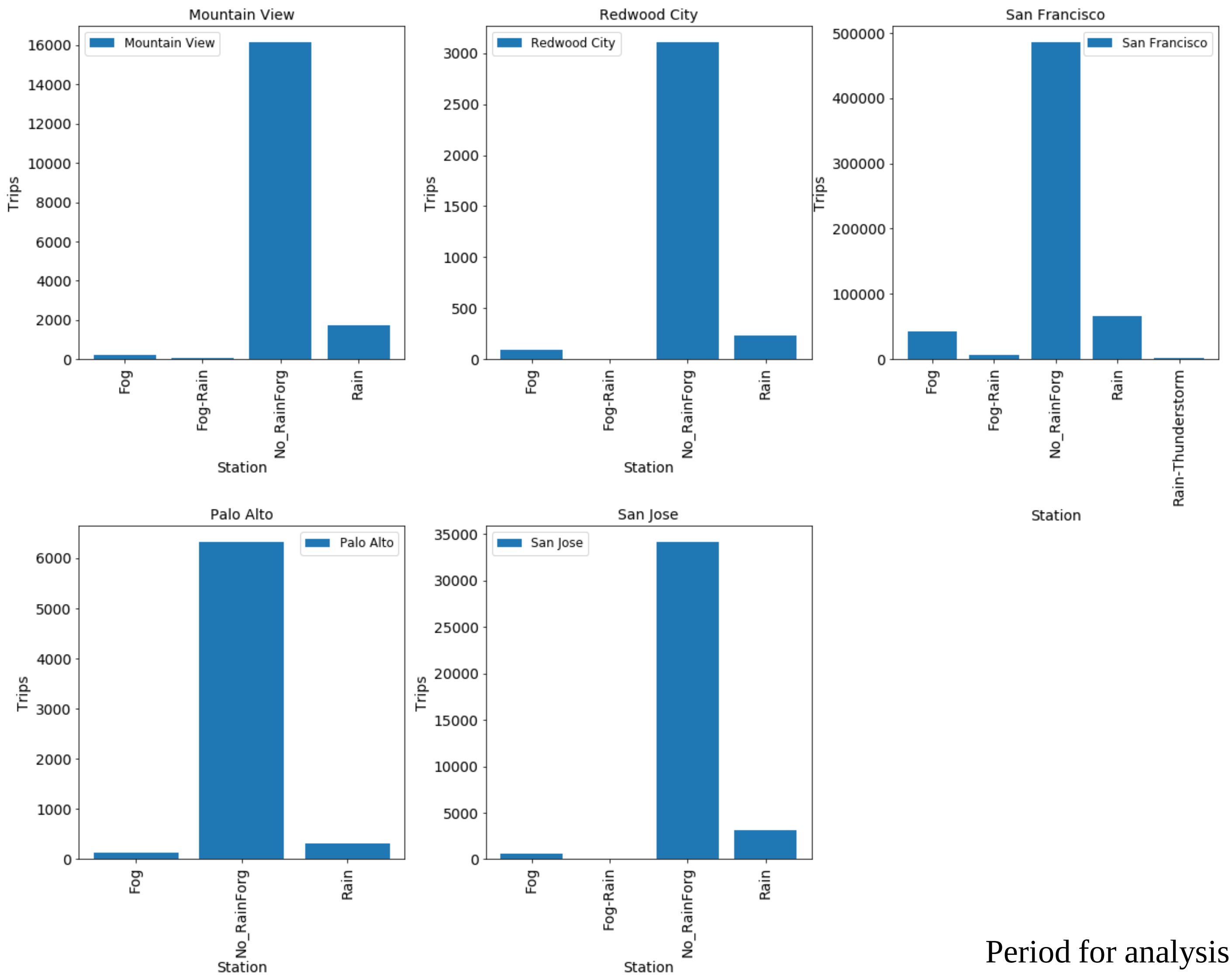






Period for analysis：
Aug. 2013/8~Aug. 2015

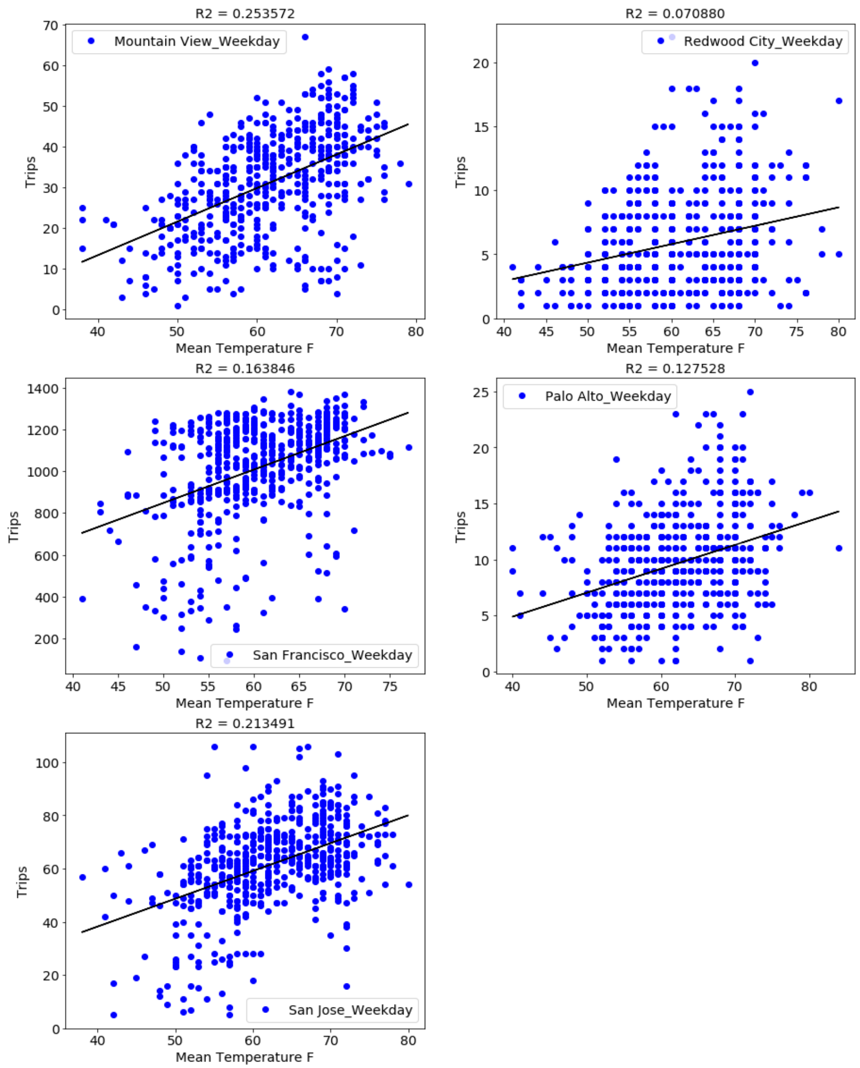
Influence factors on demands:
Most of trips concentrate on the days without weather events. (No_RainForg stands for the NaN of weather events)



