

Restaurant Visitor Forecasting

Capstone Project
Nishal Pattan

Overview

- Business Problem
- Data Wrangling
- Exploratory Data Analysis
- Inferential Statistics
- Forecast Modelling
- Conclusion / Future Work

Business Problem

The Client here is Recruit Holdings who wants to automate the process of number of visitors a restaurant can expect in future date.

It depends on various factors like location of restaurants, holiday information, historical reservations, past visitors data on a given date.

Clients : Restaurants / Hotel management team

How it is useful :

- Better Inventory management
- Effective Cost Management in purchasing ingredients, Scheduling staff by knowing by the number of visitors
- What's the most favorite genre of food that people are eating. Knowing the taste of customers much better

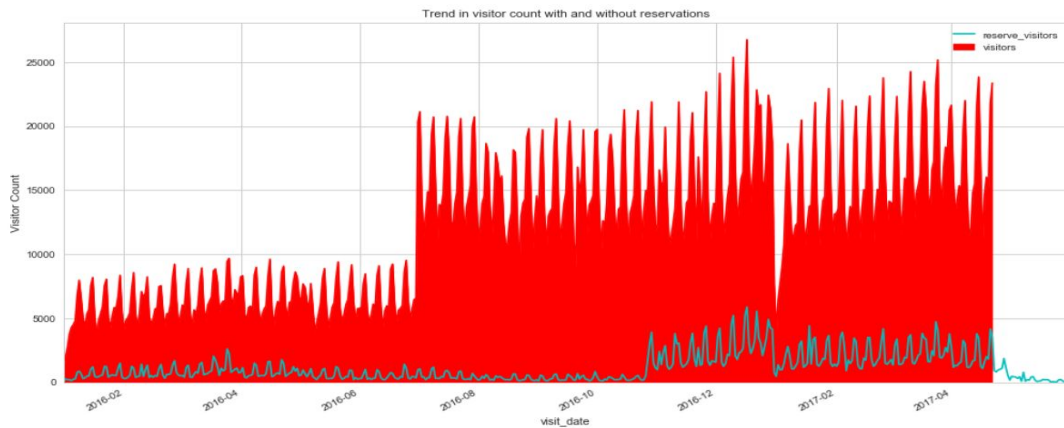
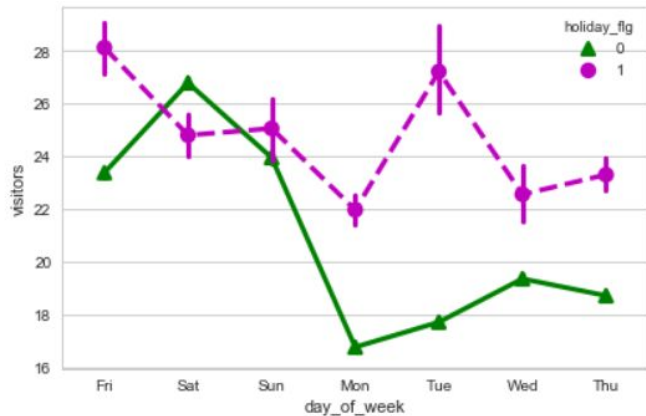
Data Wrangling / Cleaning

- Performed Inner join between Air & Hpg reservation tables to get intersection of the data
- Changed calendar dates in data from string to datetime format.
- Mapped hpg stores with their equivalent air store ids as the final test set given is having air store ids for forecasting
- After cleaning data, train and test data are stored separately as Train_data.csv, Test_data.csv files
- NaN's are filled with -1' sto treat them as outliers.
- Training data has only data till April 2017, and Test data has May to June 2017

