

Recruit Restaurant Visitor  
Forecasting  
Project 2, Semester 2, 2020

Sai Vivek Kammari  
a1807677

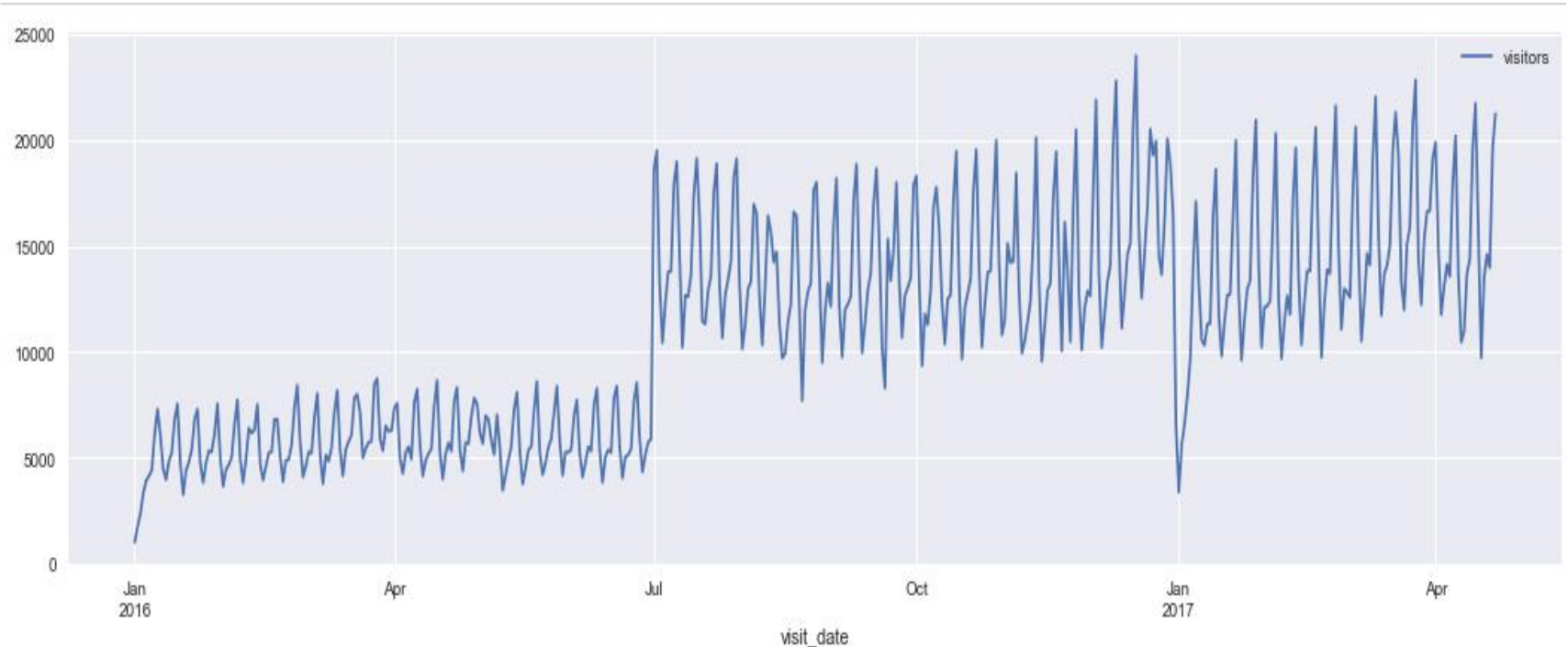
# Introduction

- In order to avoid food wastages and to attain better profitability by managing the stock and staff efficiently. Restaurants need to anticipate the number of visitors.
- In this project we try to predict the expected visitors for a give set of Japanese restaurants based on the historic visitors data