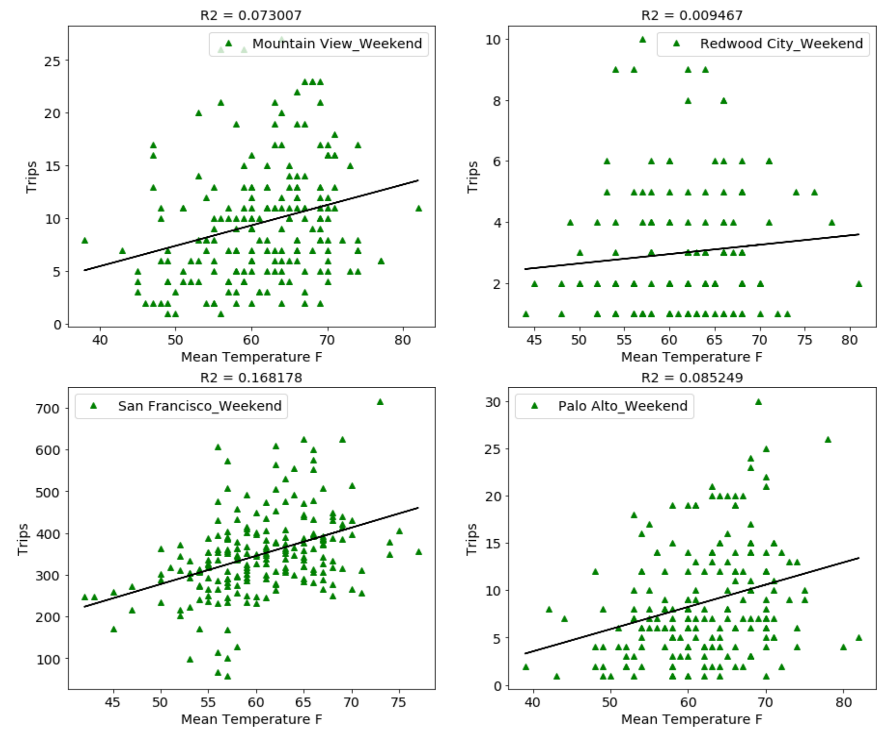
Period for analysis：
Aug. 2013/8~Aug. 2015

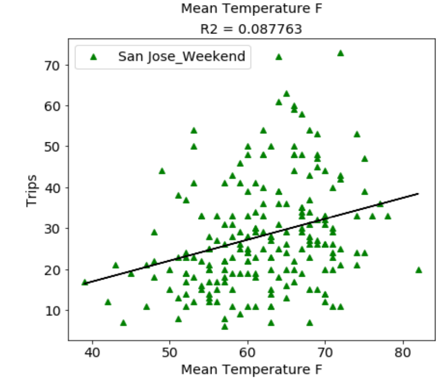
Influence factors on demands:
The mean temperature seems to have some influence on No. of trips: Higher temperature lead to more demand.

