

Bike Sharing Demand

Springboard Capstone Project - I

Project

➤ Bike Sharing Demand (Hourly)

- Bike sharing systems automate obtaining bike riders membership, bike rental and return.
- Data for capstone project is a historical log of rentals for year 2011, 2012.
- Bike share data is aggregated on Hourly basis and combined with Weather data
- This Project uses data from bike sharing systems in Washington D.C., USA.

➤ Outcome from this Project

- Provide insights into demand patterns.
- Predict the demand based on different environmental and seasonal conditions.

Clients

➤ Bike Rental Companies

- Scale bike stations to be able to meet demand
- Increase the demand of bike sharing in the City
- Help convert casual riders to membership/registered riders
- Retain existing registered riders

➤ City Transportation

- Manage city transit services based on bike demand

➤ Society

- Stress free Commute in City
- Healthy Lifestyle

Data Acquisition

- Retrieved dataset from <https://www.kaggle.com/c/bike-sharing-demand>
- The core data set is from <http://capitalbikeshare.com/system-data> aggregated on hourly and daily basis combined with seasonal & weather information
- This data is of Washington D.C area for year 2011 and 2012

Dataset Exploration

- Dataset has 17379 rows with 17 columns
- Record ID is the Index
- Data Size of 2.3 MB

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 17379 entries, 0 to 17378  
Data columns (total 17 columns):  
instant          17379 non-null int64  
dteday           17379 non-null object  
season           17379 non-null int64  
yr               17379 non-null int64  
mnth             17379 non-null int64  
hr               17379 non-null int64  
holiday          17379 non-null int64  
weekday          17379 non-null int64  
workingday       17379 non-null int64  
weathersit        17379 non-null int64  
temp             17379 non-null float64  
atemp            17379 non-null float64  
hum              17379 non-null float64  
windspeed        17379 non-null float64  
casual           17379 non-null int64  
registered       17379 non-null int64  
cnt              17379 non-null int64  
dtypes: float64(4), int64(12), object(1)  
memory usage: 2.3+ MB
```

Data Dictionary

- instant: record index
- dteday : date
