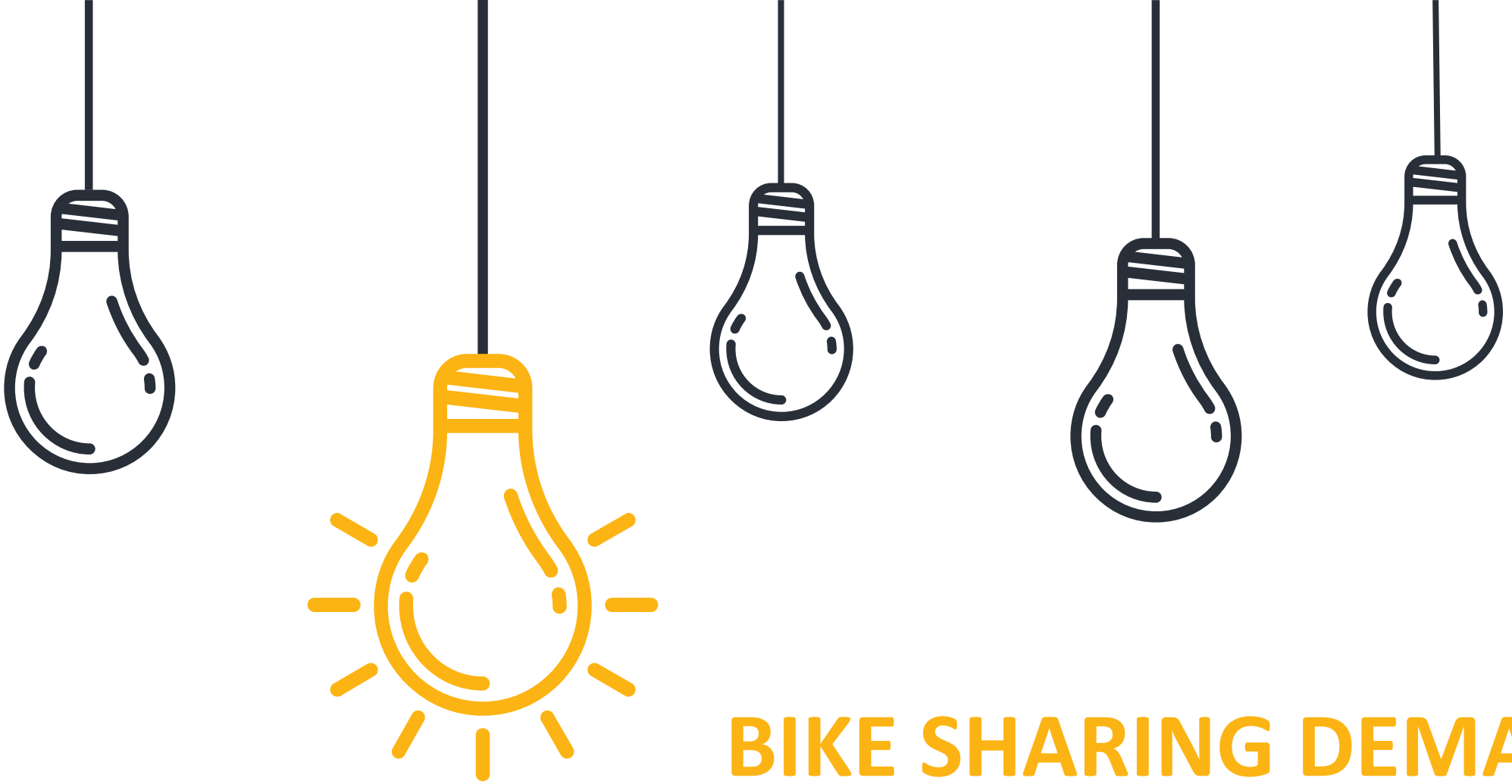


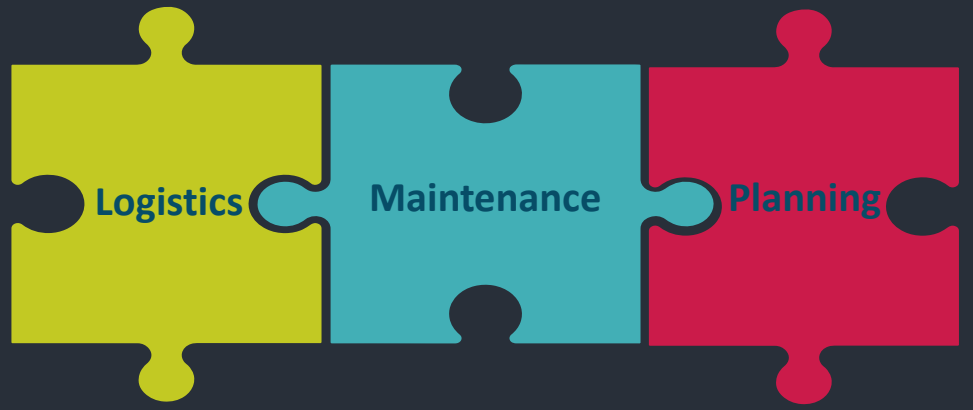
BIKE SHARING DEMAND

VICTOR FONSECA

OCTOBER 2019

The challenge

How many bikes are demanded for rental in San Francisco every hour?