Weekday



Weekend



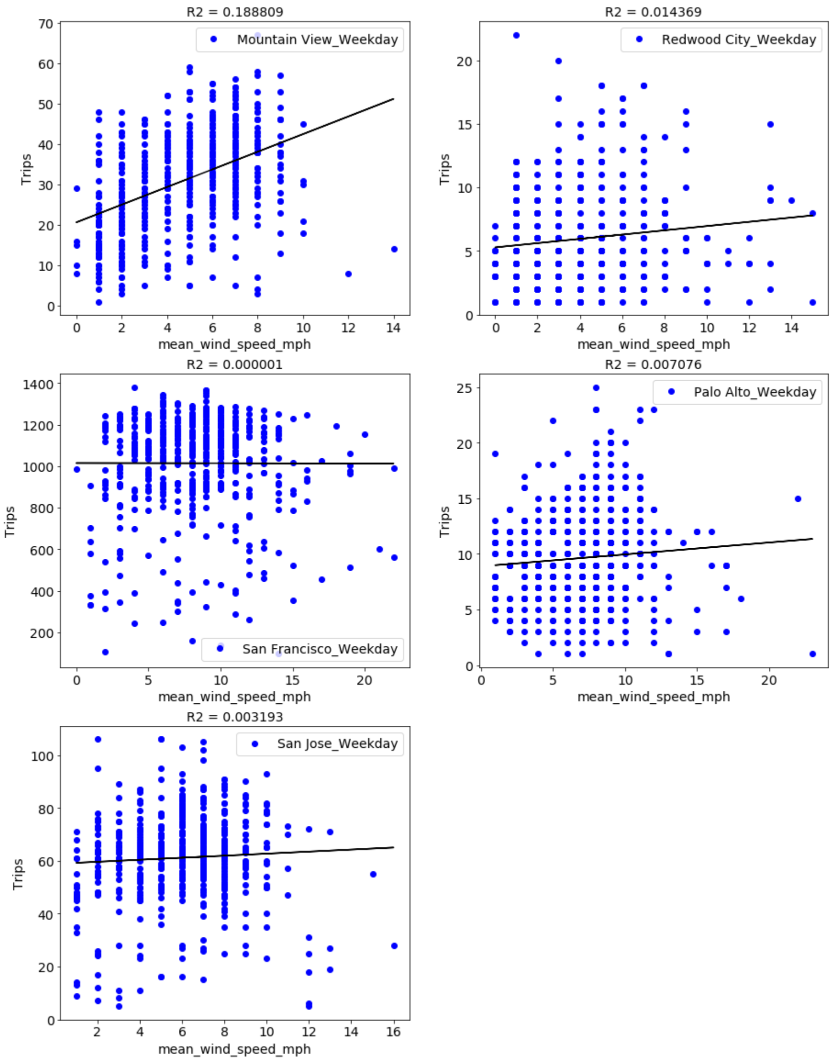


Period for analysis：
Aug. 2013/8~Aug. 2015

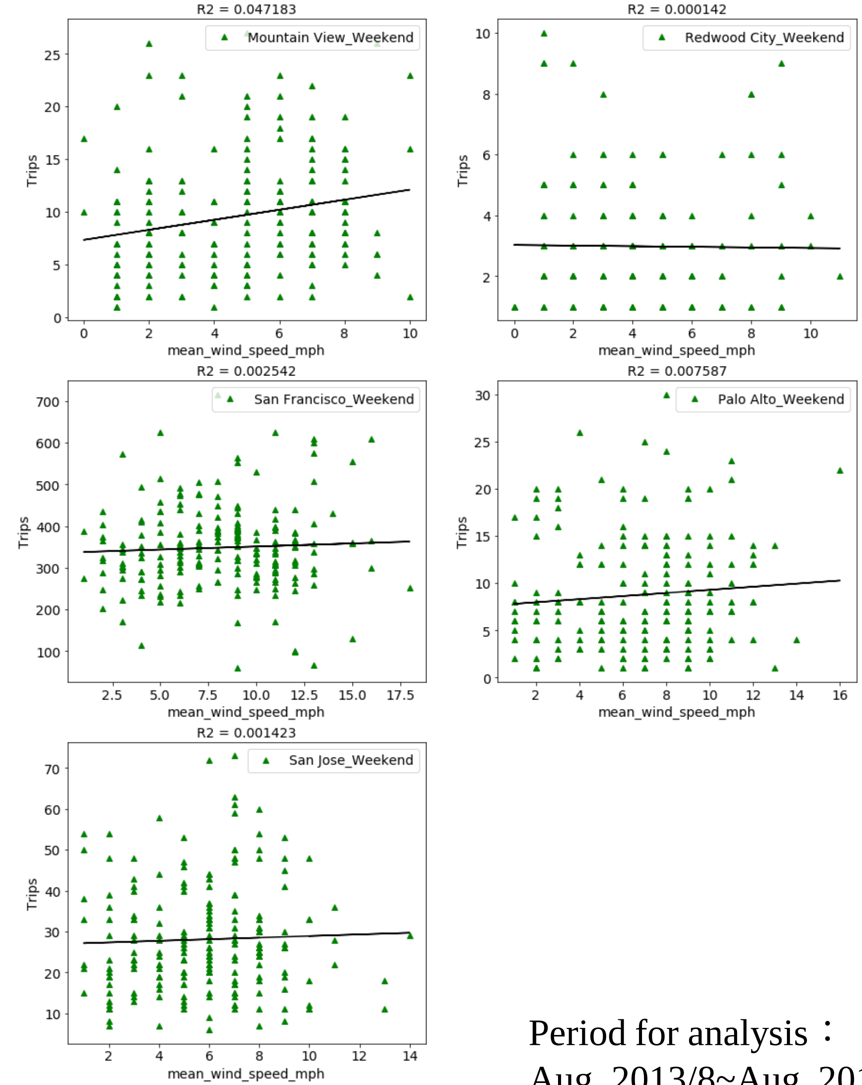
Influence factors on demands:

Ref：On the other hand, wind speed seems have little correlation with the demand.