- season : season (1:springer, 2:summer, 3:fall, 4:winter)
- yr : year (0: 2011, 1:2012)
- mnth : month (1 to 12)
- hr : hour (0 to 23)
- holiday : weather day is holiday or not (extracted from <http://dchr.dc.gov/page/holiday-schedule>)
- weekday : day of the week
- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
- + weathersit :
 - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
 - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
 - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
 - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp : Normalized temperature in Celsius. The values are divided to 41 (max)
- atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
- hum: Normalized humidity. The values are divided to 100 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)
- casual: count of casual users
- registered: count of registered users
- cnt: count of total rental bikes including both casual and registered

Data Wrangling

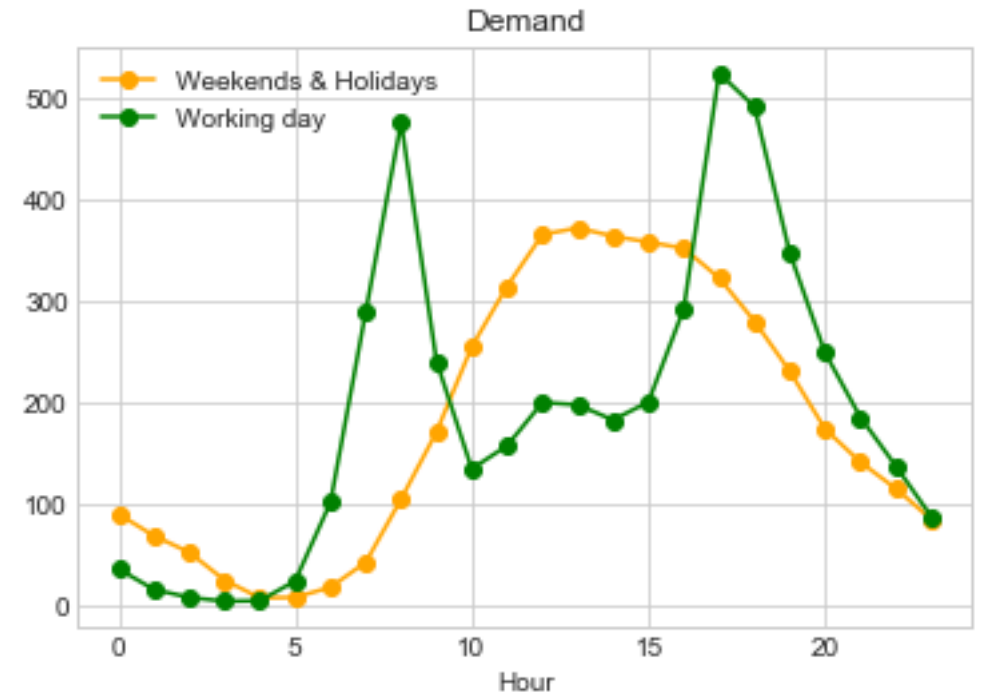
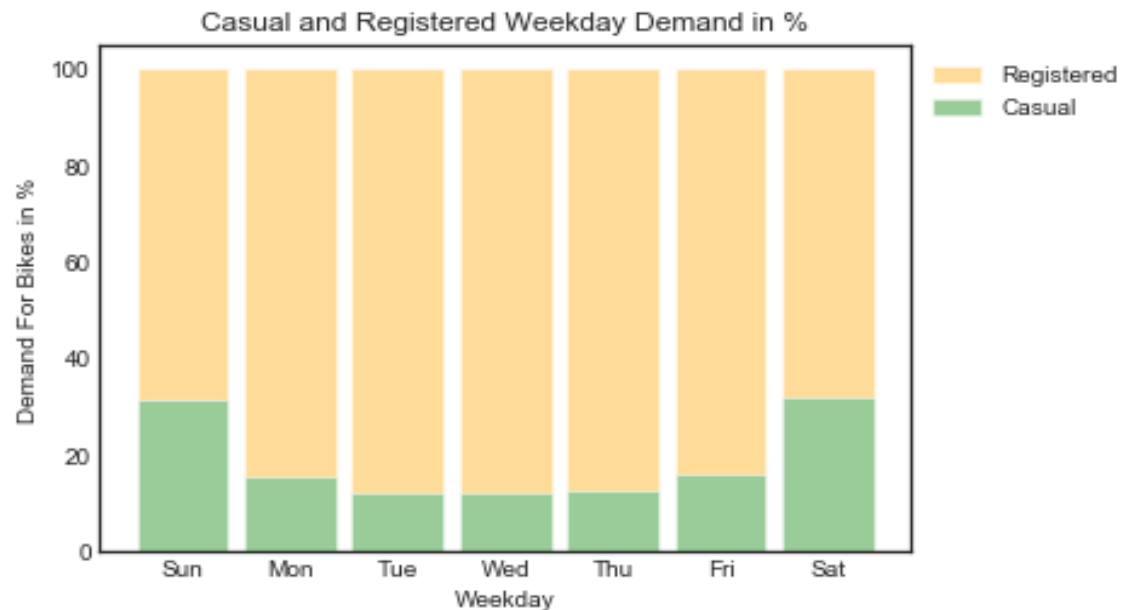
- Rename column names
- Combine **Date** and **Hour** columns as Index
- Delete **instant** column
- Create **Part_Of_Day** column
(divides a day in 6 windows/periods)
 - 0-4 : 0
 - 4-8 : 1
 - 8-12 : 2
 - 12-16 : 3
 - 16-20 : 4
 - 20-23 : 5

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17379 entries, 0 to 17378
Data columns (total 18 columns):
Instant                17379 non-null int64
Date                  17379 non-null object
Season                17379 non-null int64
Year                  17379 non-null int64
Month                 17379 non-null int64
Hour                  17379 non-null int64
Holiday               17379 non-null int64
Weekday               17379 non-null int64
Workingday            17379 non-null int64
Weather_Condition     17379 non-null int64
Normalized_Temperature 17379 non-null float64
Normalized_Feels_Temperture 17379 non-null float64
Humidity              17379 non-null float64
Windspeed             17379 non-null float64
Casual                17379 non-null int64
Registered            17379 non-null int64
Demand               17379 non-null int64
Part_Of_Day           17379 non-null category
dtypes: category(1), float64(4), int64(12), object(1)
memory usage: 2.3+ MB
```

