

# 機械学習を活用した 自転車シェアリングの需要予測 及び価格変動モデルのご提案

日産自動車株式会社  
市場情報統括本部  
趙 惟恒

**1. 自転車シェアリングビジネスの現状に対する理解**

**2. 需要変動の影響要因分析**

**3. 機械学習による需要予測モデルの構築**

**4. 需要予測に基づく価格変動モデルの提案**

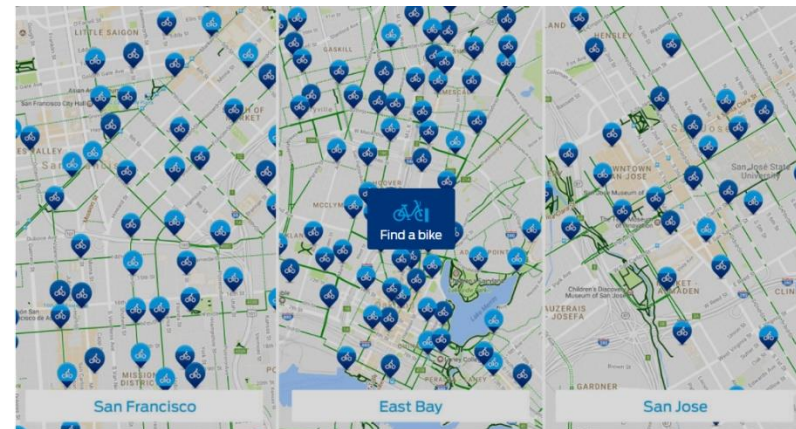
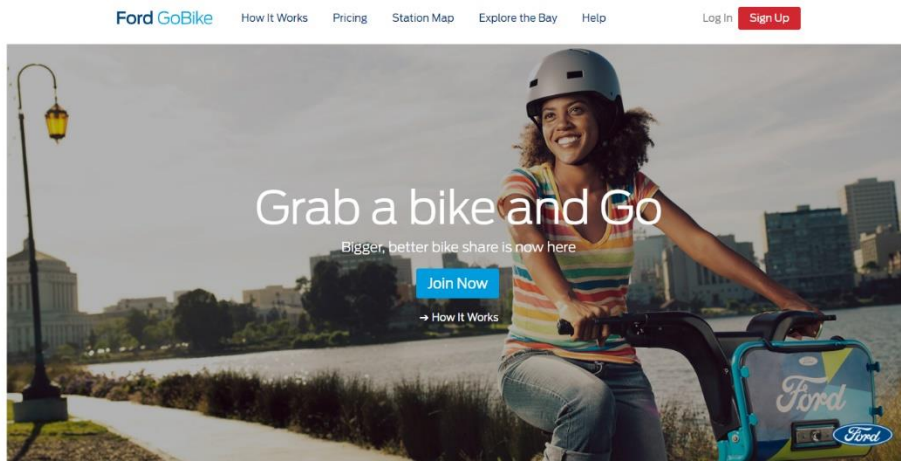
## **1. 自転車シェアリングビジネスの現状に対する理解**

## **2. 需要変動の影響要因分析**

## **3. 機械学習による需要予測モデルの構築**

## **4. 需要予測に基づく価格変動モデルの提案**

# 貴社 (Ford Motor) はバイエリアで自転車のシェアリングのビジネスを展開している



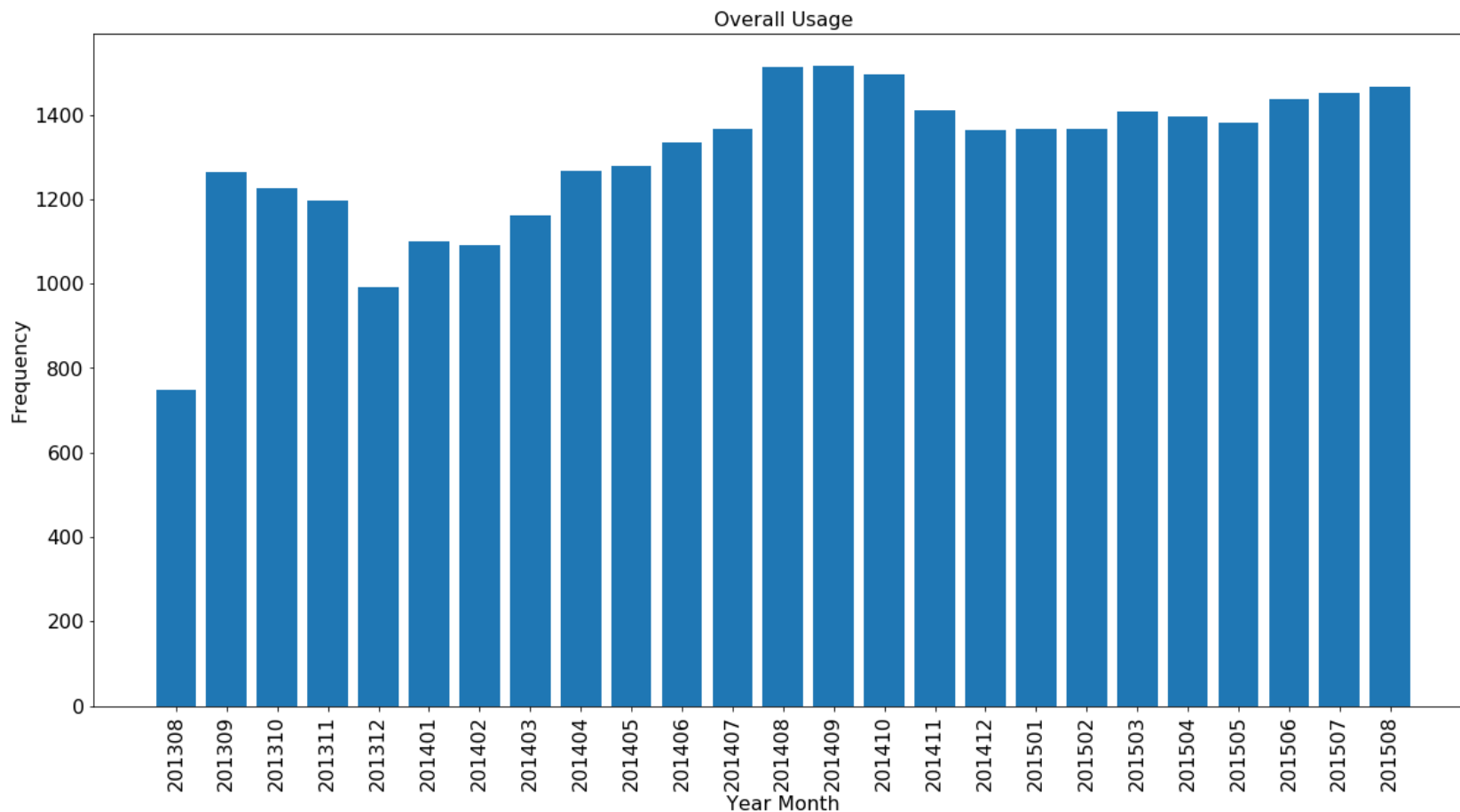
ステーション数 (2015年8月時点)



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

# 2014年10月頃をピークに需要は頭打ち傾向

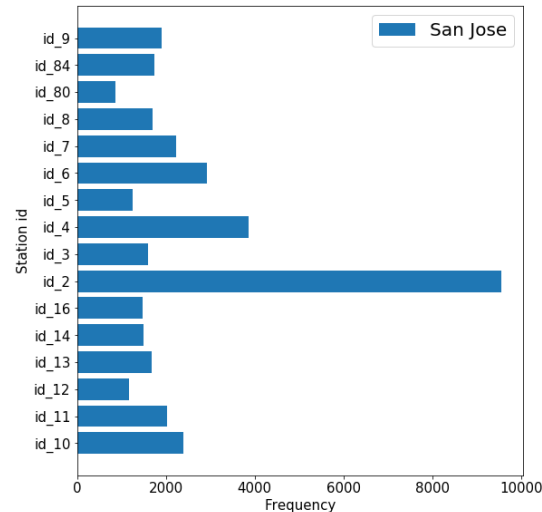
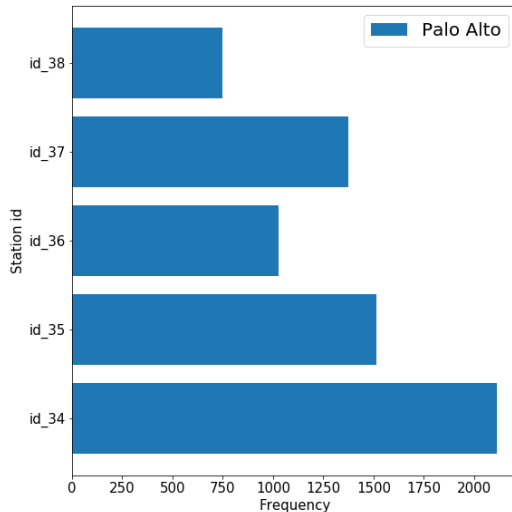
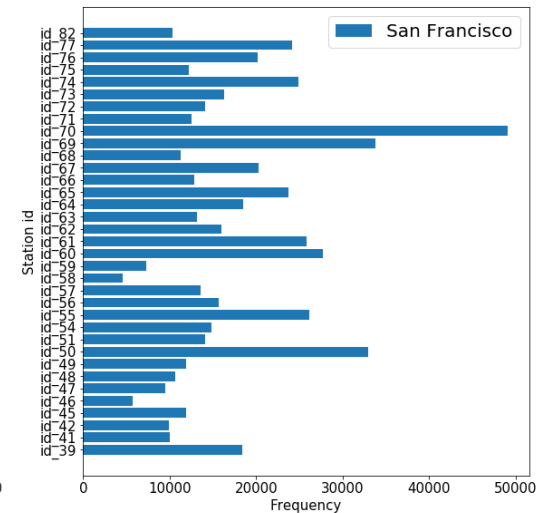
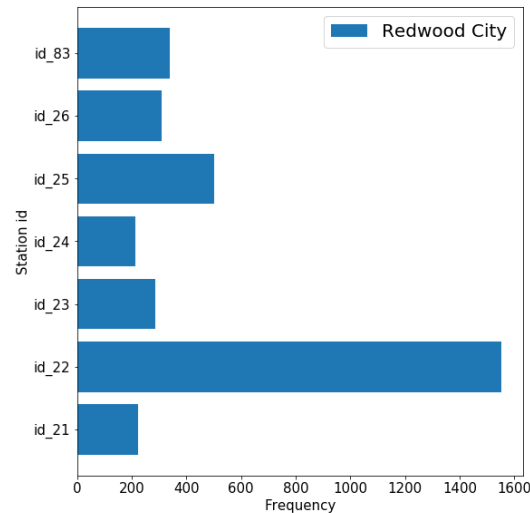
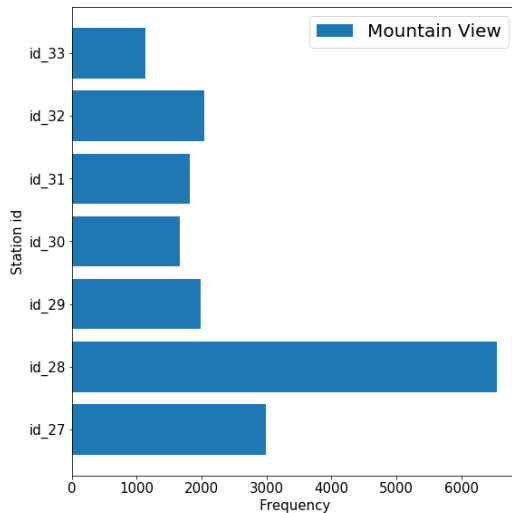
## 1ヶ月ごとの利用数推移 [回]