Ford 2 Go - Lyft

- Bike sharing service working in San Francisco Bay area: bikes available 24/7/365 in San Francisco, Oakland, Berkeley, Emeryville and San Jose (CA)
- More than **300 stations** and more than **6,000 bikes** available for rental
- Affordable prices start at $2/30 minutes ride, or $5/year (membership)
- Recently acquired by Lyft and rebranded to **Bay Wheels**

Available data

FordGoBike 2017-2019

Cleaned and merged bike sharing transactions from July-17 to April-2019

- Start and end station
- Date and time
- 6M transactions
- 16k stations-hour

World Weather Online API

Weather Data and API for Businesses and Developers

- Weather conditions history for San Francisco between 2017 and 2019
- Hourly data: 16k observations

Project walkthrough

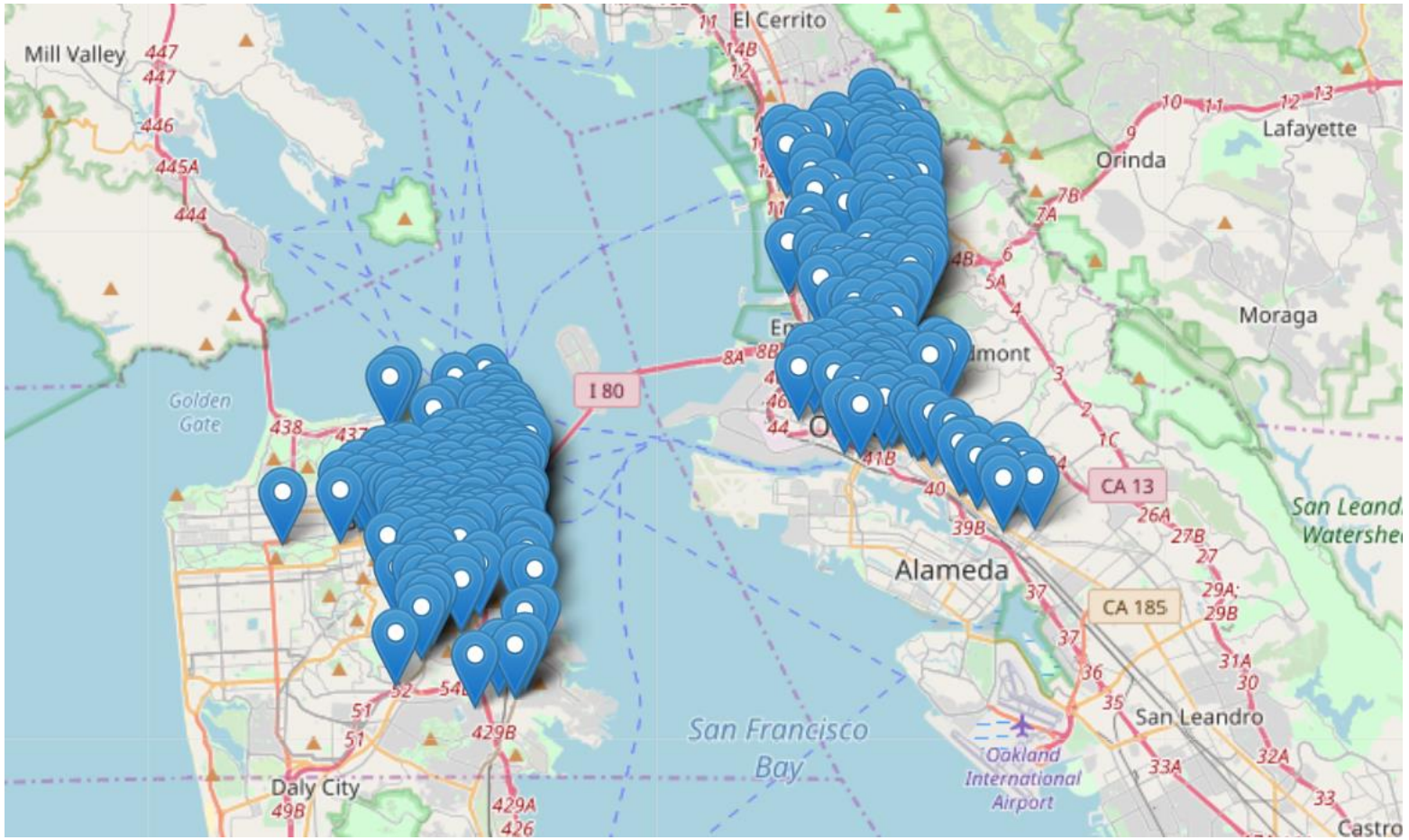
Data collection
- Download the Kaggle dataset
- Build the API to collect weather information
  - **Requests**
  - **Beautiful Soup**
  - **Time**