Weekday



Weekend



Period for analysis：
Aug. 2013/8~Aug. 2015

1. Basic Understandings for the Share Bike Business

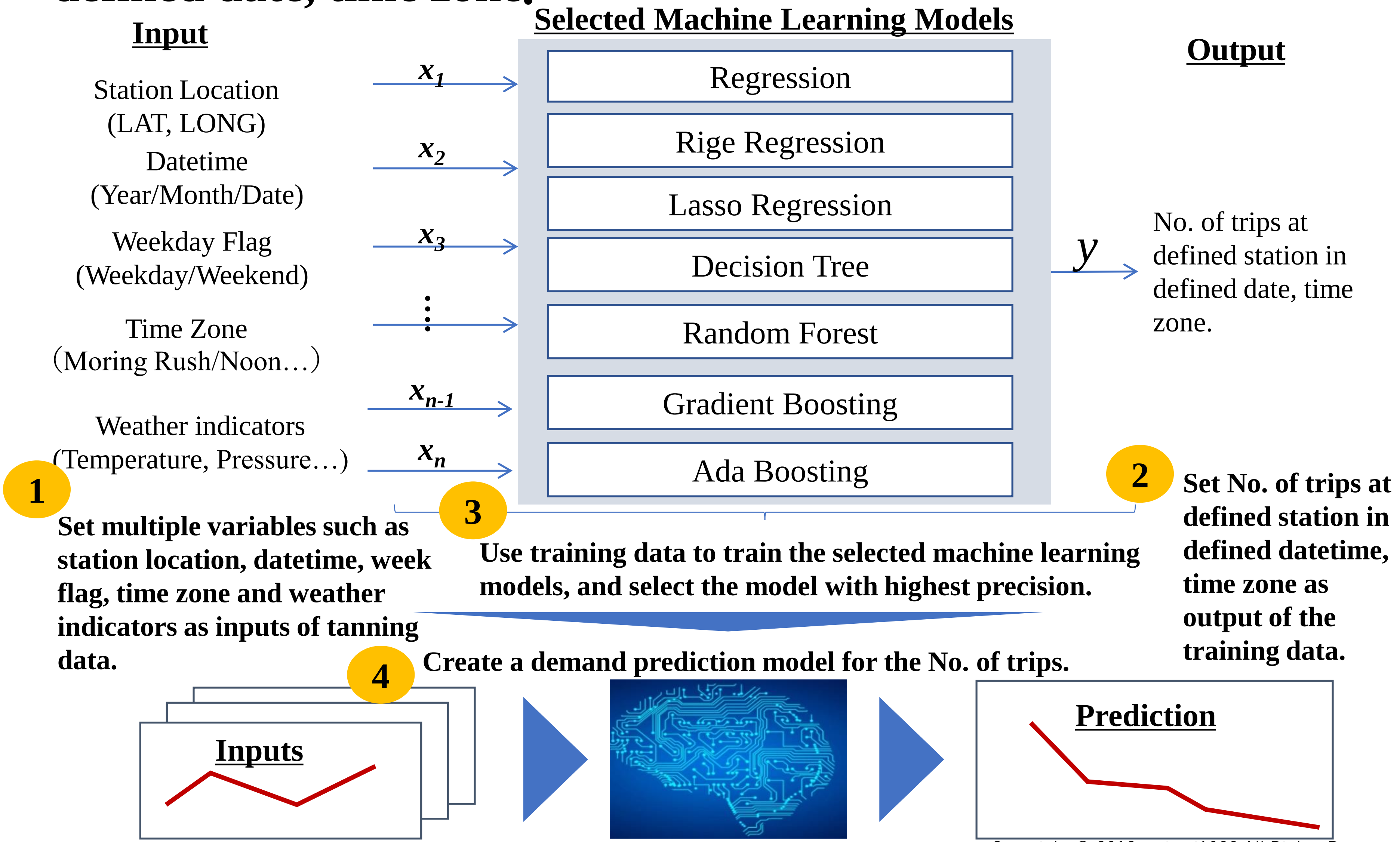
2. Analysis on the Influence Factors on demands

3. A Demand Prediction Model Utilizing Machine Learning

4. The Dynamic Pricing Model Based on the Demand Prediction Model

The Demand Prediction Model:

A demand prediction model based on machine learning was proposed to forecast the No. of trips at defined station on defined date, time zone.



Preconditions of the tranining:

Explanatory and explained variables

Period for analysis：Aug. 2013/8~Aug. 2015

*the NaN of weather.csv were filled with fillna(method = ‘pad’)

Explanatory variables	Factor/Unit	Data Source
Latitude of the Station (LAT)	deg.	station.csv
Longitude of the Station (LONG)	deg.	station.csv
Year	2013/2014/2015	trip.csv
Month	1/2/3/4/5/6/7/8/9/10/11/12	trip.csv
Day	1-31	trip.csv
Weekflag	Weekday（Mon/Tue/Wed/Thu/Fri） Weekend（Sat/Sun）	Variable create using conduction branch from “Weekday” in trip.csv
Timezone	Moring Rush (6-10) Noon(11-15) Evening Rush(16-20) Night(21-5 on next day）	Variable create using conduction branch from “Weekday” in trip.csv
Weather Events	No_RainFog(No events: NaN)/ Rain/Fog-Rain/Rain-Thunderstorm	weather.csv*
Mean temperature	F	weather.csv*
Mean humidity	%	weather.csv*
Mean wind speed	mph	weather.csv*
Mean sea level pressure	inches	weather.csv*
Cloud cover	0/1/2/3/4/5/6/7/8	weather.csv*
Mean visibility	miles	weather.csv*
Precipitation	inches	weather.csv*
Explained Variable	**Factor/Unit**	**Data Source**
No. of Trips	times	Grouped by explanatory variables from trips.csv

Preconditions of the trainning:

Selected Models, Datasets for Tanning and Validation.

- **Selected Machine Learning Models**
  - Regression
  - Rige Regression
  - Lasso Regression
  - Decision Tree
  - Random Forest
  - Gradient Boosting
  - Ada Boosting
- **Datasets for Training**
  - 80% of random samples from Aug. 2013 to Aug. 2015
- **Datasets for Validation**
  - 20% of random samples from Aug. 2013 to Aug. 2015
- **Cross Validation**
  - 5 times
- **Python library for Tanning**
  - Scikit learn

Model Validation：

Radom Forest Regression was selected as prediction model because of the highest precision scores.

Selected Models	Mean R2	Negative Mean Squad Error
Regression	0.24	-40.02
Rige Regression	0.24	-40.02
Lasso Regression	0.00	-52.65
Decision Tree	0.68	-16.68
Random Forest	**0.85**	**-8.22**
Gradient Boosting	0.60	-21.05
Ada Boosting	0.40	-31.53

Parameter Tuning：
Random forest regressor was tuned with listed parameters using grid search.

```
tuned_parameters_rdfr = {
                     "max_depth": [2,3, None],
                     "n_estimators":[100, 200, 300],
                     "max_features": [1, 3, 5],
                     "min_samples_split": [2, 3, 10],
                     "min_samples_leaf": [1, 3, 10],
                     "bootstrap": [True, False],
                        }
```