需要が頭打ちになる要因：

そもそも需要数は場所などによってオーダーが大きく異なる

例えば、San Francisco, San Joseなどのダウンタウンでは利用数5万回を超えているところもあるが、Palo Altoのような学生街は2千回前後。

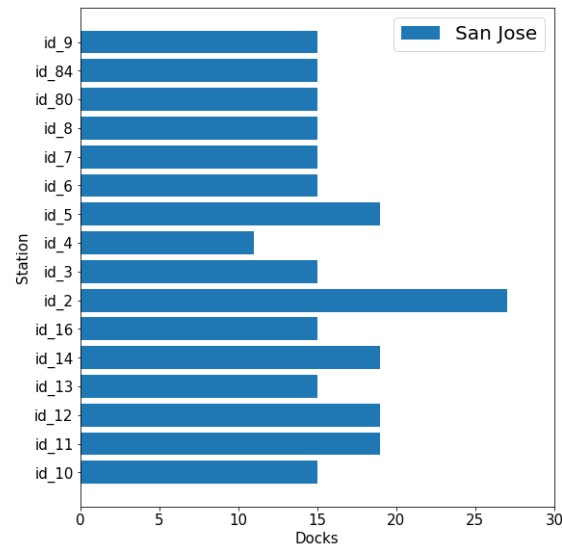
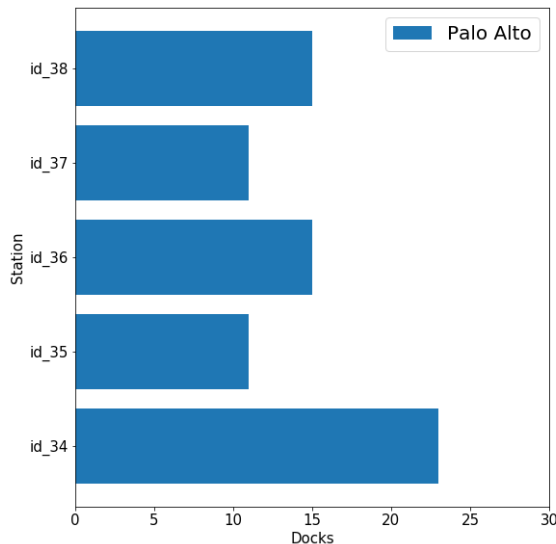
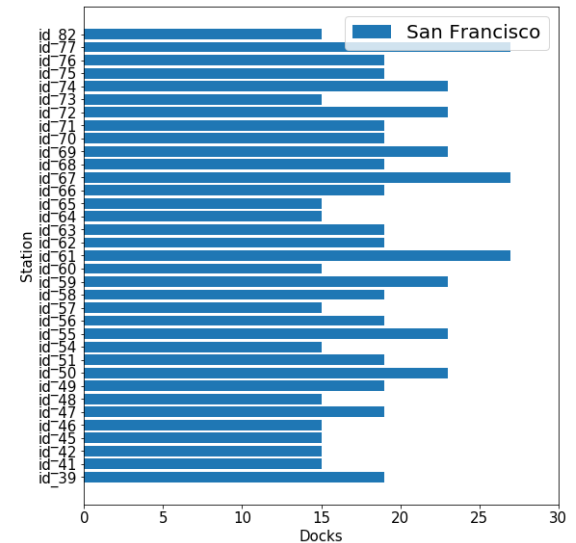
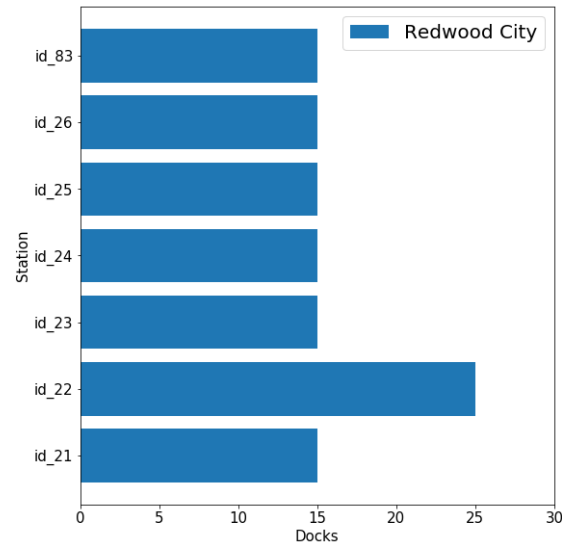
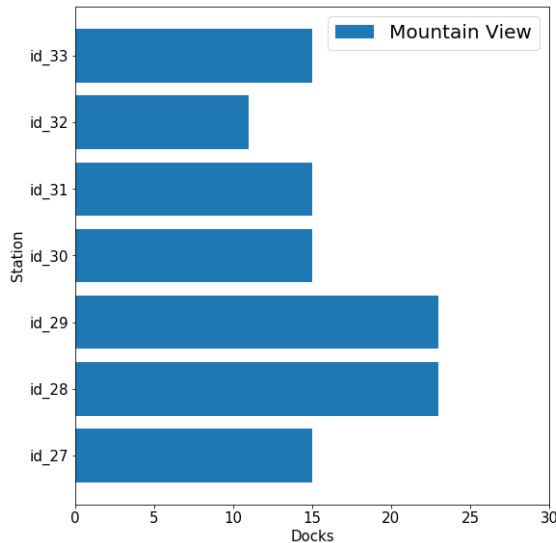


ほかにも、時間帯、月、位置、天気など様々な影響要因が考えられる

集計期間：2013/8~2015/8の3年間

需要が頭打ちになる要因：

しかし、ステーションのdock数は各都市20前後で平準化されており、需要に合わせた数を配置できていない



集計期間：2013/8~2015/8の3年間

# 需要が頭打ちになる要因： 価格プランも膠着的、需要に合わせた柔軟な設定 が出来ていない

現状の価格設定（非会員）

一回：3ドル(30min上限)

一日：9.95ドル(24時間；30min制限)

一年：149ドル（年間；45min制限）

## Choose your plan

Only available in the app

### Single Ride

\$3 / trip

Ride from point A to B  
with 30 minutes of ride  
time.

[See Details](#)

Perfect for the explorer

### Day Pass

\$9.95 / day

Explore the city with  
unlimited 30-minute trips  
in a 24-hour period.

[See Details](#)

Best deal for locals

### Annual Membership

\$149/year

Unlimited 45-minute trips.

[Join Now](#)

→ Or pay in installments  
(\$14.90/month)

Best Value

### Bike Share for All

Low-income residents qualify  
for a discounted membership.

[→ Learn More](#)

### More Options

[→ Corporate Memberships](#)  
[→ Bulk Passes](#)