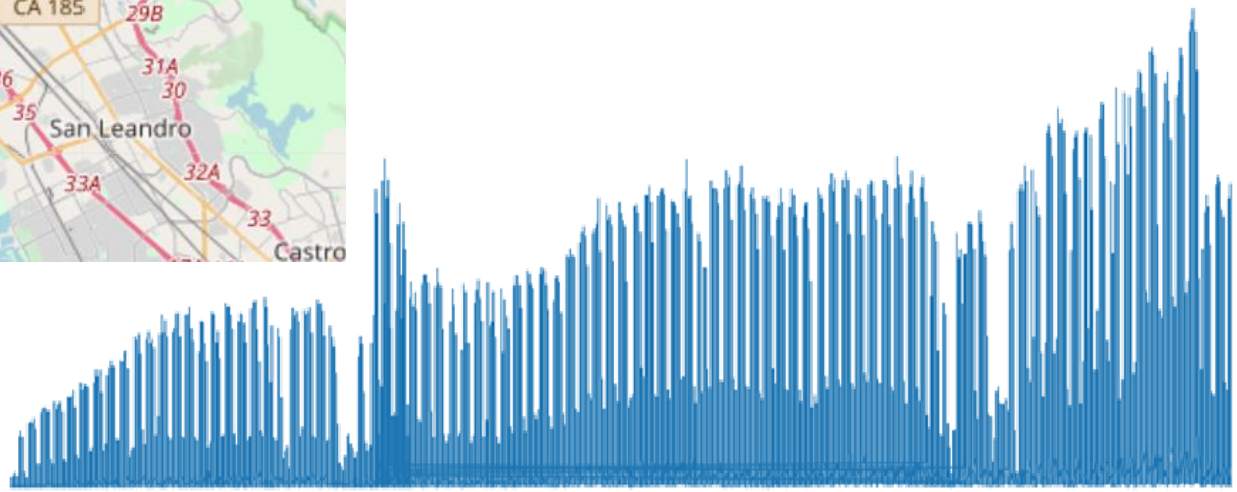
Exploratory analysis
- Charts to describe the predictors and raise hypothesis
  - **Matplotlib**
  - **Seaborn**
  - **Folium**

Data Preparation
- Aggregate transactional data by hour
- Merge weather conditions (temperature, weather description, rain)
- Create additional variables (holidays, season, month, day, working day, weekday, hour, city, log)
  - **Pandas**
  - **Numpy**

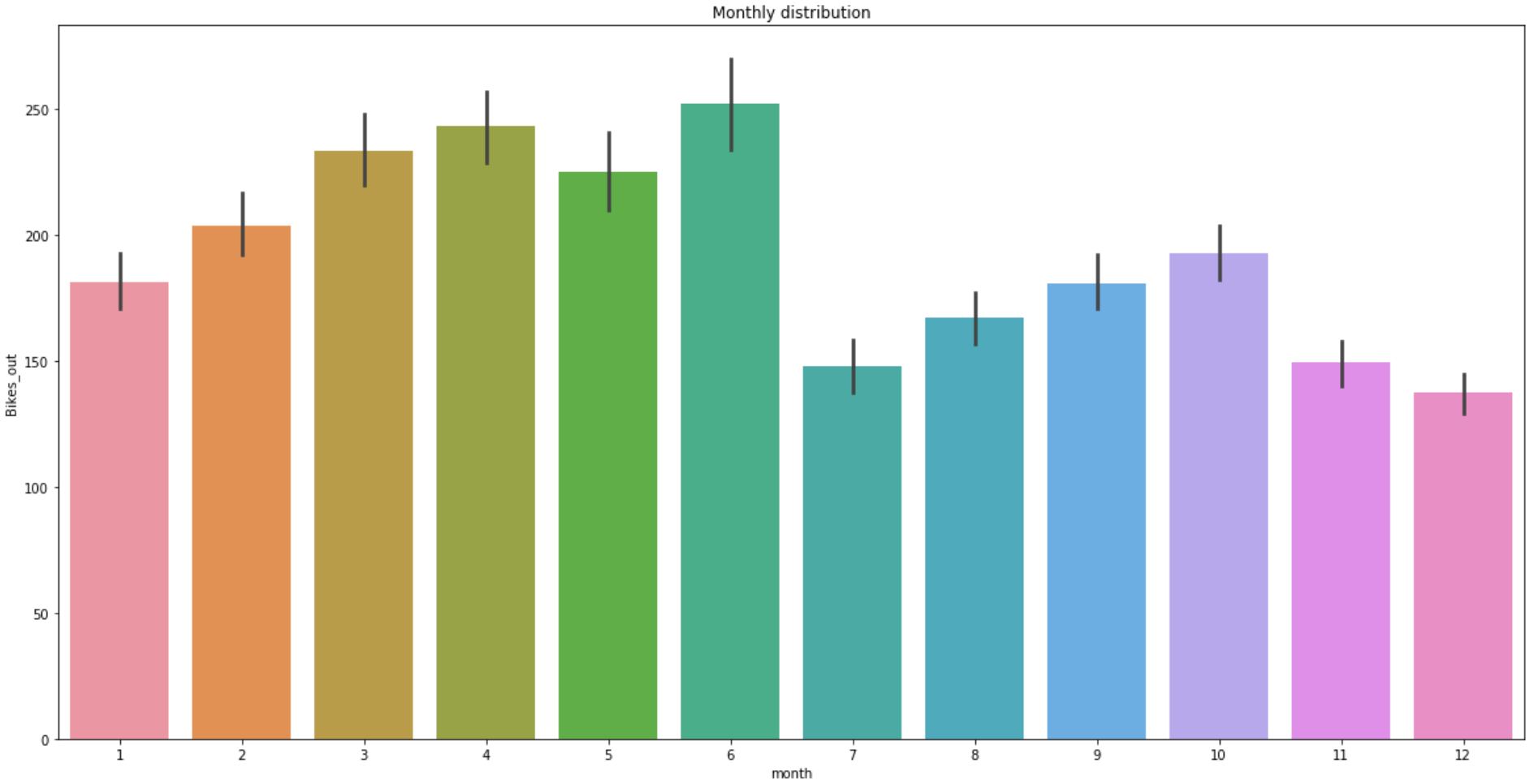
Modelling
- Linear Regression
- Decision Trees
  - **Scikit-learn**