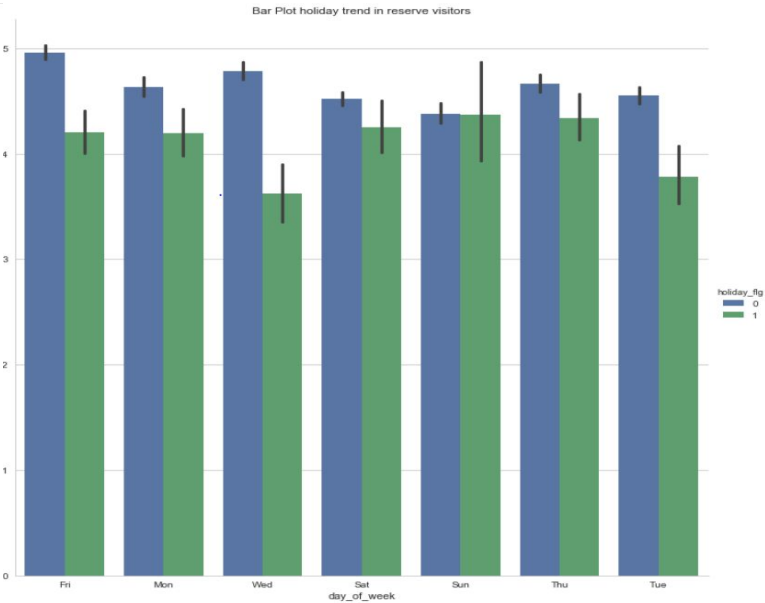
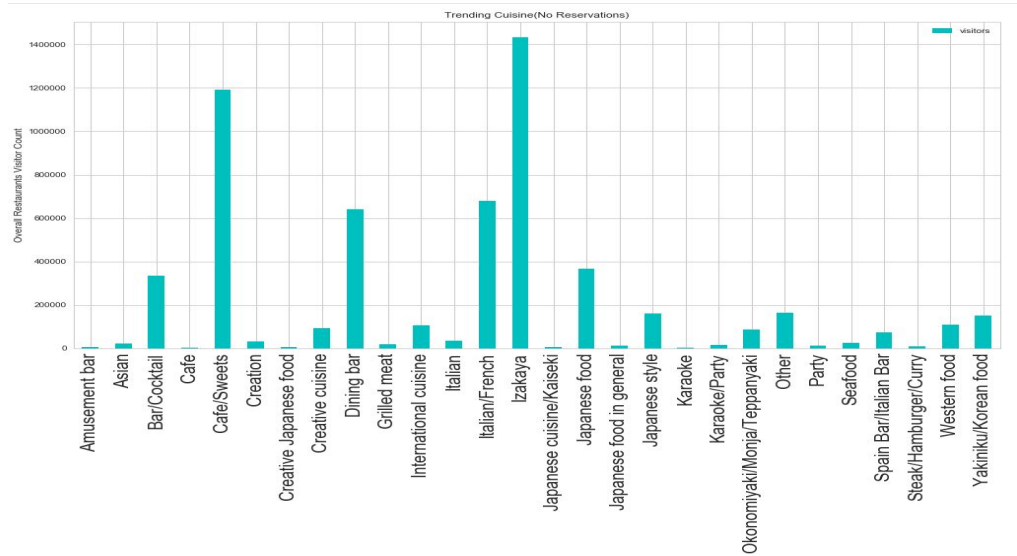
Feature Engineering

- Added holiday feature using date_info table based on calendar date
- Included min_visitors, max_visitors, median_visitors, reserve_visitors feature
- Added reserve_time_diff feature which represents the time difference between reservation date and visited date.