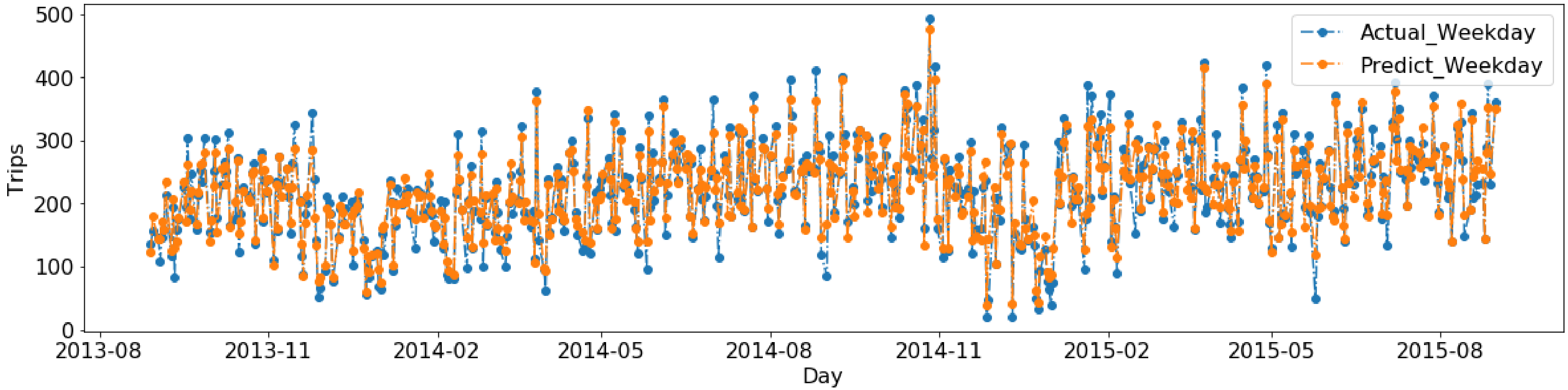
Regressor was tuned using
GridSearchCV method (CV=5)

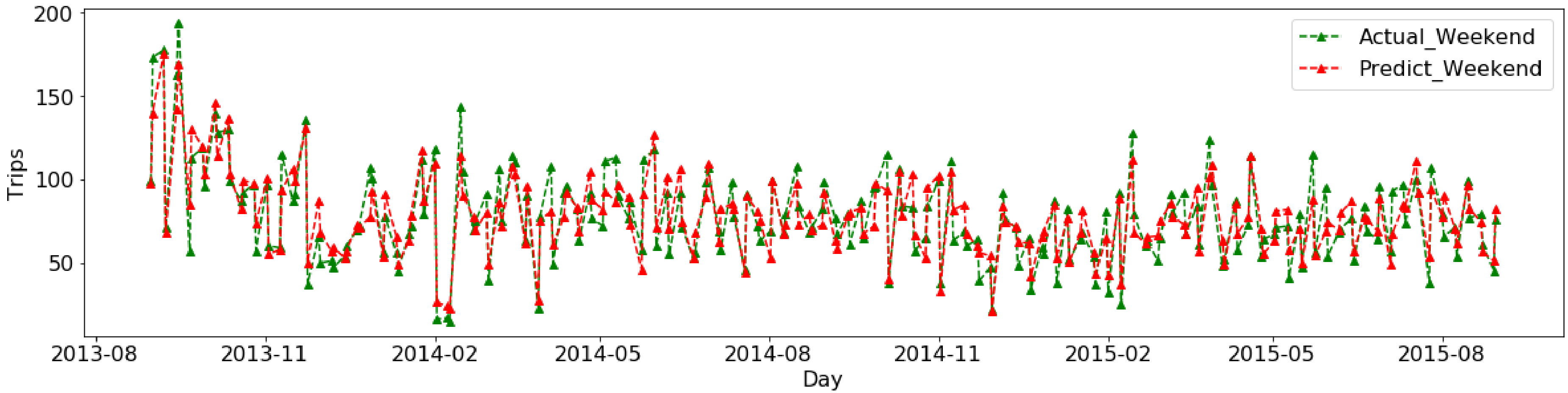
```
max_depth=None, max_features=5, max_leaf_nodes=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1, min_samples_split=10,
min_weight_fraction_leaf=0.0, n_estimators=300
```

Prediction results:

The No. of trips by each day were relatively well predicated by the machine learning model for practical usage.