Weekdays Data Analysis

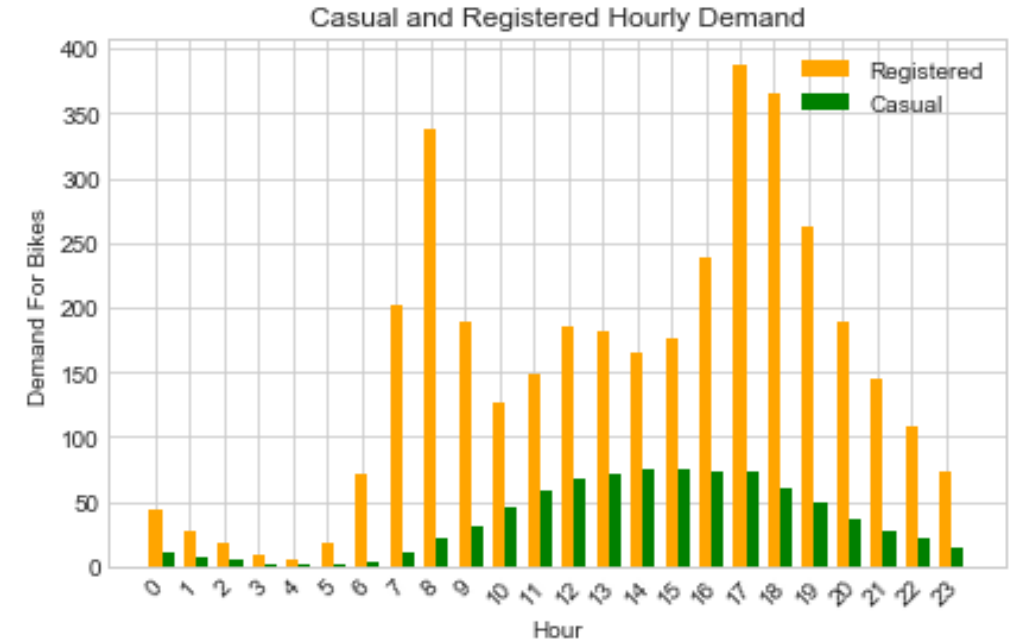
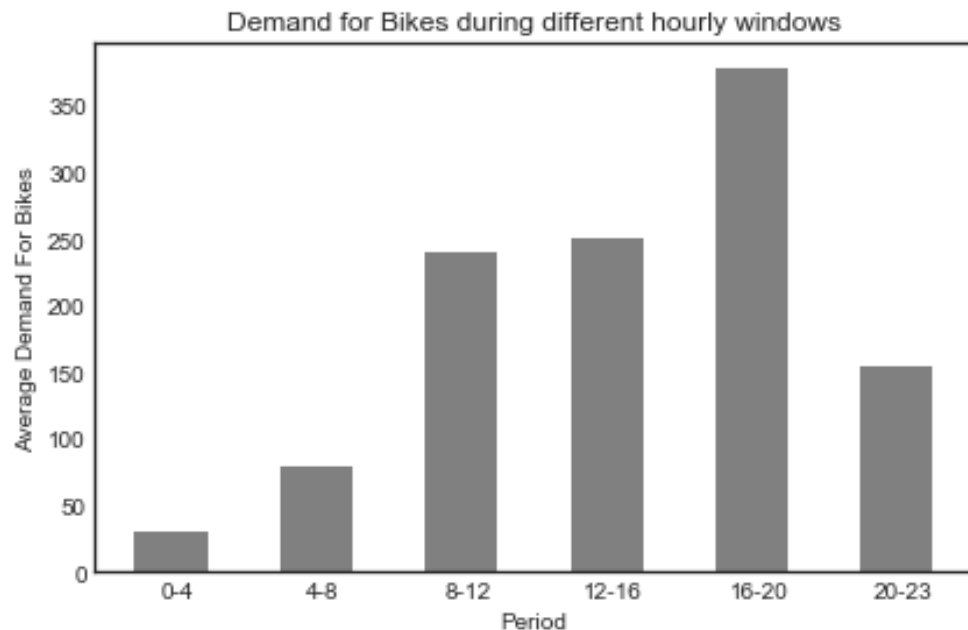
- Weekends & Holidays
 - Demand peaks in afternoon
- Workday
 - Demand peaks at Morning and Evening Hours