# Dataset

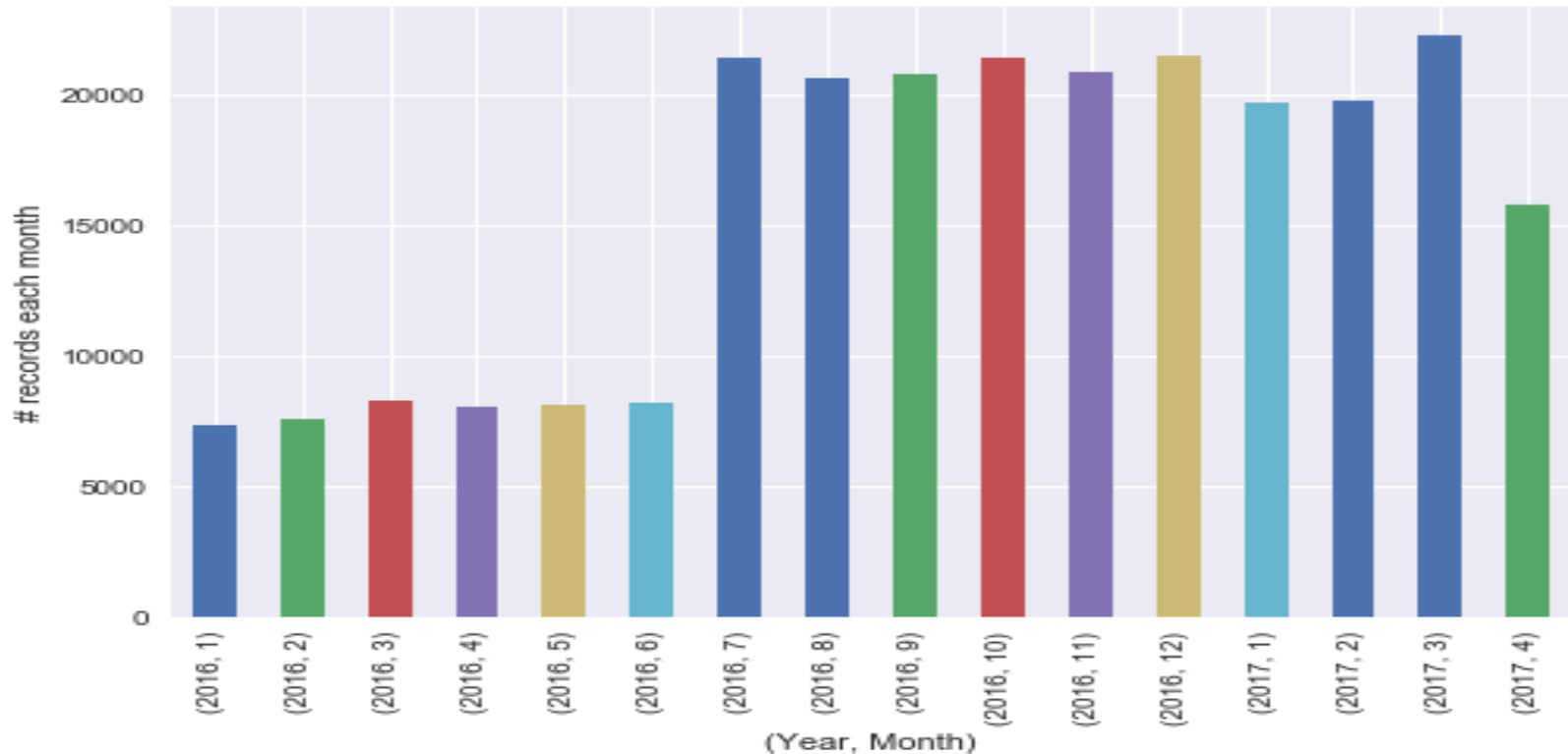
- The datasets are extracts from two websites
  - Hot Pepper Gourmet
  - AirREGI / Restaurant Board
- The datasets 'air\_reserve.csv', 'air\_store.csv' and 'air\_vist\_data.csv' are extracted from the AirREGI
- The 'hpg\_reserve.csv' and 'hpg\_store\_info.csv' are the extracts from the website Hot Pepper Gourmet.

# Exploratory Data Analysis



Aggregate to all the restaurants visitors on each day of the historic period. The gap in the figure is the visitor numbers that are to be predicted (from last week of April 2017 to the end of May 2017).

# Exploratory Data Analysis



Summarizes the total number of visitors to all the restaurants on a monthly basis. From the above figure it is evident that the number of visitors to the restaurants has grown significantly from the month of July 2016.

# Feature Engineering

- Developing a Final Dataset by combining all the given datasets.
- Taking special care of the Holidays.
- Adding Time Series Features
- Encoding the Categorical Features.