Ford GoBike System

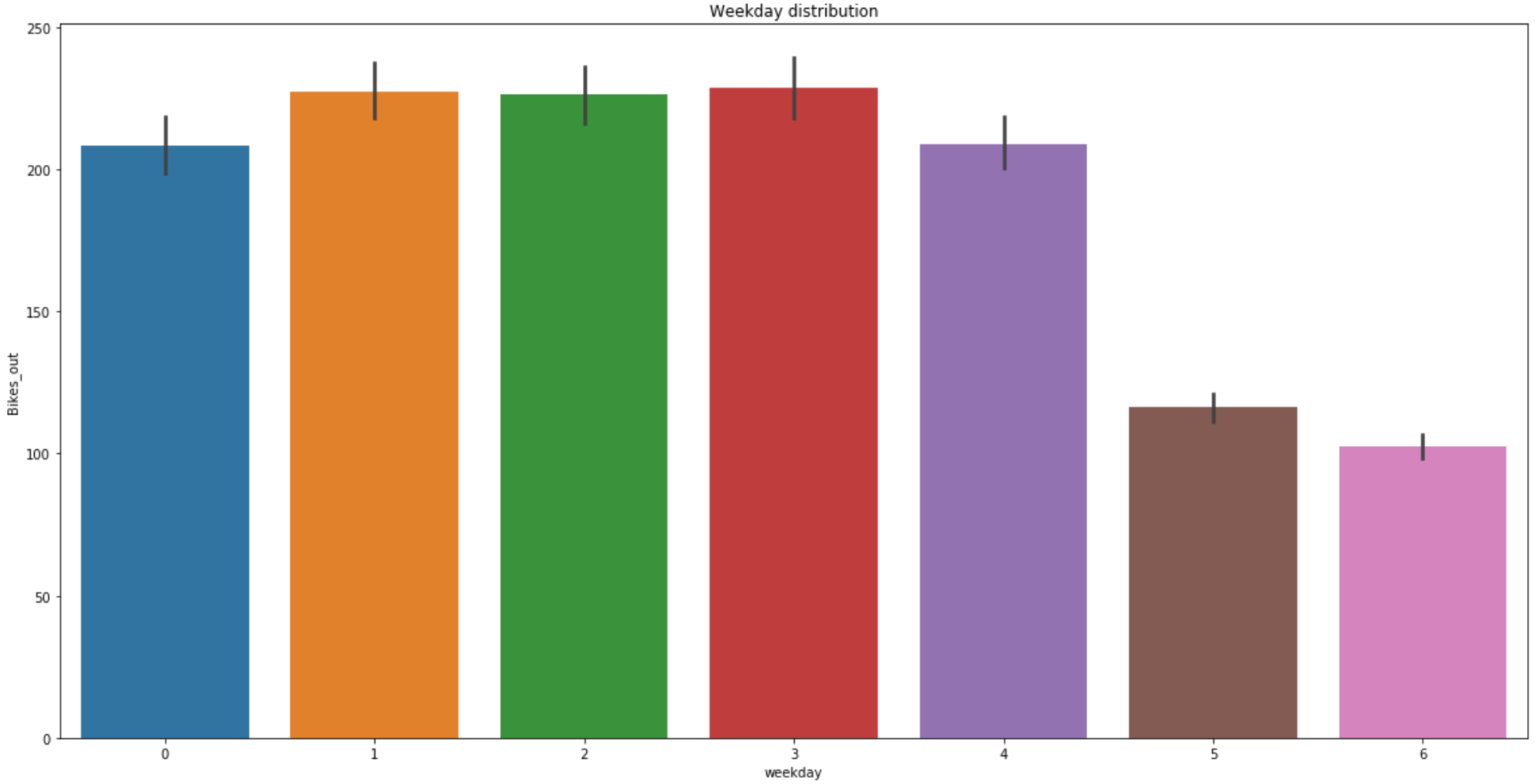




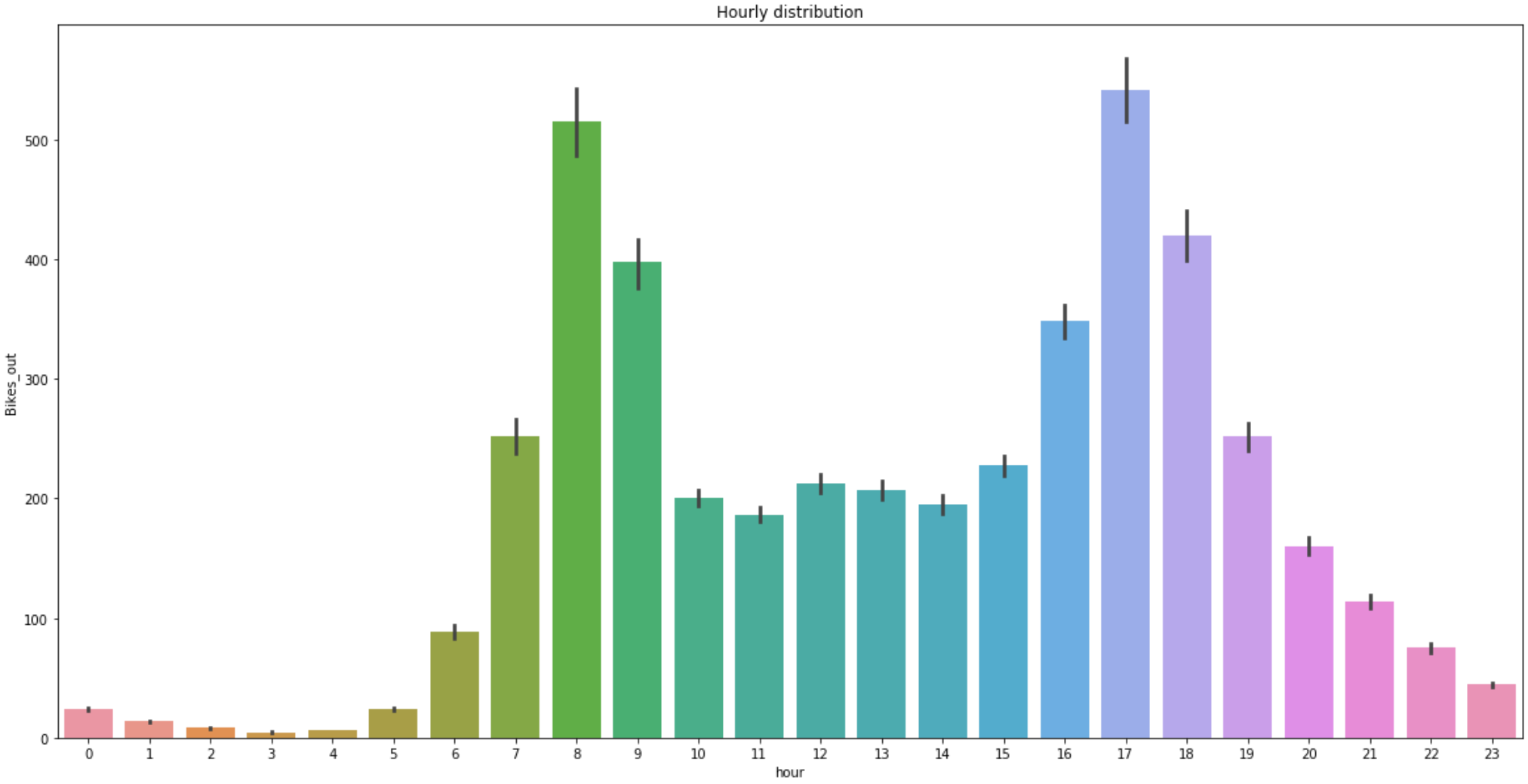
Rentals by month



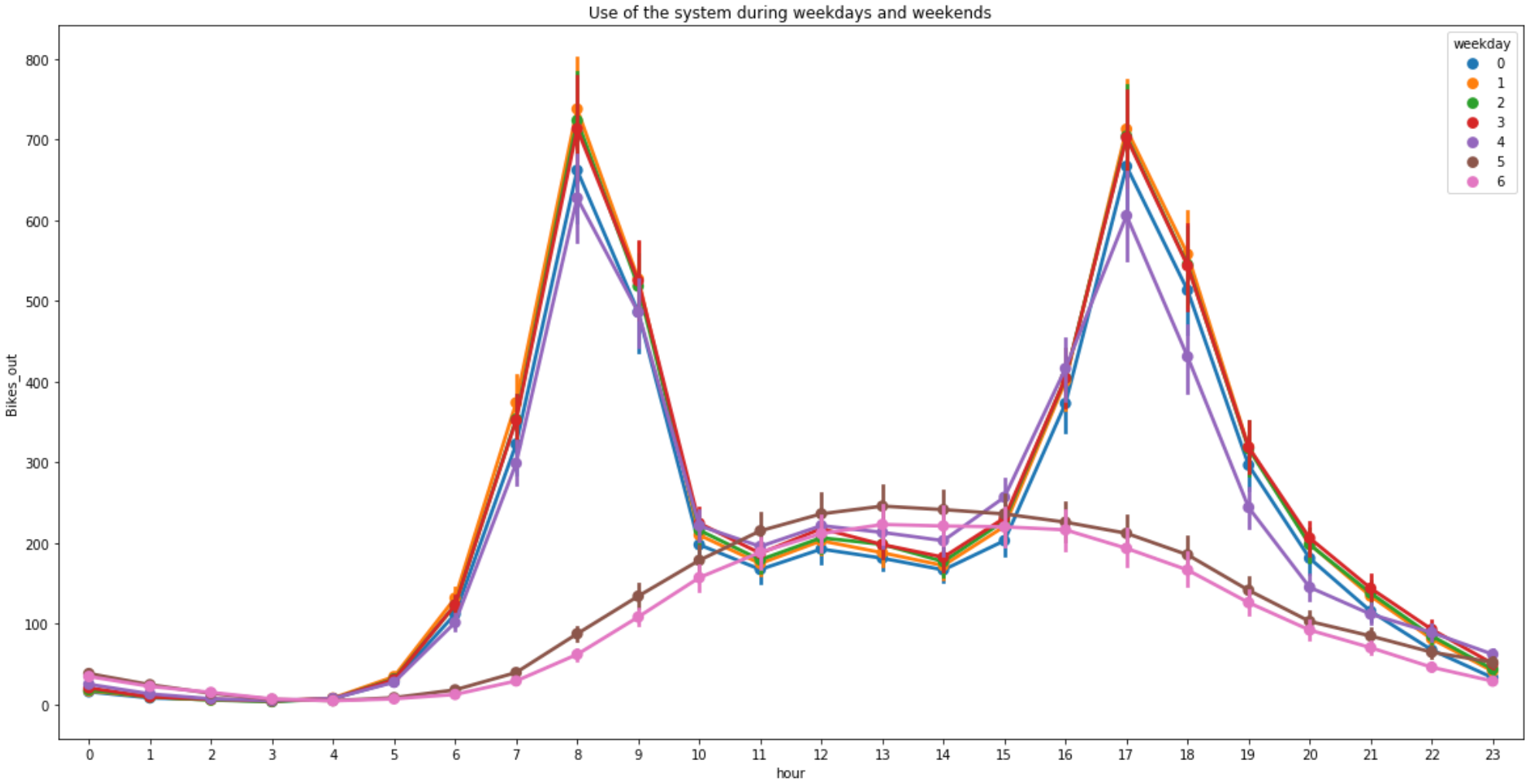
Rentals by weekday