# 利用数の頭打ち傾向を改善するには、需要を正確に把握し、 利用ニーズの取りこぼし/無駄を解消する必要がある

需要数は月、日、場所など様々な要因によって変動する

ステーション**docks**数、  
価格設定は膠着的

受給のアンマッチが発生し、お客さんの利用ニーズ  
を取りこぼし/無駄を引き起こしている

近い未来の利用数を正確に予測し、  
需要に合わせた柔軟な価格、**docks**数設定が必要

機械学習による需要予測モデルのご提案

**1. 自転車シェアリングビジネスの現状に対する理解**

**2. 需要変動の影響要因分析**

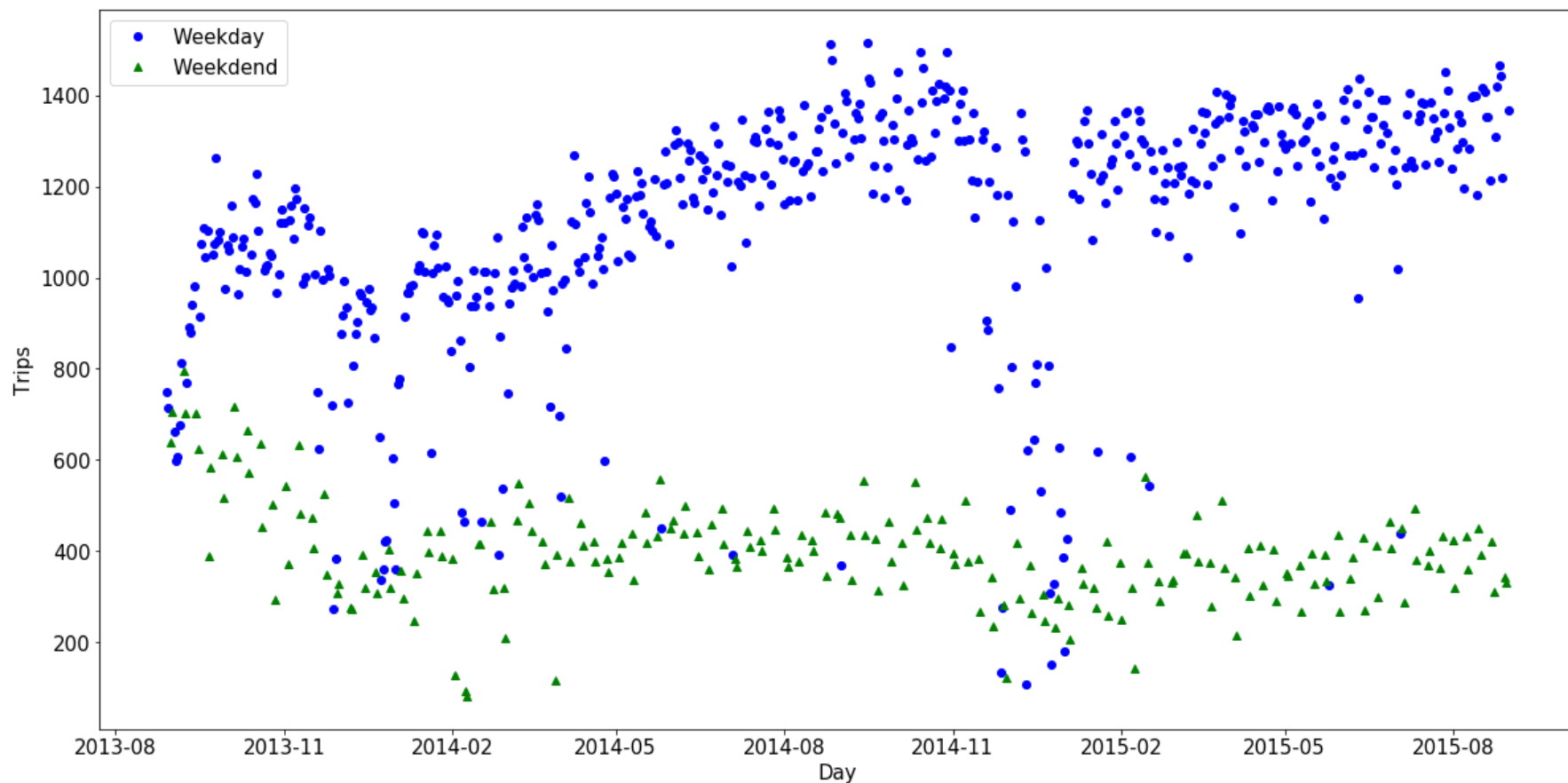
**3. 機械学習による需要予測モデルの構築**

**4. 需要予測に基づく価格変動モデルの提案**

**需要変動の影響要因：**  
**そもそも、平日/休日によって需要が大きく異なる**  
**平日の需要は休日の2-3倍前後**

**一日ごとの利用数推移 [回]**

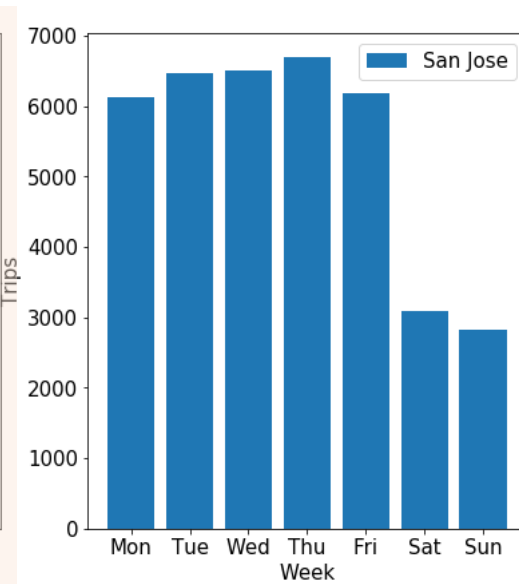
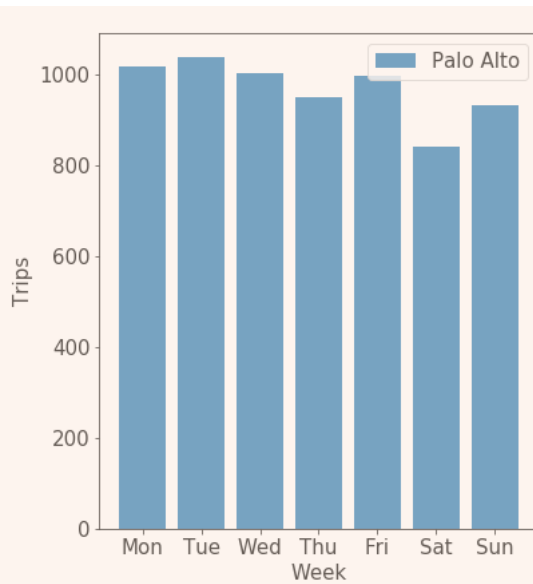
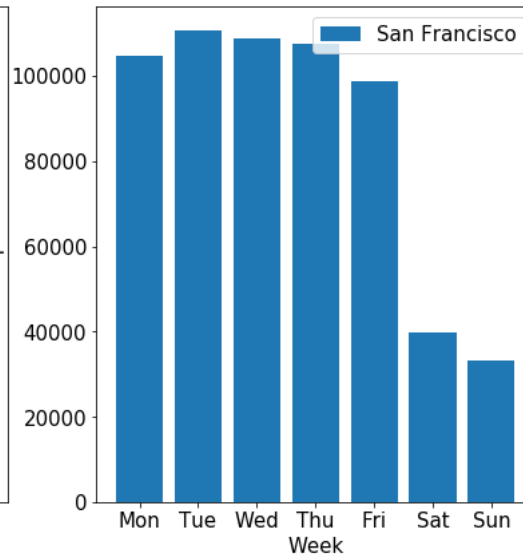
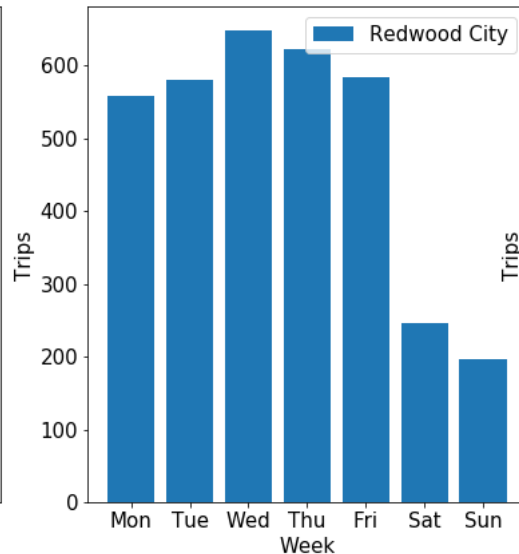
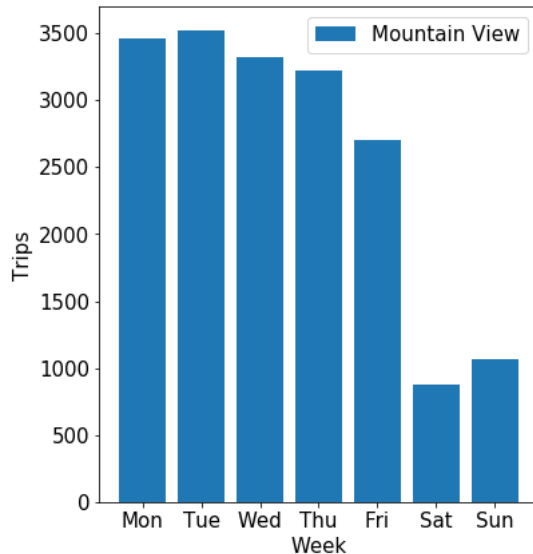
Weekday: Mon/Tue/Wen/Thu/Fri  
Weekend: Sat/Sun



集計期間：2013/8~2015/8の3年間

需要変動の影響要因：

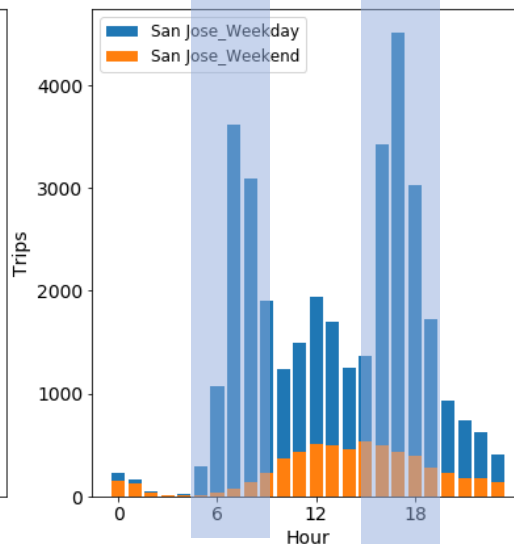
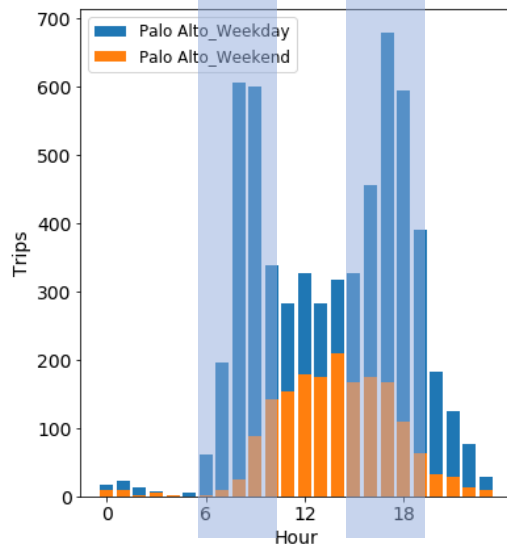
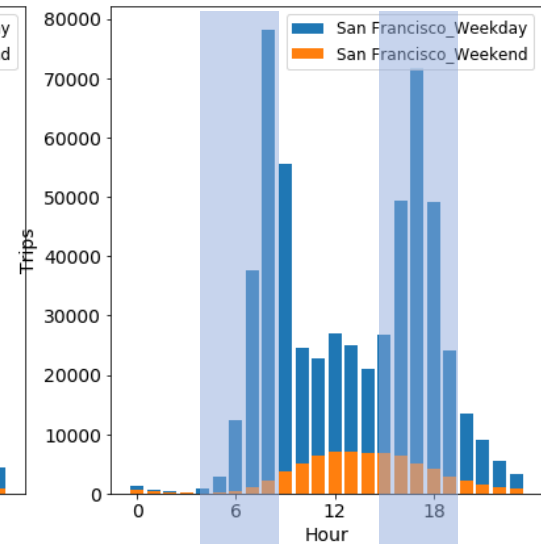
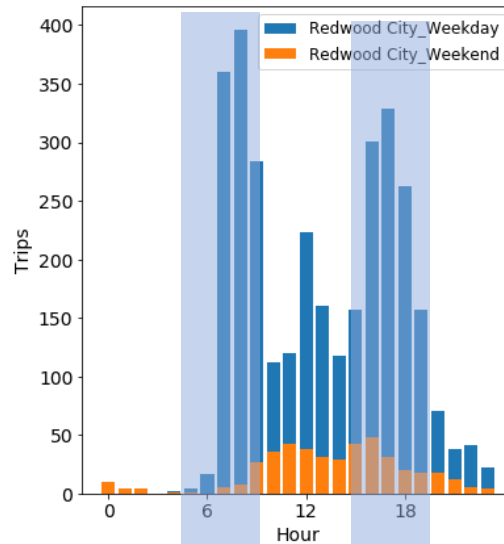
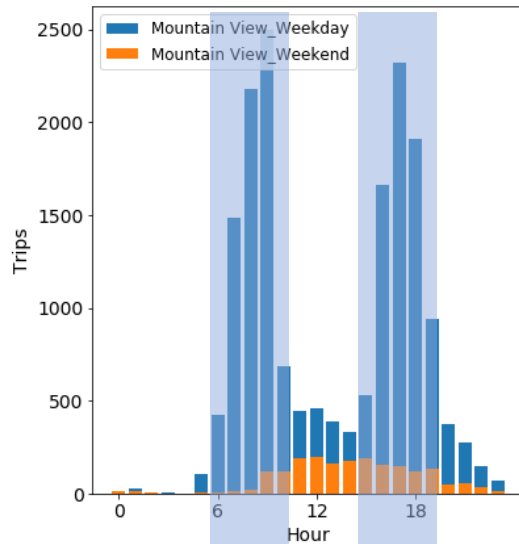
曜日の影響は都市によってもことなる。ダウンタウンは休日の利用が激減するが学生街は減らない