# Machine Learning Models

- XGBoost Regression  
Parameter Tuning- Bayesian Optimization
- Random Forest Regression  
Parameter Tuning- Grid Search Optimization
- KNN Regression

# Evaluation

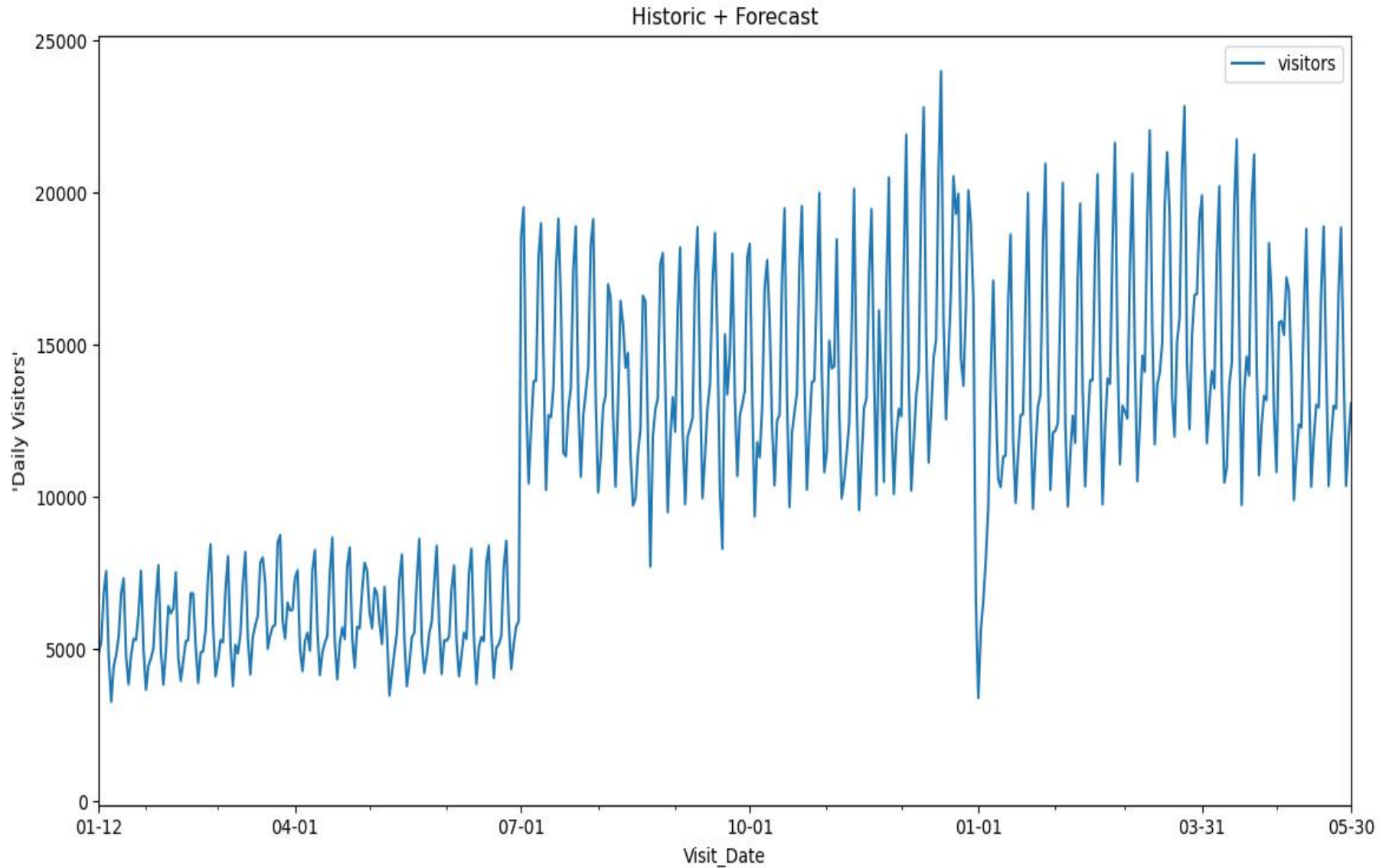
RMSLE-  $\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(x_i+1) - \log(y_i+1))^2}$

REGRESSION MODEL	RMSLE
XGBOOST	0.489975
RANDOM FOREST	0.492949
KNN	0.608833

- Rmsle value for the XGBoost model is lowest and hence XGBoost is selected for the final implementation.



# Final Implementation



# Future Work

- Weather also plays a key role in the business of restaurants and in this whole experiment I have not analyzed the impact of weather. In future, we can use the external weather data for the entire time period and add this to the final dataset. I expect this implementation will definitely give us more precise forecast.
- We can use better tuning techniques for Random Forest, tuning the Random Forest parameters using the random search and then narrow down the grid of value. After narrowing down to the grid levels, we can then perform grid search cv to finalize the parameters. This procedure would also give better predictions from Random Forest model.
- We can implement ensemble techniques for improving the predictions. I plan to improve my model over the future after gaining significant knowledge in this direction.

THANK YOU!