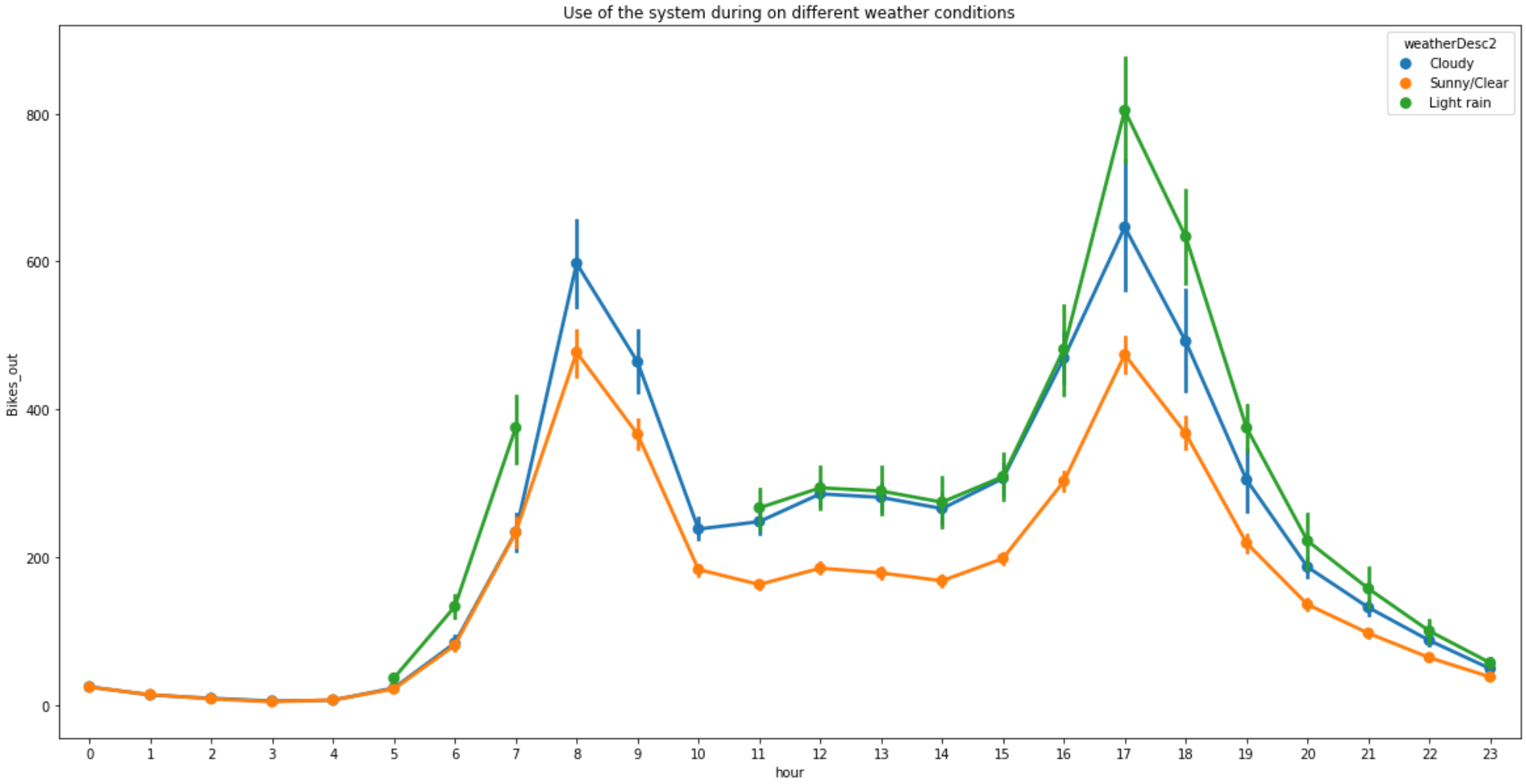
Rentals by hour



Rentals by weekday and hour

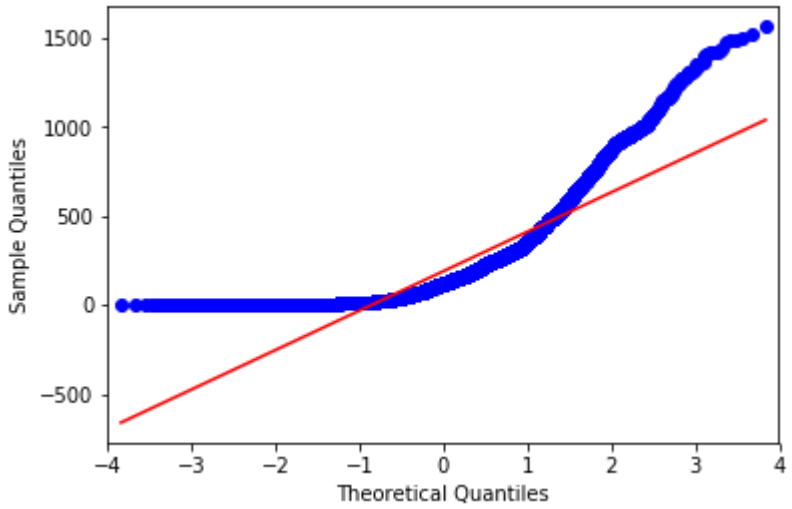


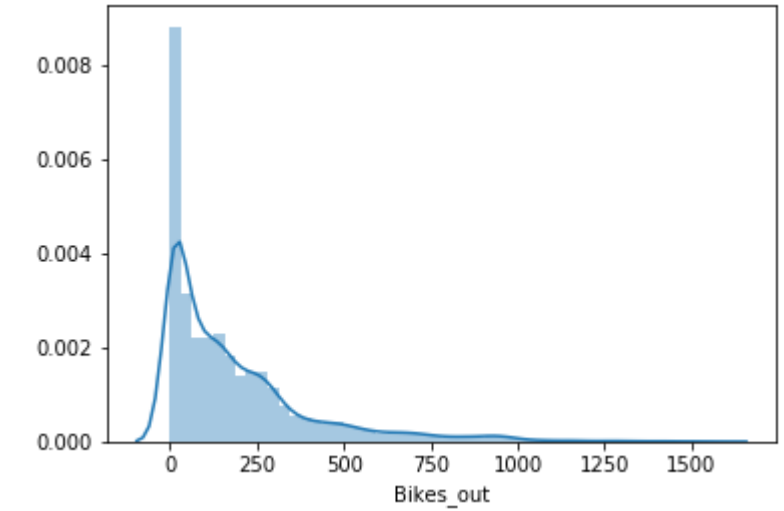
Rentals by weather condition