集計期間：2013/8~2015/8の3年間

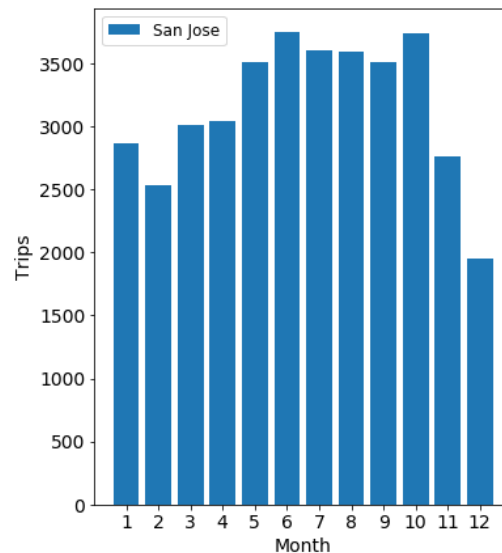
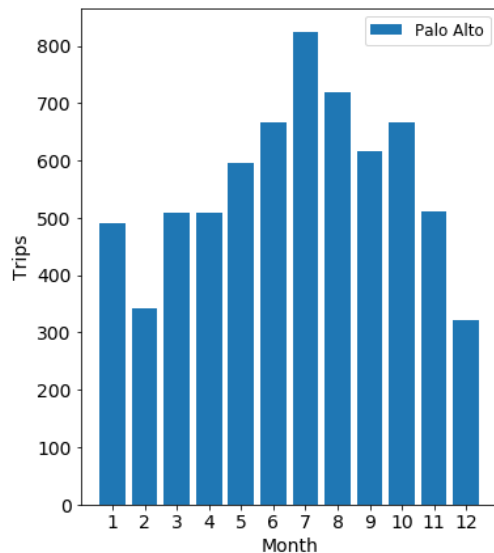
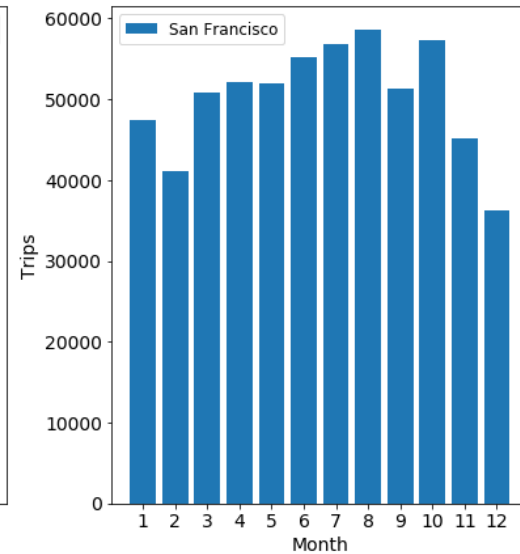
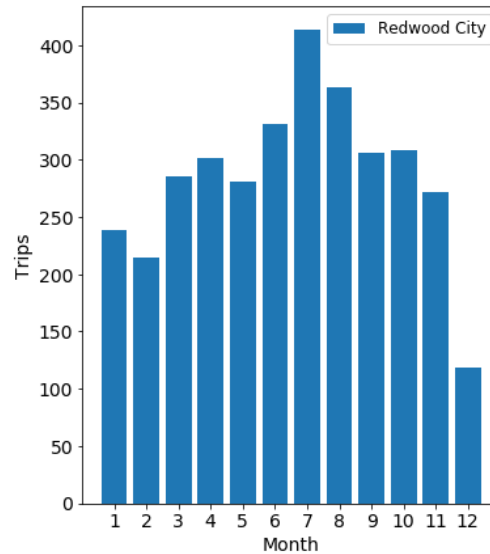
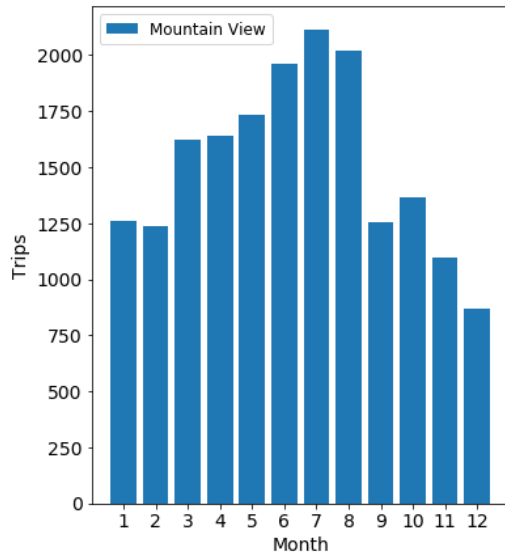
需要変動の影響要因：

時間帯によっても大きく変動する、平日の朝・夜ラッシュ時は利用数が高く、昼と夜は利用が少ない



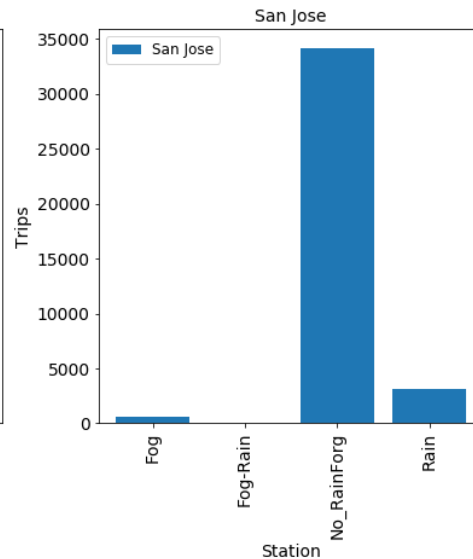
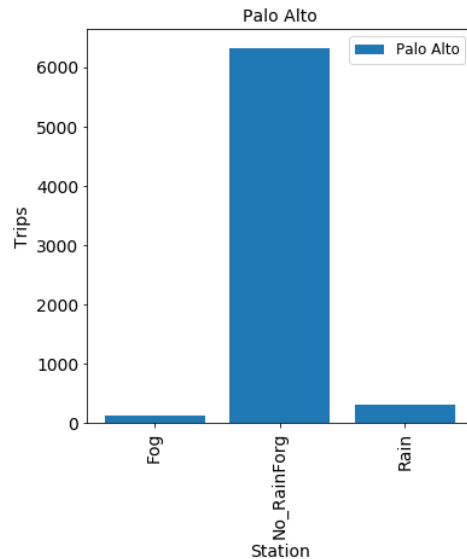
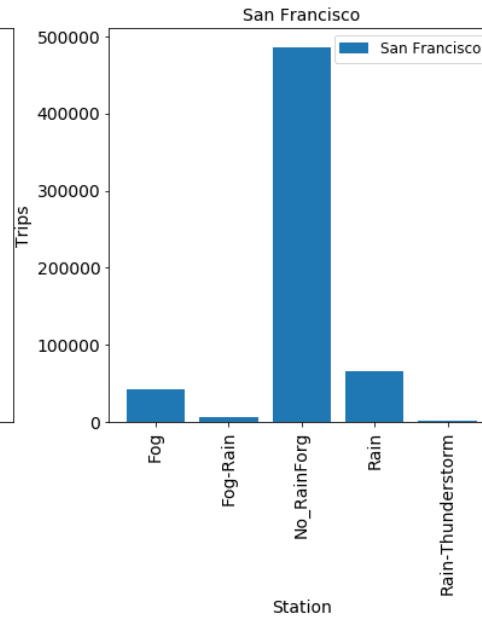
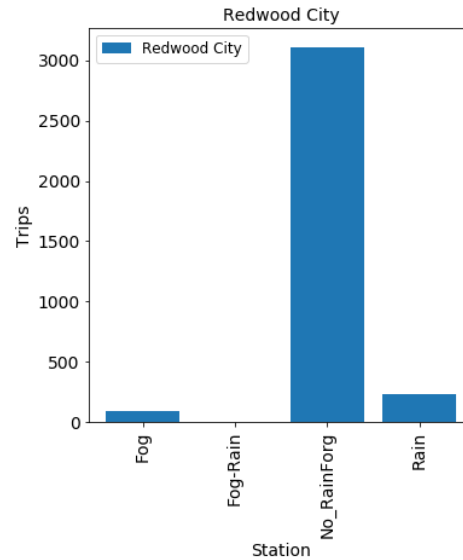
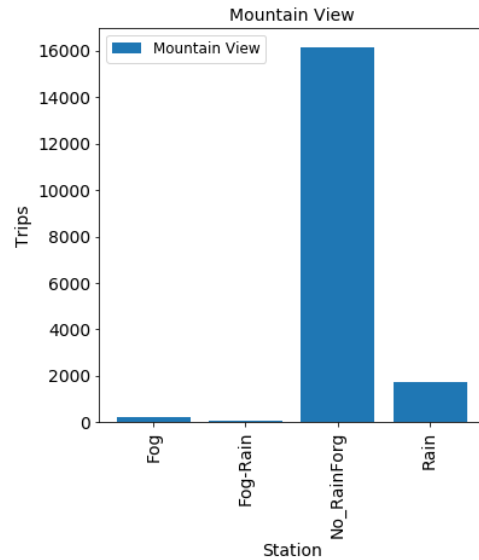
集計期間：2013/8~2015/8の3年間