Exploratory Data Analysis

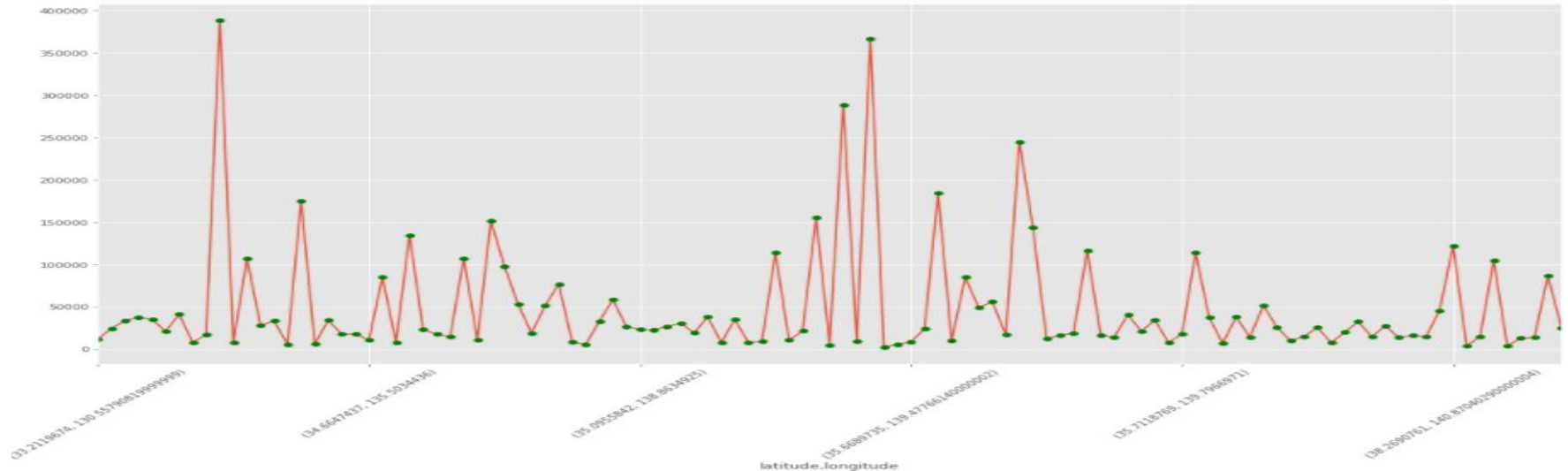


Food Genre Preference



Key Findings

- When there is no holiday for a restaurant, more number of visitors are expected in weekends(Saturday, Sunday)
- The location where sum of latitude and longitude is 175 , having majority of visitors
- In both holiday and no-holiday days Monday has lower number of visitors
- Most of the visitors in all restaturants prefer to order **Izakaya** food, the visitor count is more than 1400000
- The third most eaten food in Japanese restaturants is **Italian/French** cuisine

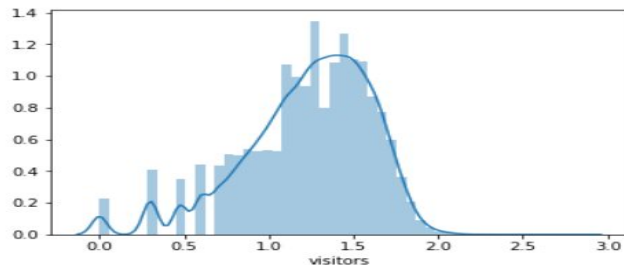


Inferential Statistics

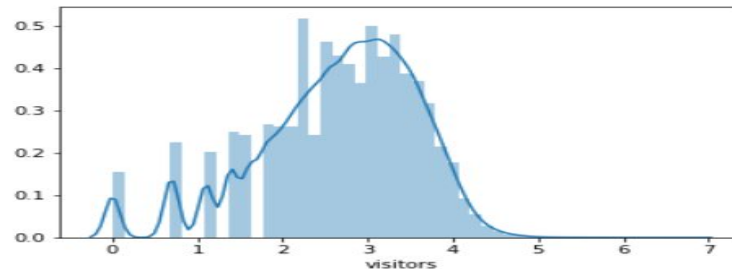
Forming Hypothesis

- **Null Hypothesis** : : Average number of visitors ordering Izakaya food is equal to average number of visitors ordering Other food types
- **Alternante Hypothesis** : : Average number of visitors ordering Izakaya food is greater than the average number of visitors ordering Other food types
- The threshold value of α is assumed to be 0.05. Assuming Null Hypothesis is true.
- It is evident from t-test, mean number of customers eating Izakaya food is statistically significant.

Izakaya Distribution



Other Food genre distribution



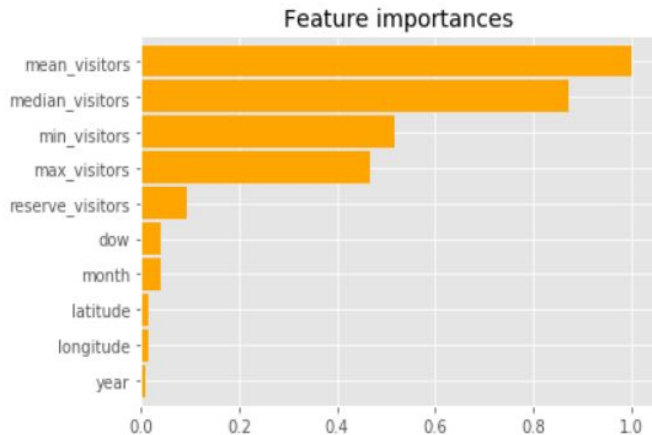
Forecast Modelling

- Used Linear Regression as baseline model to keep baseline score for further models.
- Hyperparameters are tuned with 2-Fold cross validation

Model	Parameters	Root Mean Square(rmse) on final test set
Linear Regression(Baseline model)		0.568
Random Forest Regressor	'max_depth': 15, 'max_features': 10, 'min_samples_leaf': 1, 'n_estimators': 100	0.512
XG Boost Regressor	n_estimators=100, learning_rate=0.2	0.505

Feature importance

Feature importances from Random Forest Regressor. Here considered top ten features.



Future Work

Though Random Forest model, as given rmse score of 0.505 with hyper parameter tuning using KFold training.

From the Random Forest model, mean visitors, max visitors, date of week, year, month , reserve visitors are playing key role in forecasting.

Expecting to implement Neural Network model adding more data if possible to improve the rmse test score.