PARKING LOT MODEL

Problem Statement

- The management team of ABC mall wants to solve the issues that they faced during the Time-out sale in the previous year. As the sale is going to begin this year, the management has started the preparation of the sale by collecting the parking data from the past few months.

- The management wants to understand if they must increase the parking capacity or divert the vehicles for sale this year.

Task

- The task is to forecast the occupancy of vehicles in the parking space.

PARKING LOT MODEL

About the dataset

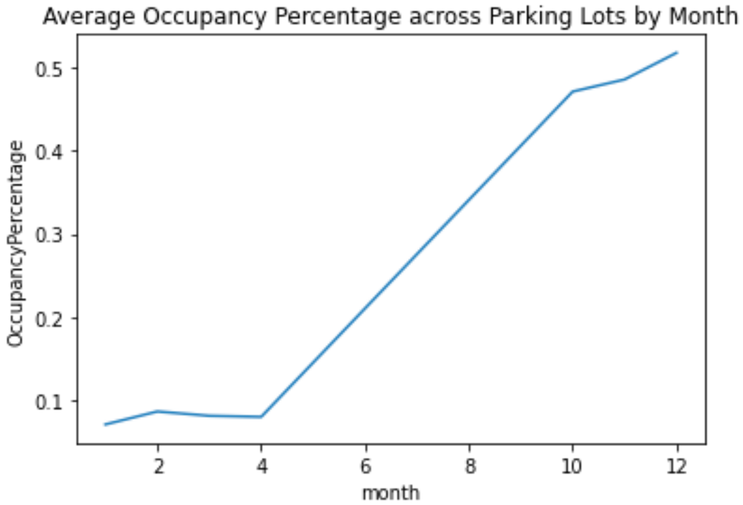
- SystemCodeNumber: ID code for the parking area
- Capacity: The number of parking spots available
- ExitRate: Represents the exit rate of cars
- FillRate: Represents the fill rate of cars
- OccupancyPercentage: Represents the occupancy percentage
- OccupancyTrend: Represents the trend in the car occupancy
- Occupancy: The number of parkings spots actually occupied at the time
- LastUpdated: The date and time of the observation

Shape of the dataset

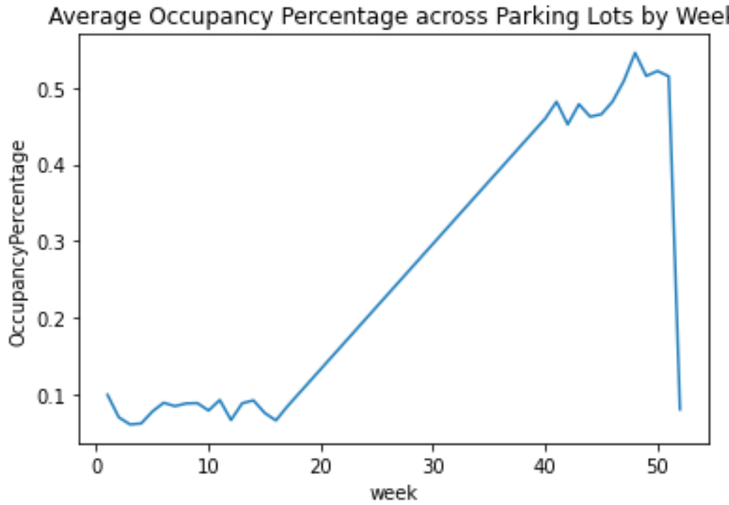
- 38,034 instances and 8 features in the dataset