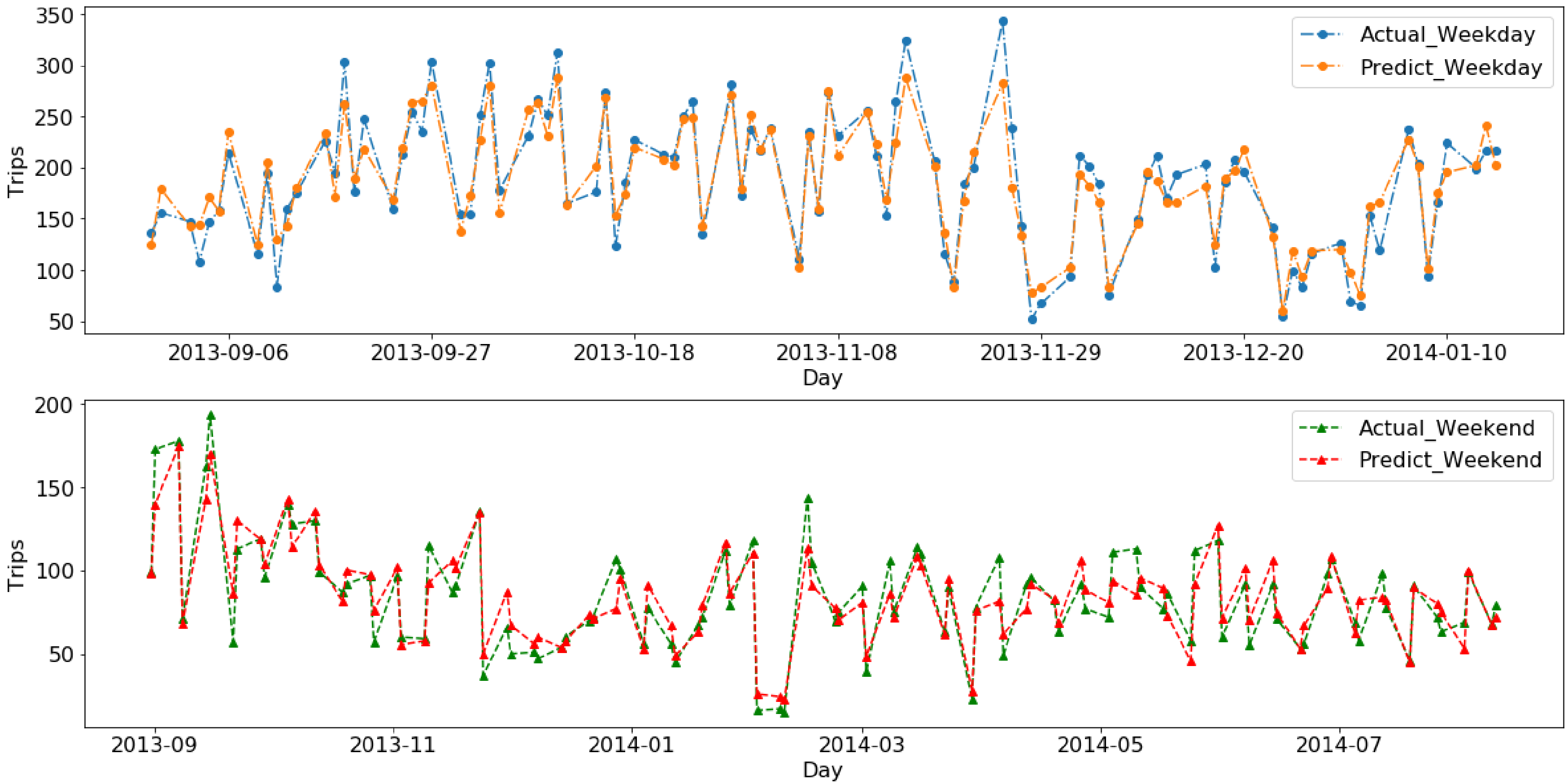
Actual and predicted No. of trips by day (cumulative sum for all stations)





Prediction results:
Ref: Results for head 100 datasets.

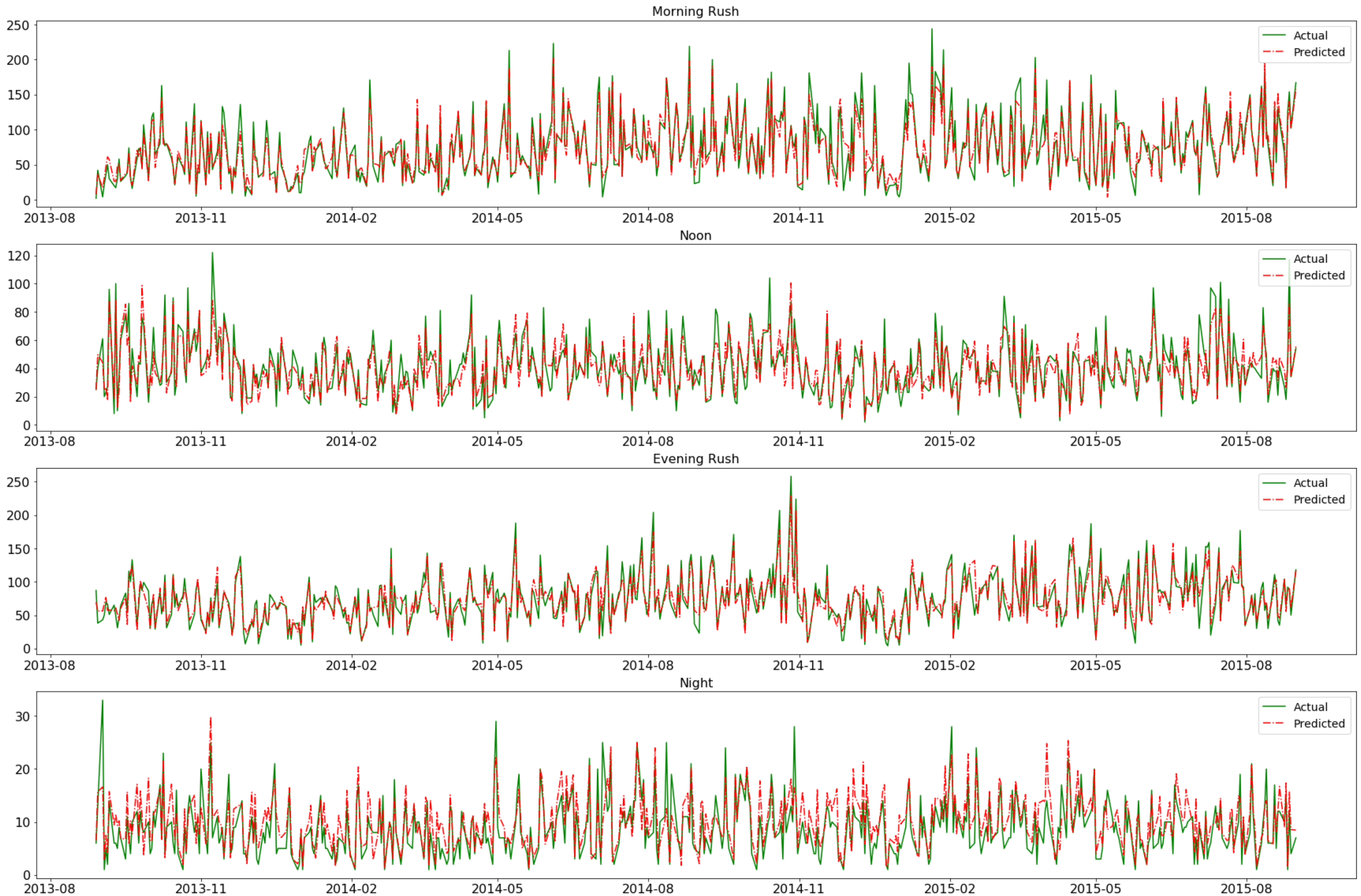
Actual and predicted No. of trips by day (cumulative sum for all stations)



Prediction results：

The No. of trips by each time zone were relatively well predicated by the machine learning model for practical usage.