# 需要変動の影響要因： 全体として夏季、特に7月の利用数が高く、 冬季、特に12月の利用が少ない



集計期間：2013/8~2015/8の3年間

# 需要変動の影響要因： 雨、霧、雷など気象イベントがない日（No\_RainForg） の利用が多い

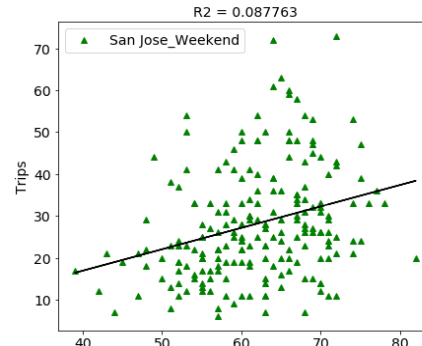
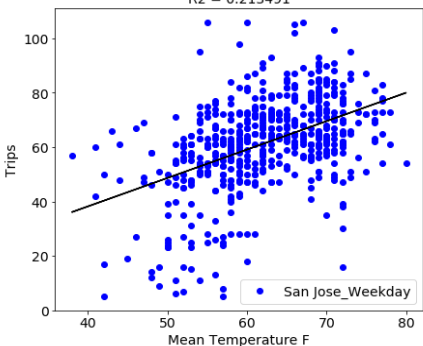
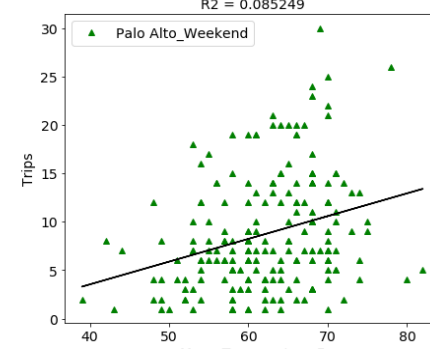
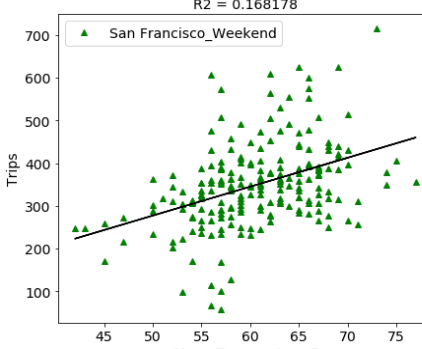
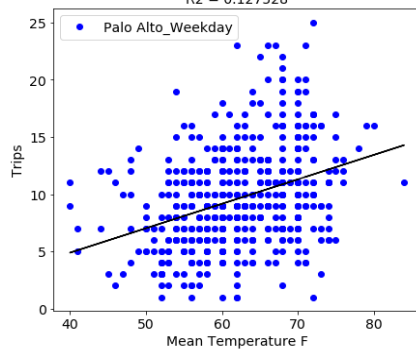
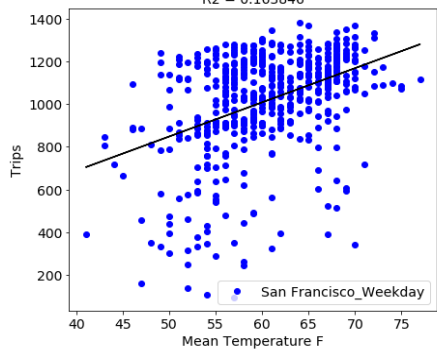
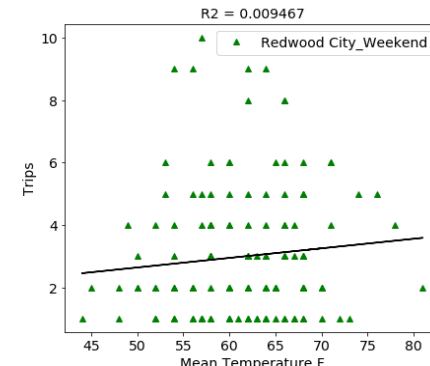
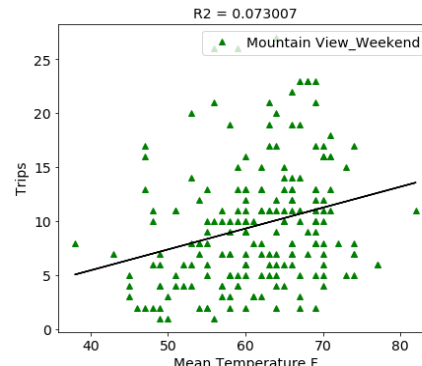
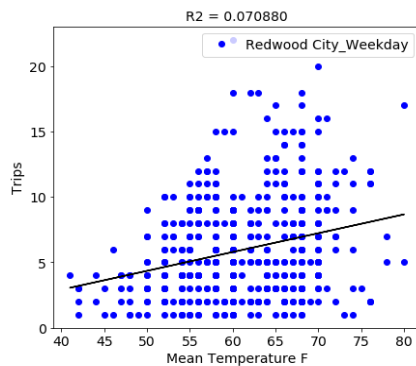
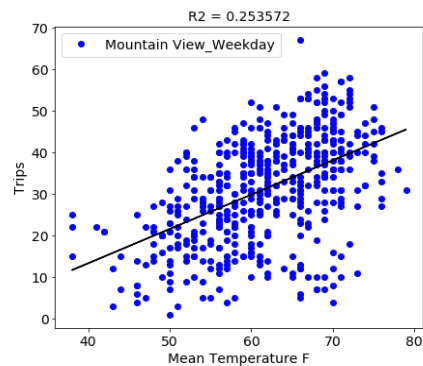


集計期間：2013/8~2015/8の3年間

# 需要変動の影響要因： 気温による影響も若干ある模様 気温が高いほど利用数が若干増える傾向にある

平日

休日



集計期間：  
2013/8~2015/8の3  
年間

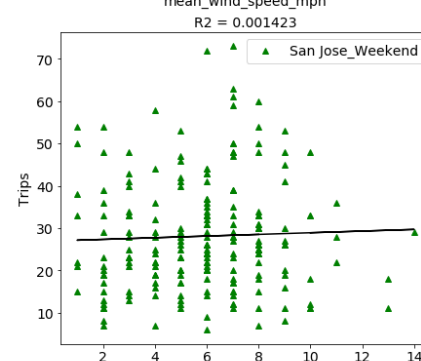
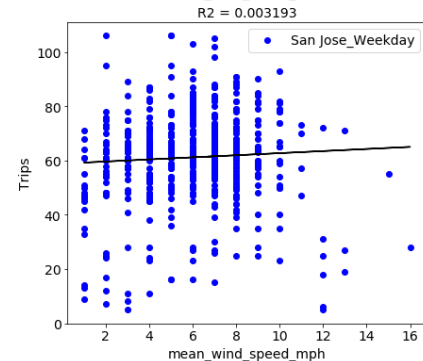
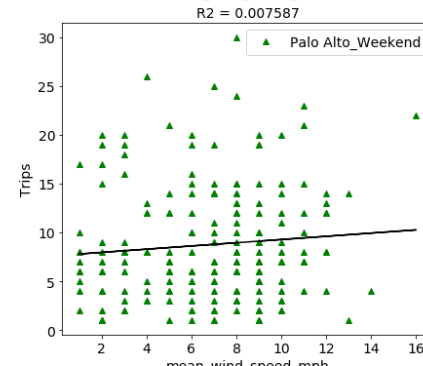
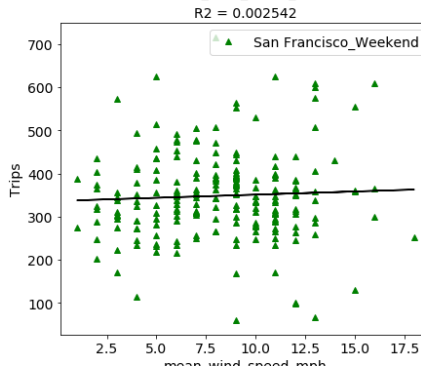
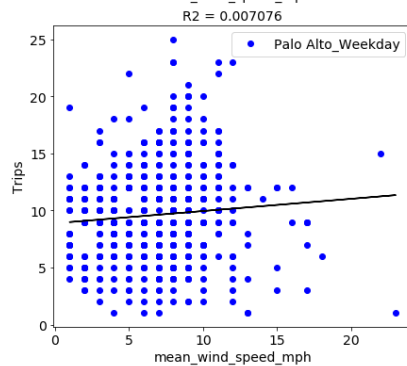
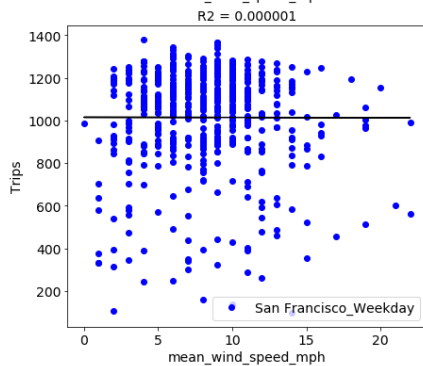
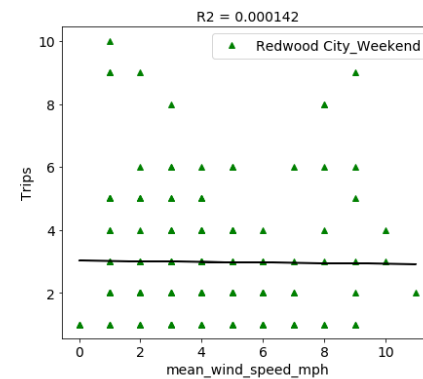
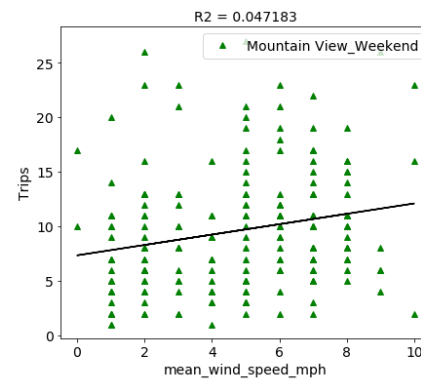
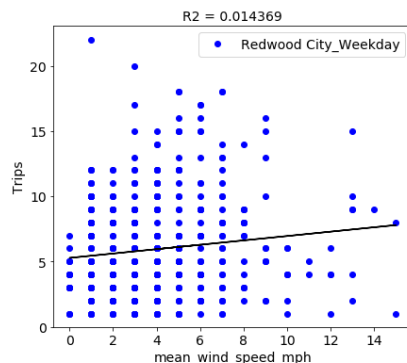
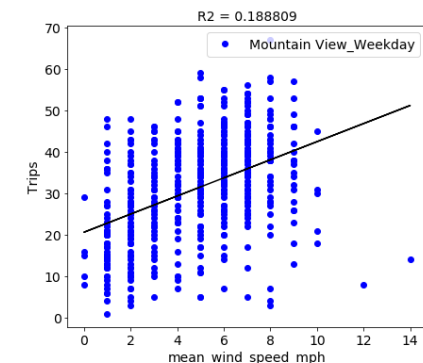