- Registered riders contribute higher percentage demand

Hourly Data Analysis

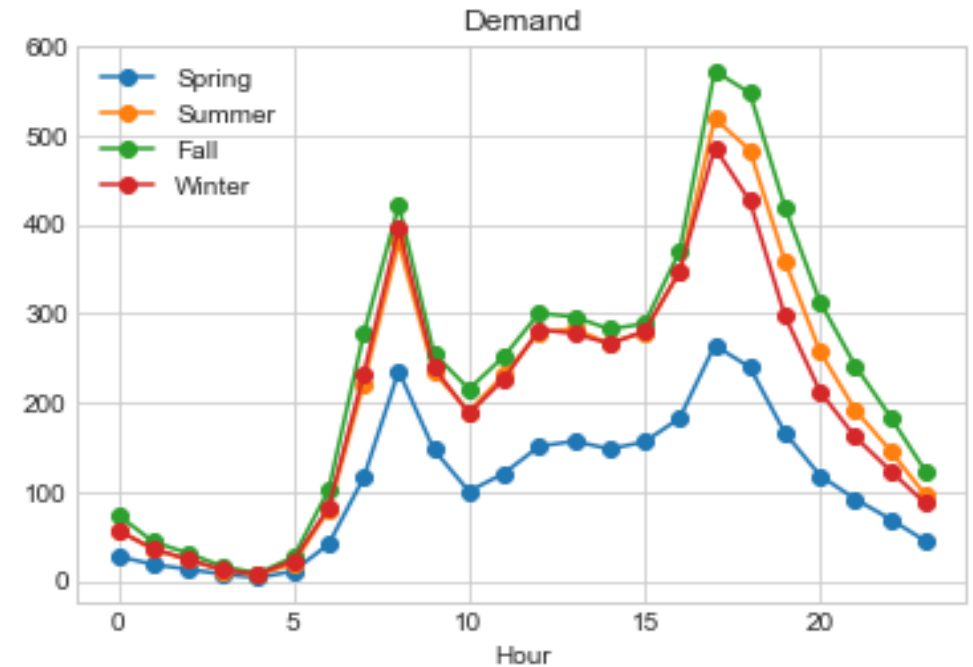
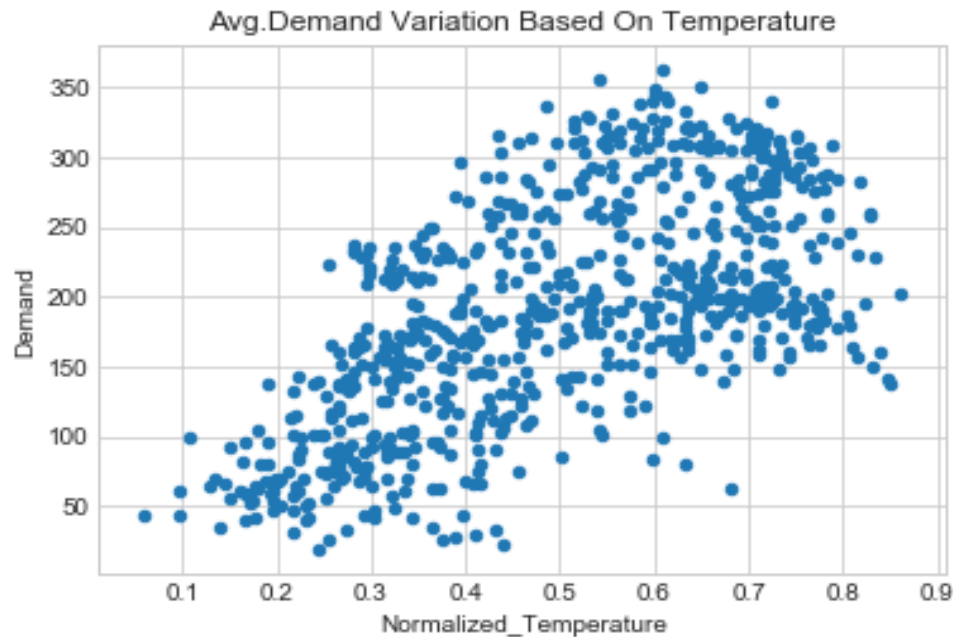
- Evening Hours have higher demand than Mornings