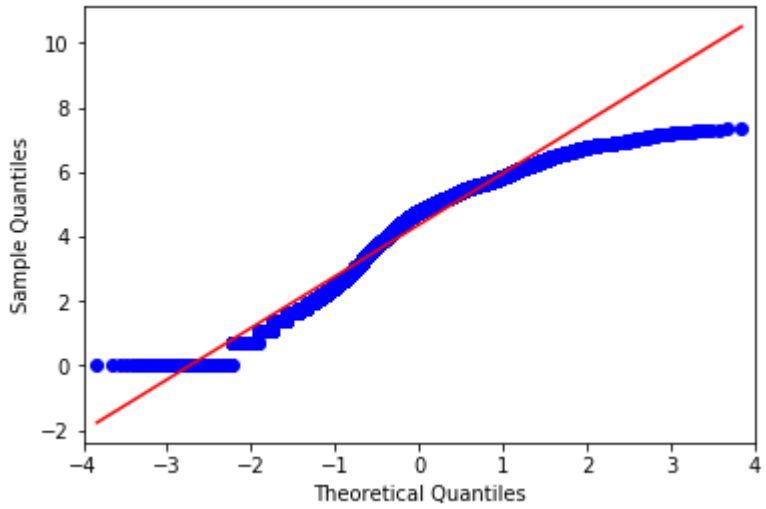


Linear Regression

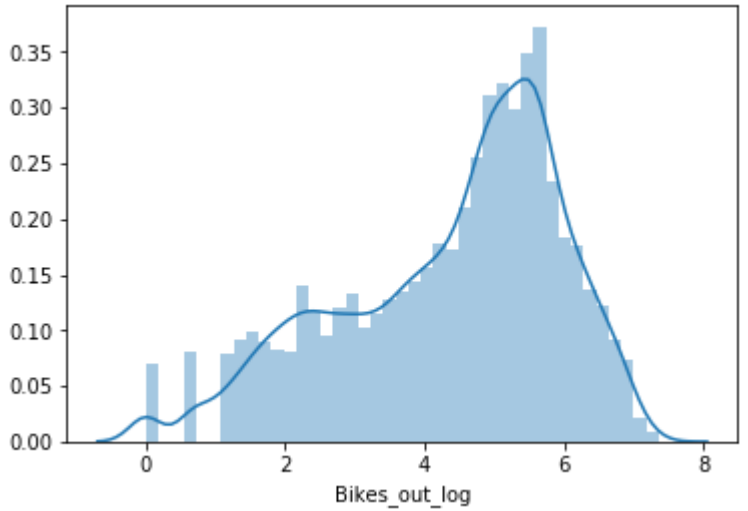
Original data





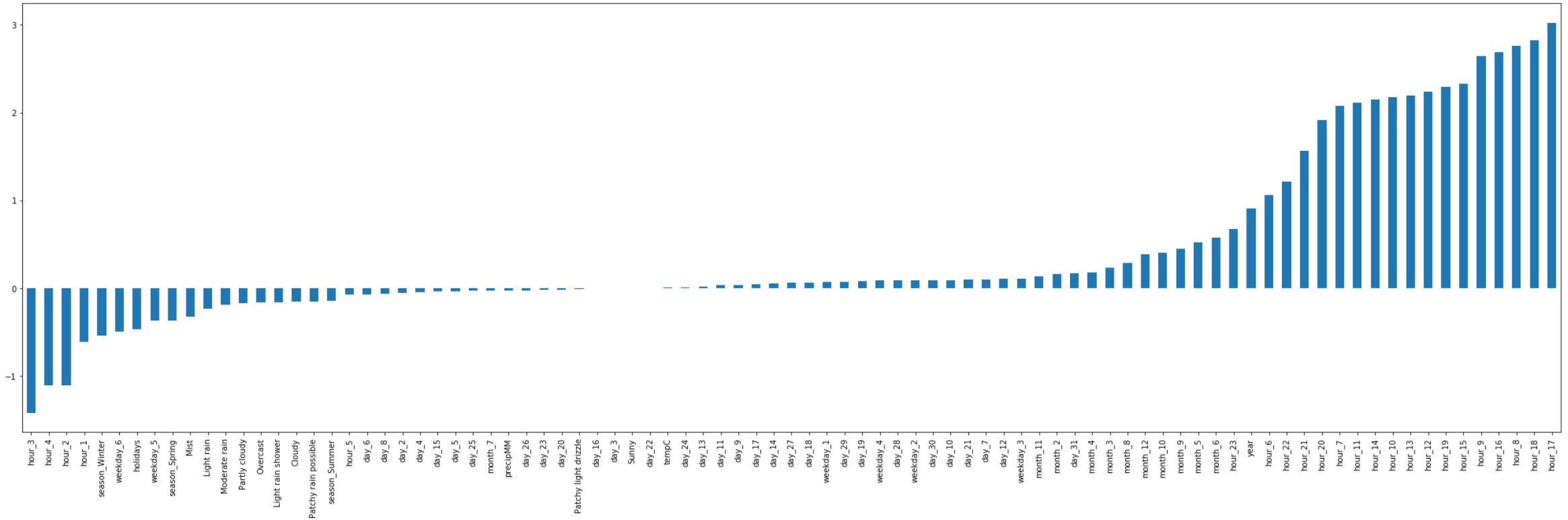


Log transformation