Actual and predicted No. of trips by time zone (San Francisco, Weekday）

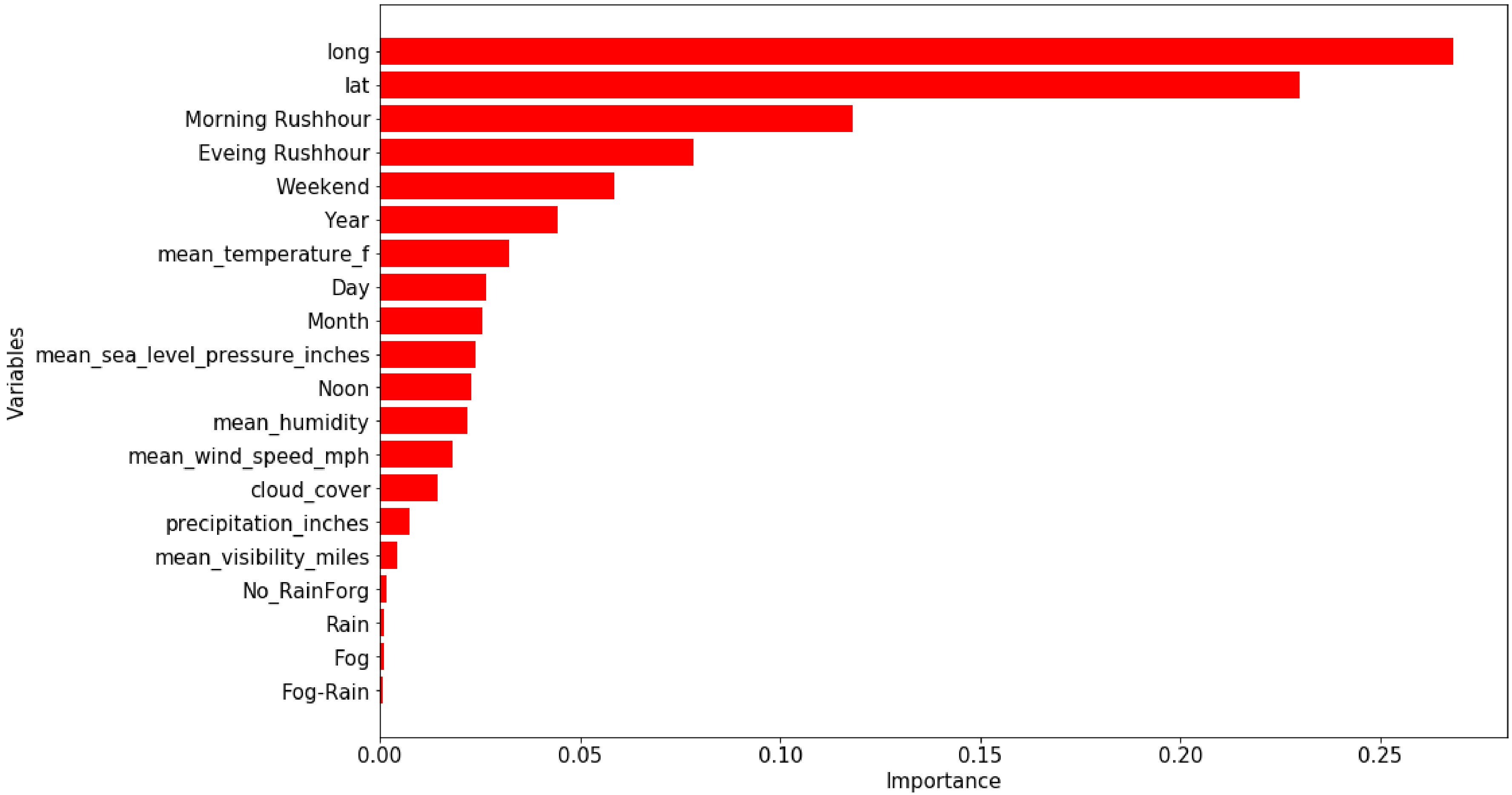


Prediction results:
Ref: Results for head 100 datasets.

Actual and predicted No. of trips by time zone (San Francisco, Weekday）

Importance score:
Variables with relatively high importance score are: Location(Long&Lat), Timezone, Weekflag, Year and Temperature