- Demand Peaks between 4 and 8 PM

Seasonal Data Analysis

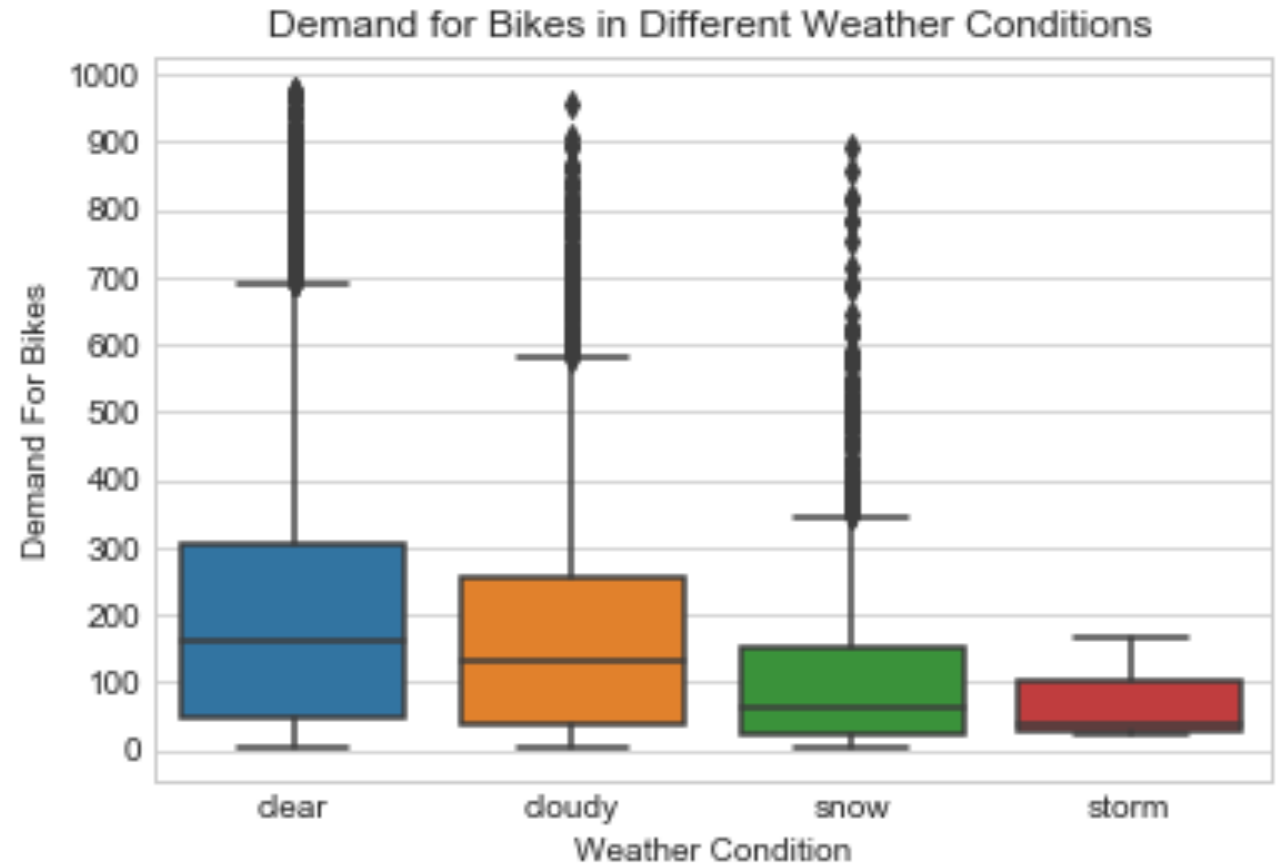
- Demand is highest in Fall Season and lowest in Spring Season



- Demand increases with temperature and it starts declining at very high temperature.

Weather Data Analysis

- Average Demand is highest in Clear Weather
- Demand is lowest in Stormy Weather



Model Evaluation & Selection

- Evaluated below 3 Linear Regression Models

	Linear Regression	Ridge Regression	Lasso Regression
R^2	0.40	0.40	0.38
RMSE	139.97	140.23	139.85

- Its evident from above metrics, that linear regression is not the best model for this project