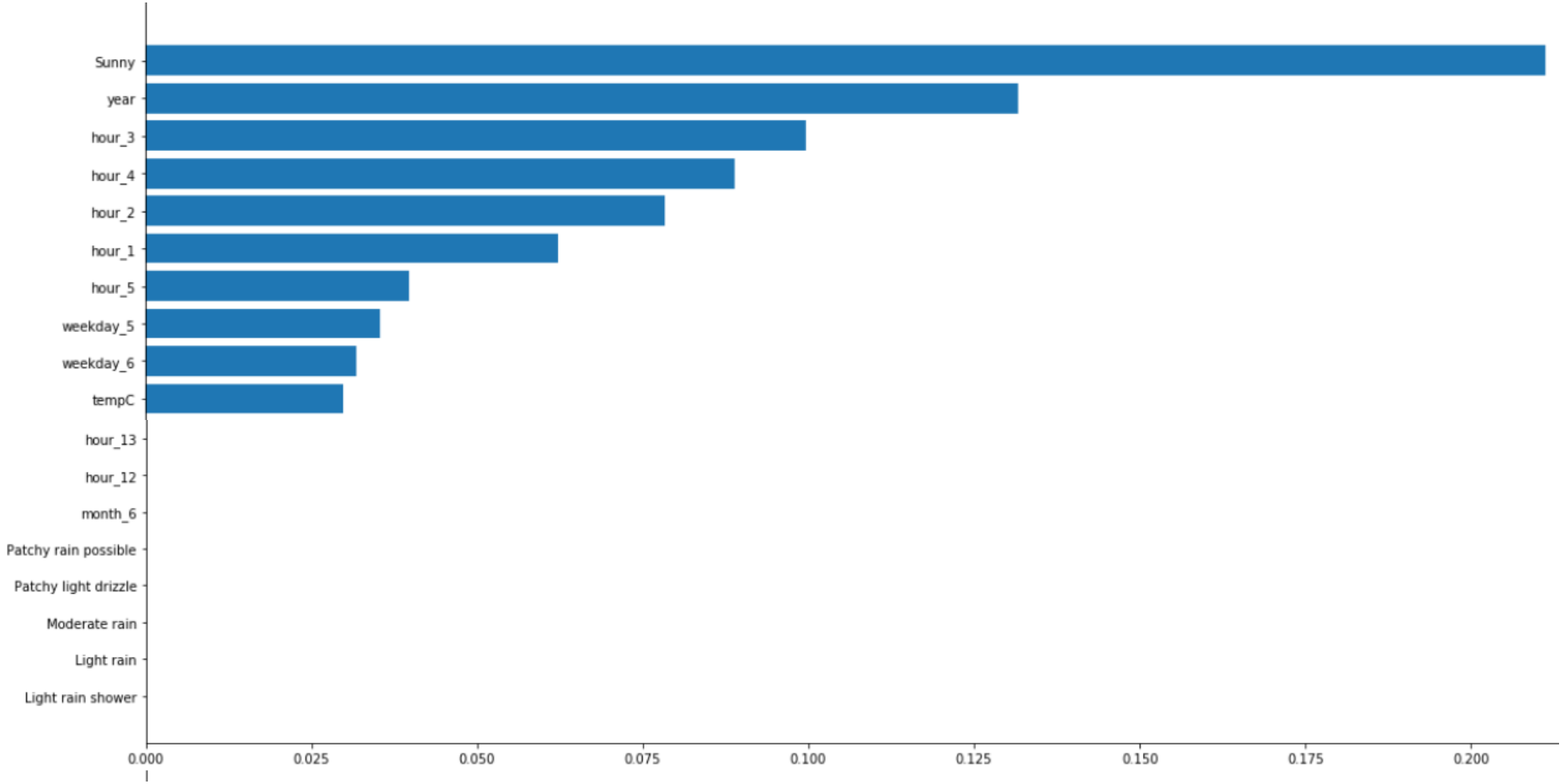
Linear Regression

R2 - Train	R2 - Test	MSE
0.8707	0.8667	0.3417



Decision Tree

R2 - Train	R2 - Test	MSE
1.0000	0.9095	0.2320



Recommendations

- Investigate the models – both seem to be good for predictions
- Incorporate new information to the model (school calendar, weather specifics, membership, etc)
- Stations demand forecast model
- Stochastic models