Importance score of each explanatory variables



1. Basic Understandings for the Share Bike Business

2. Analysis on the Influence Factors on demands

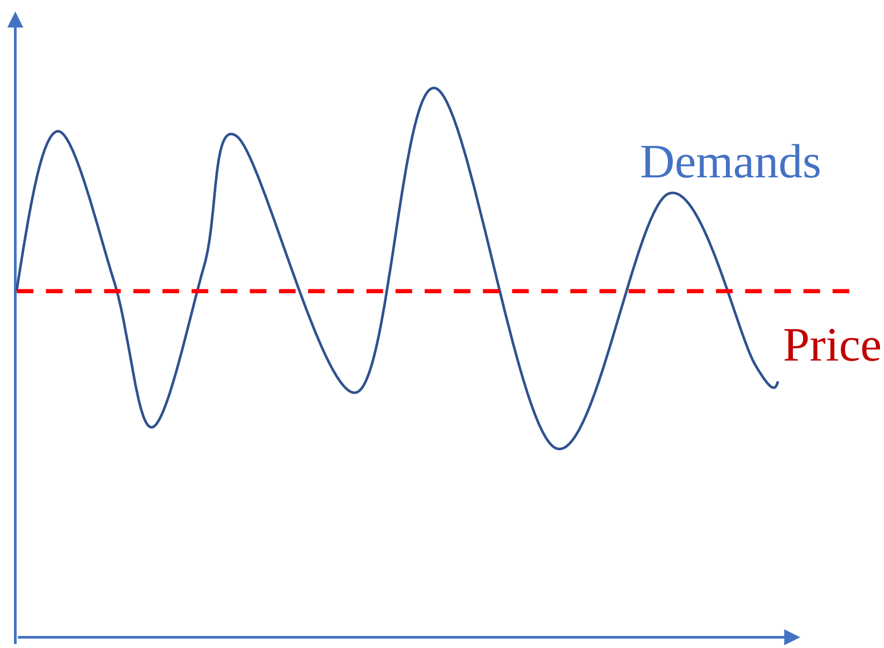
3. A Demand Prediction Model Utilizing Machine Learning

4. **A Dynamic Pricing Model Based on the Demand Prediction Model**

Dynamic Pricing Model:

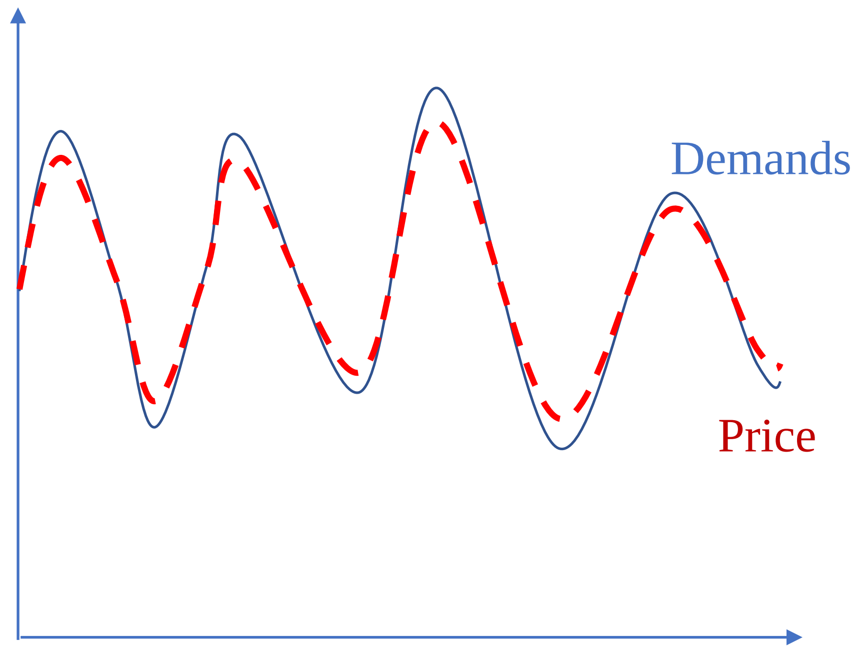
A Dynamic pricing model based on the demand prediction model instead of current inflexible pricing plan was proposed to solve the demand supply un-matching problem.

Current Pricing Plan



Price was constant in spite of fluctuate demand, which lead to demand supply gap.

Proposed Pricing Model



Price will be set based on the demand prediction results by machine learning, which could balance the demand and supply.

Model Example:

$$\textbf{Price Fluctuation Rate\%(t)} = \textbf{(1/P.E.)} \times \textbf{Demand Fluctuation Rate\%(t)}$$

Price Fluctuation Rate = Set Price / Current Constant Price
Demand Fluctuation Rate = Predicted Demand / Average Demands
P.E. : Price Elasticity

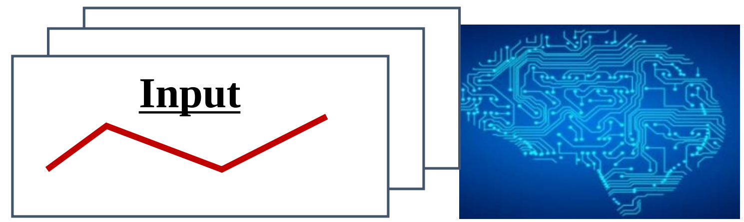
How to benefit customers?

Integrate the dynamic pricing model to the APP, Provide customers with real-time optimized price.

1 **Get the necessary data from user app.**

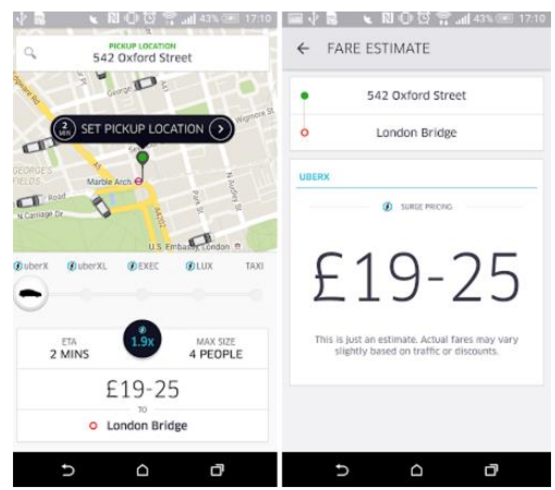
Station Location/
Date/Weather…

2 **Predict the demand at the defined condition**



3 **Use the dynamic pricing model to calculate the optimized price at this point.**

4 **Offer the price on the app, allow customers to make payments at real time.**



Ex: Price fluctuation rate to offer.

Price Fluctuation Rate (San Francisco, Weekday）

Price Elasticity = 10