需要変動の影響要因：

参考：風の強さとはあまり関係がない模様

このように、天気と利用数も関連があると思われる

平日



休日

集計期間：  
2013/8~2015/8の3  
年間

**1. 自転車シェアリングビジネスの現状に対する理解**

**2. 需要変動の影響要因分析**

**3. 機械学習による需要予測モデルの構築**

**4. 需要予測に基づく価格変動モデルの提案**

# 需要予測モデルのご提案：

ステーション位置/月/日付/時間帯/天気などを説明変数とし、ステーションの時間帯別利用数の予測するモデルを機械学習で構築

## 入力

ステーション位置  
(LAT, LONG)  
年月日  
(2013/2014/2015)  
平日/休日  
(Weekday/Weekend)  
時間帯  
(ラッシュアワーなど)  
...  
気温、湿度など  
天気データ

$x_1$   
 $x_2$   
 $x_3$   
 $\vdots$   
 $x_{n-1}$   
 $x_n$

## 機械学習モデル



## 出力

$y$  → ステーション別、  
時間帯別利用数

1

ステーション位置、年月日、平日休日フラグ、時間帯フラグ、各種天気データを入力データにセット

3

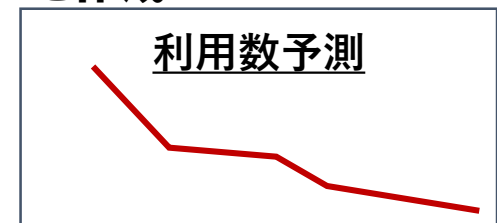
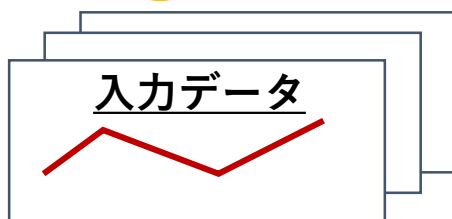
入力と出力を元に、複数モデルによる機械学習を実施し、予測精度が最も高いモデルを選択

2

出力データには、ステーション別時間帯別需要数をセット

4

日々の時間帯別利用数を予測するモデルを作成



# 各種前提条件： 説明変数・被説明変数一覧

2013年8月～2015年8月までのデータを結合、集計

\*weatherデータは欠損値が存在するため、時系列であることを考慮し、直前のデータで補間

説明変数	要因/単位	取得先
ステーション緯度(LAT)	度	Stationデータより取得
ステーション経度(LONG)	度	Stationデータより取得
年	2013/2014/2015	Tripデータ
月	1/2/3/4/5/6/7/8/9/10/11/12	Tripデータ
日	1-31日	Tripデータ
曜日	Weekday (月曜日 ～ 金曜日) Weekend (土曜日、日曜日)	Tripデータより条件分岐で作成、 変数ダミー化
時間帯	Moring Rush (6-10時) Noon(11-15時) Evening Rush(16-20時) Night(21-翌日5時)	Tripデータより 条件分岐で作成 変数はダミー化
天気イベント	No_RainFog(イベントなし)/Rain/Fog-Rain/ Rain-Thunderstorm	Weatherデータ*
平均気温	F	Weatherデータ*
平均湿度	%	Weatherデータ*
平均風速	mph	Weatherデータ*
平均気圧	inches	Weatherデータ*
雲量	0/1/2/3/4/5/6/7/8	Weatherデータ*
可視度	Miles	Weatherデータ*
降水量	Inches	Weatherデータ*
被説明変数	要因/単位	取得先
利用数 (Trips)	回	Tripsデータより、説明変数で groupbyして集計

## 各種前提条件：

# 選択モデル、学習・検証データ、精度検証条件

### • 選択モデル

- 重回帰
- Ridge回帰
- Lasso回帰
- 決定木
- Random Forest
- Gradient Boosting
- Ada Boosting

### • 学習データ

- 2013年-2015年データのうち、ランダムに抽出された8割前後のデータ・セット

### • 検証データ

- 2013年-2015年データのうち、ランダムに抽出された2割前後のデータ・セット

### • 精度検証条件


- 5回のクロスバリデーションによって検証

モデルの選択：  
クロスバリデーションの結果、Random Forestの精度  
が一番高かったため、予測モデルに使用

検証モデル	Mean R2	Negative Mean Squad Error
重回帰	0.24	-40.02
Rige回帰	0.24	-40.02
Lasso回帰	0.00	-52.65
決定木	0.68	-16.68
<b>Random Forest</b>	<b>0.85</b>	<b>-8.22</b>
Gradient Boosting	0.60	-21.05
Ada Boosting	0.40	-31.53

## ハイパーパラメータチューニング： Random Forestモデルについて、以下のパラメータ条件で グリッドサーチによるパラメータチューニングを実施

```
tuned_parameters_rdfnr = {  
    "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],  
    }
```

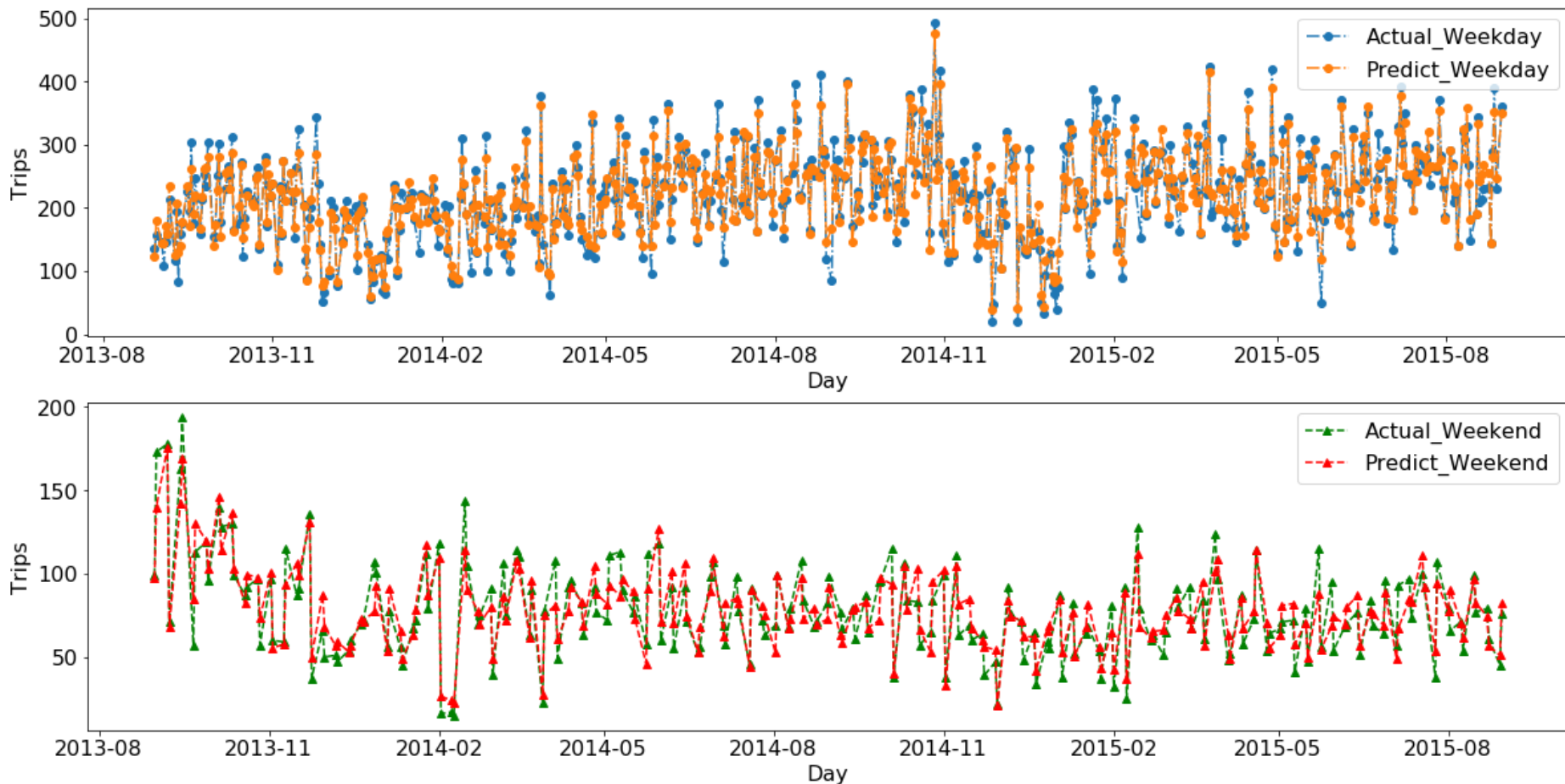


各5回ずつのクロスバリデーションの結果、  
最も精度の良い条件を抽出

```
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
```