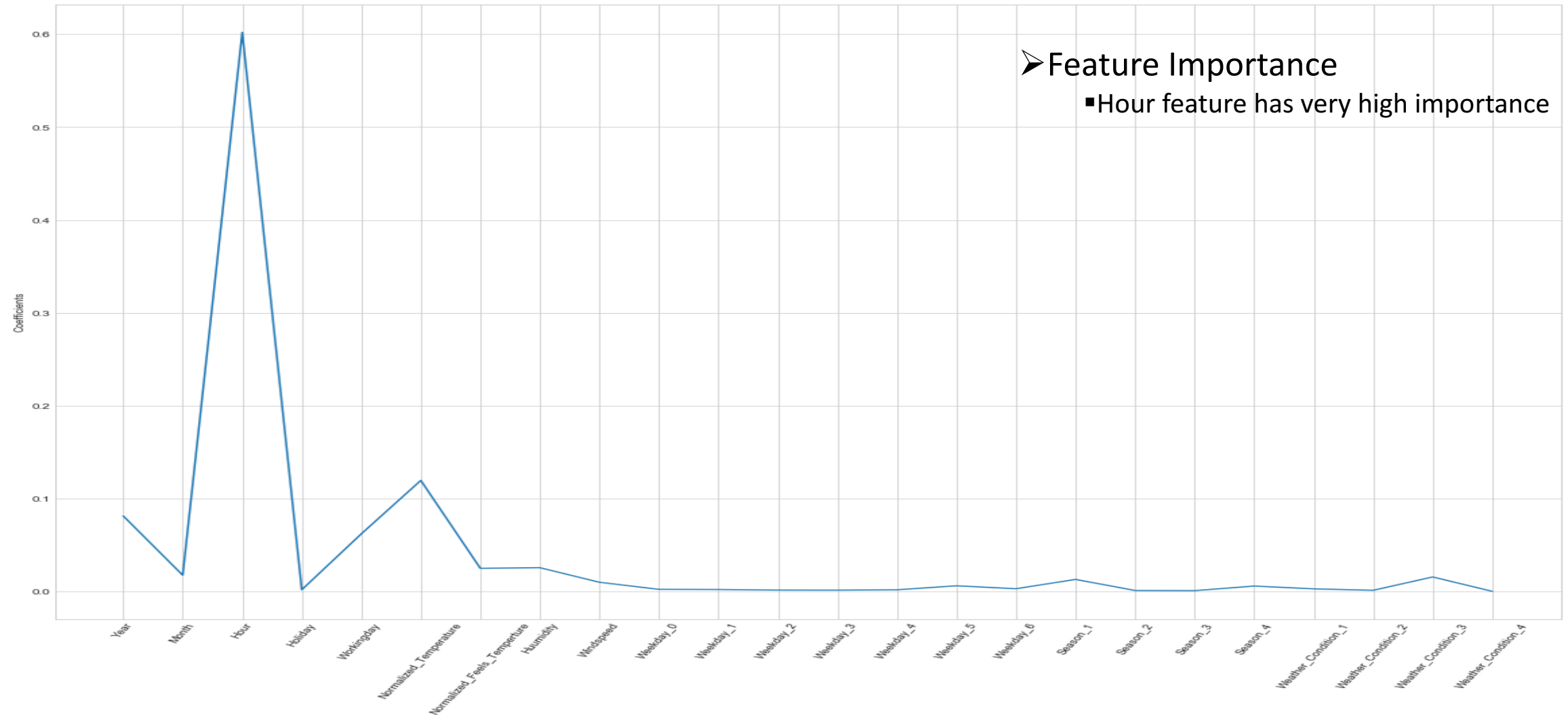
Model Evaluation & Selection

- Random Forest with One Hot Encoding
- Determine important parameters to build a final model with RandomizedSearchCV
 - max_depth=25
 - min_samples_leaf= 1
 - min_samples_split= 2
 - n_estimators = 100

	Decision Tree Regression	Random Forest Regression
R^2	0.90	0.94 ←
RMSE	57.12	44.48

- Random Forest regression with R^2 score of 0.94 is a better fit for this dataset

Model Evaluation & Selection



Recommendations to Clients

- Factors Influencing increase in demand for bike sharing in city
 - Registered Ridership
 - Working Day
 - Evening Commute Hours (4 and 8 PM)
 - Clear Weather Conditions
 - Season
 - Moderate Temperatures
- Recommendations to Bike Rental Companies
 - Convert Casual Riders to Registered
 - Adapt bike stations to meet evening demand
- City Transportation
 - Increase services during bad weather conditions as bike rental demand decreases

Thank You.