**予測結果の確認：**  
**一日ごとの予測結果は、実際の需要をうまく追従できて**  
**いて、実用域に達している**

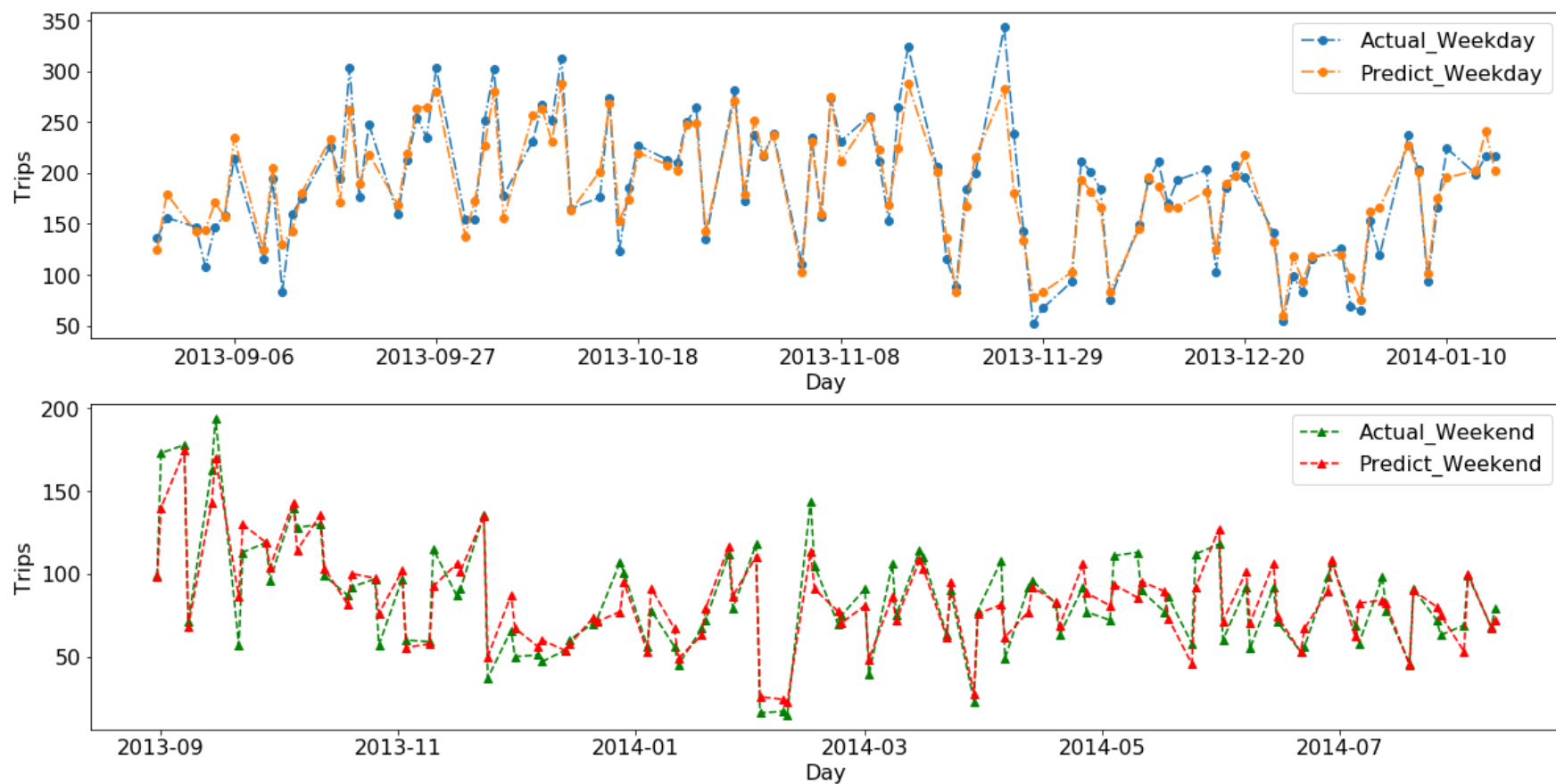
### 一日ごとの予測利用数と実利用数[回]（総数集計）





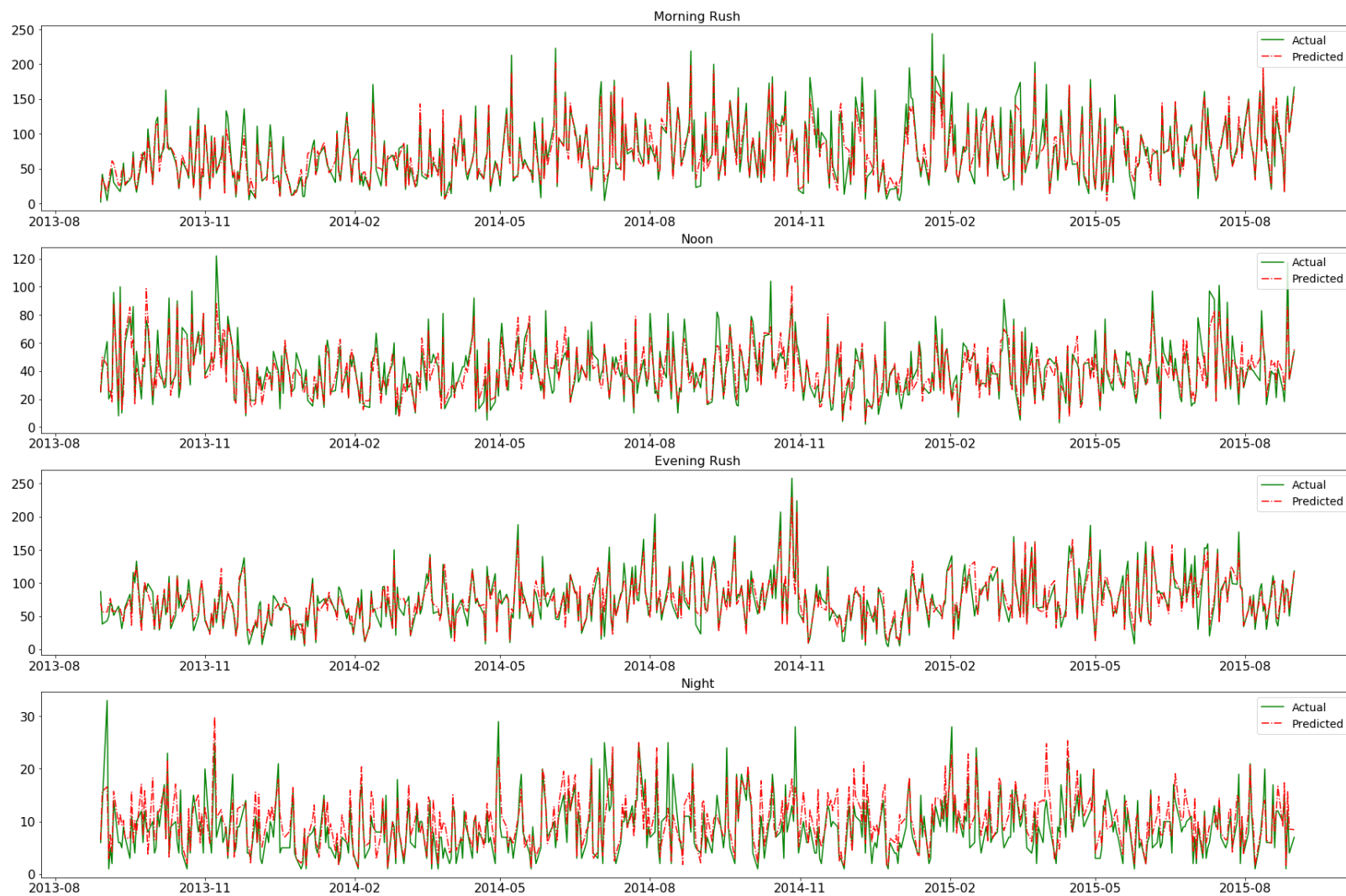
# 予測結果の確認： 参考：100セットだけ取り出した場合の結果

## 一日ごとの予測利用数と実利用数[回]（総数集計）



# 予測結果の確認： 時間帯ごとの予測結果も実需要を比較的うまく追従でき ていて、実用域に達していると思われる

## 時間帯ごとの予測利用数と実利用数〔回〕（San Francisco市、平日）



# 予測結果の確認： 参考：100セットだけ取り出した場合の結果

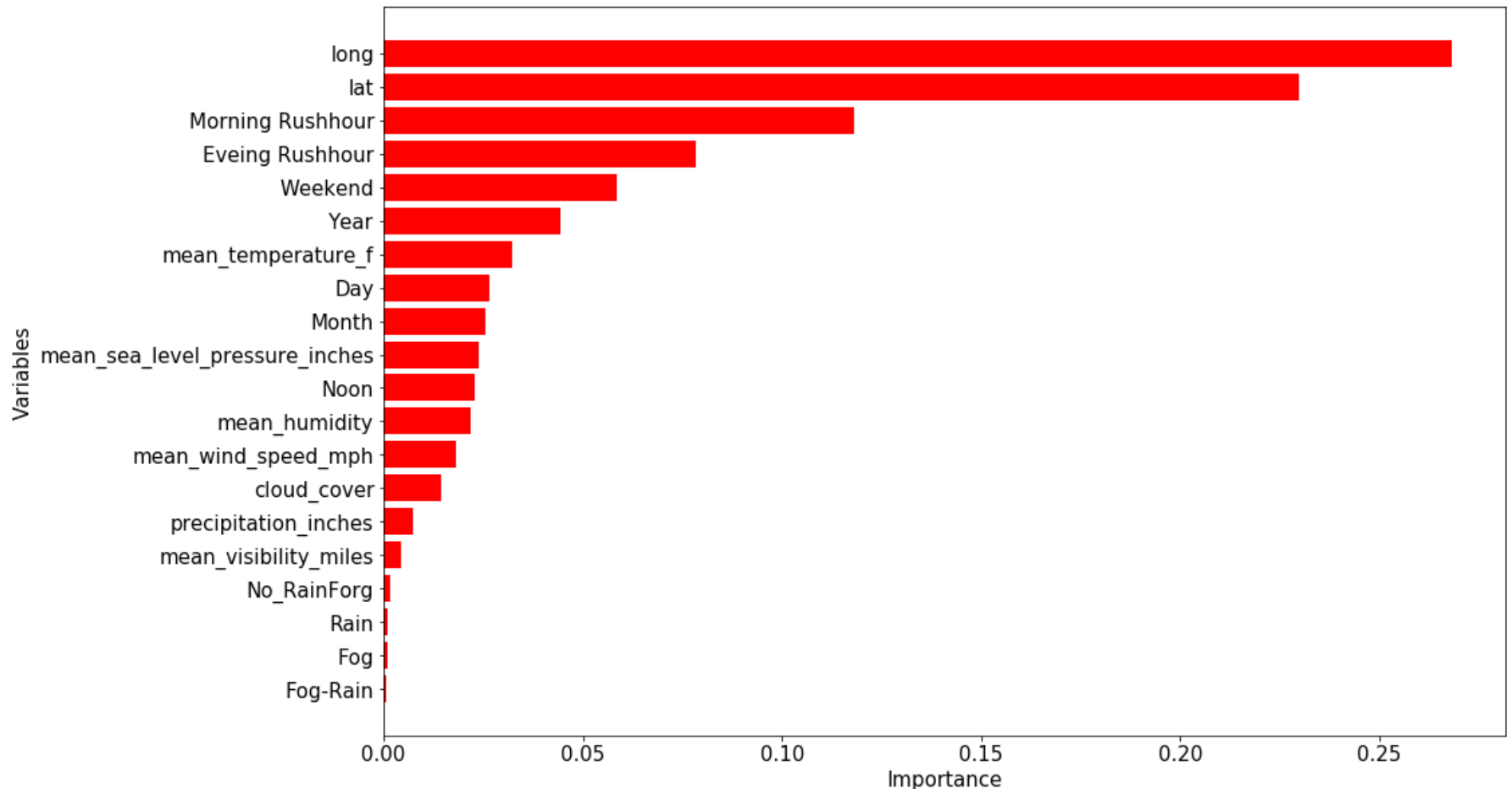
## 時間帯ごとの予測利用数と実利用数〔回〕（San Francisco市, 平日）



**予測結果の確認：**

**参考：変数の重要度が高いのは、位置（ステーションを  
区別）、朝/夜ラッシュ、平日/休日、気温など**

**変数ごとの重要度（ランダムフォレスト）**



**1. 自転車シェアリングビジネスの現状に対する理解**

**2. 需要変動の影響要因分析**

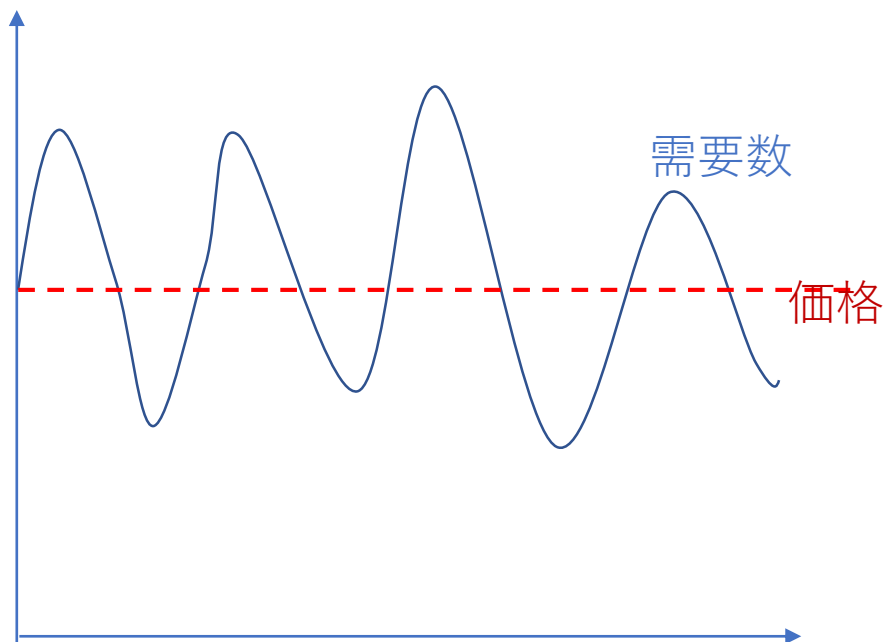
**3. 機械学習による需要予測モデルの構築**

**4. 需要予測に基づく価格変動モデルの提案**

## 変動価格モデルのご提案：

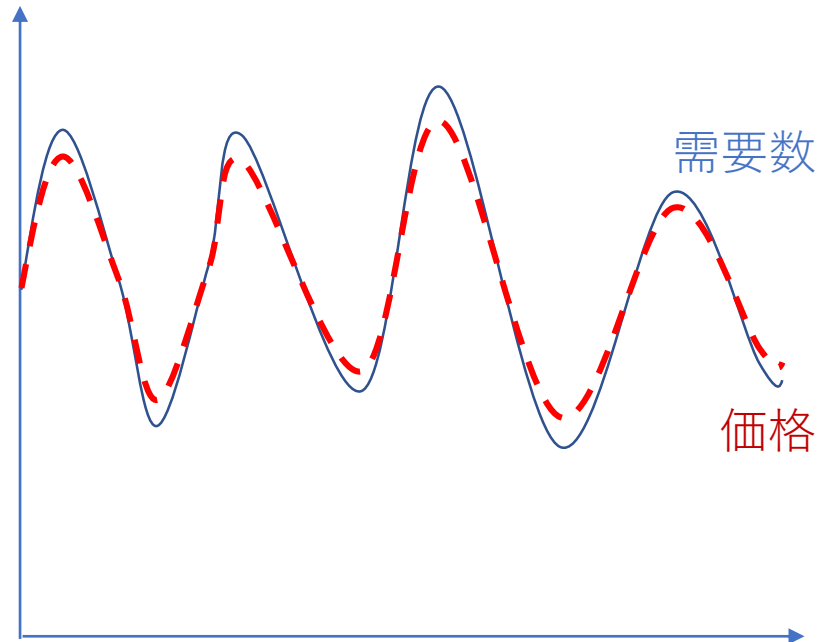
現在の膠着的な価格体系を改め、機械学習で予測された需要数をもとにその日、その時間帯、そのステーションの価格を柔軟に変動させる

現在



変動する需要に対して価格が一定値であり、需要が多い・少ないときの売上の取りこぼしが発生する

新価格モデル



機械学習で予測された需要のもと、ステーション、日付、天気、時間帯の違いに応じ、価格を上下させることで、需給にあった価格体系にする

モデル  
(例)：

$$\text{価格変動率}\%(t) = (1/\text{価格弾性値}) \times \text{需要変動率}\%(t)$$

価格変動率：設定価格/現在の一定価格  
需要変動率：予測需要数/平均需要数

価値の提供方法：

変動価格モデルをスマホアプリに組み込み、お客様に最適価格をアプリ内決済によってリアルタイムで提供

1

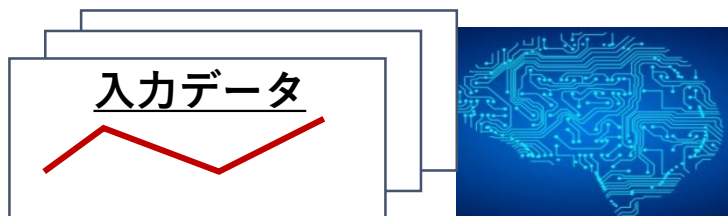
アプリから必要な  
変数データを取得



Dock位置  
日付/曜日  
天気など

2

機械学習によってその条件  
下での需要数を弾き出す

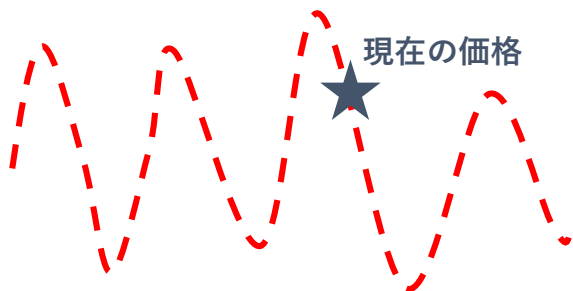


3

予測された需要数を価格変動モデルに当てはめ、その  
時の最適価格を計算

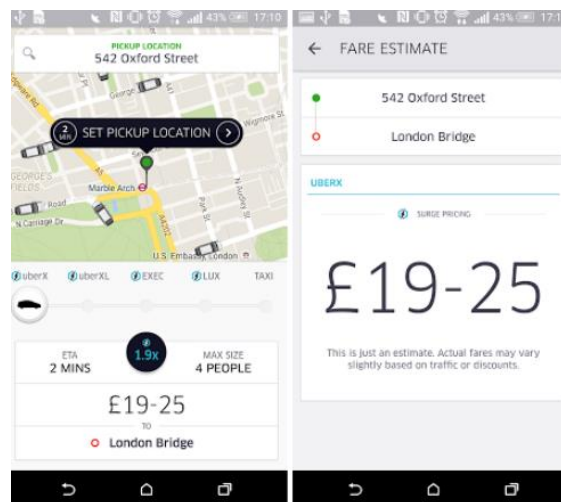
その時の価格を予測

現在の価格



4

最適価格をアプリで即提案  
アプリ内決済による支払い  
を実施



# 需要予測に基づいた価格変動率の計算例

価格弾性値：10  
と仮定した場合

## 時間帯ごとの価格変動率例（San Francisco市, 平日）





# APPENDIX

- 時間があったら（PCスペックが足りていたら）やってみたかったこと
  - ステーションごとの稼働率の予測
    - PCスペック不足により、status.csvの集計、処理に時間がかかりすぎてしまい、時間が足りず（MYSQL, SQLiteで扱っても同じ）。
  - SVM、ニューラルネットワークなどのモデルの適用
    - Fitメソッドの呼び出しに10分かかるため、クロスバリデーション、ハイパーパラメータチューニングをする時間が足りず。
  - 複数モデルのハイパーパラメータチューニング
    - PCスペック不足により時間が足りず。
  - 価格変動モデルの精緻化
    - 思いつく範囲の線形モデルで立ててみたが、実際にはより実情を反映したモデルがあると思われる。