PARKING LOT MODEL

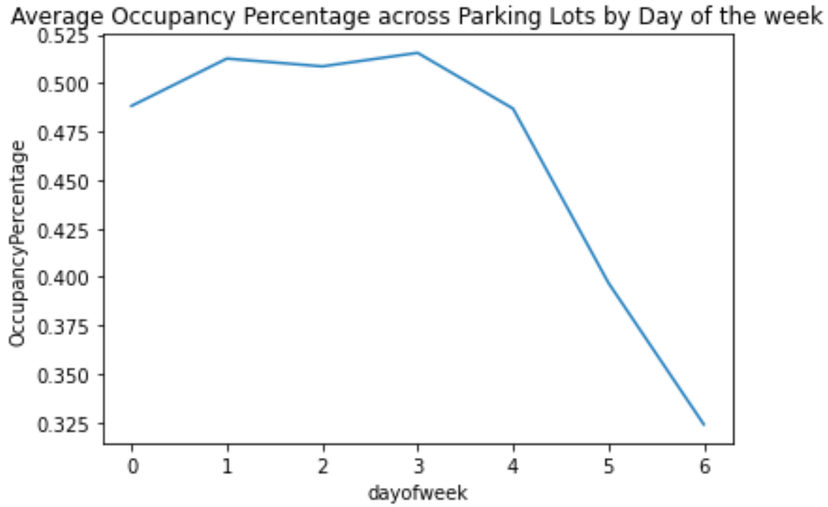
Exploratory Data Analysis



We understand that Parking Demand rises during festive / winter months



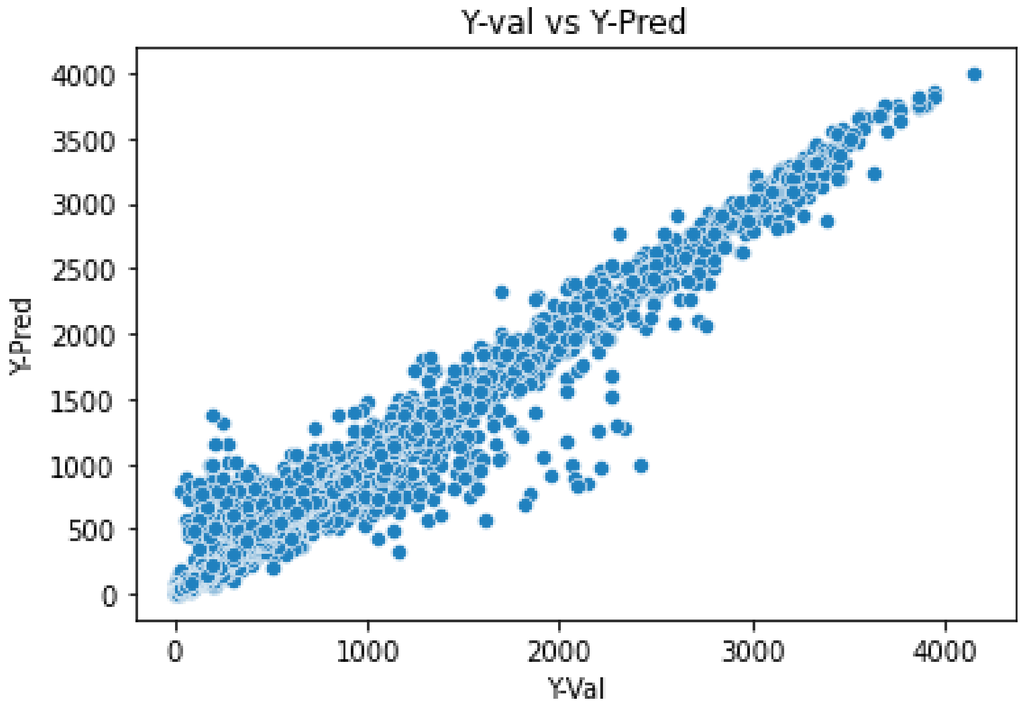
As we look in more detail, we understand that parking demand falls sharply at the end of the year during the X-mas week



As expected, Parking demand rises falls sharply during weekends. It could be the case that the Parking lots are near Office area

PARKING LOT MODEL

Decision Tree Regressor

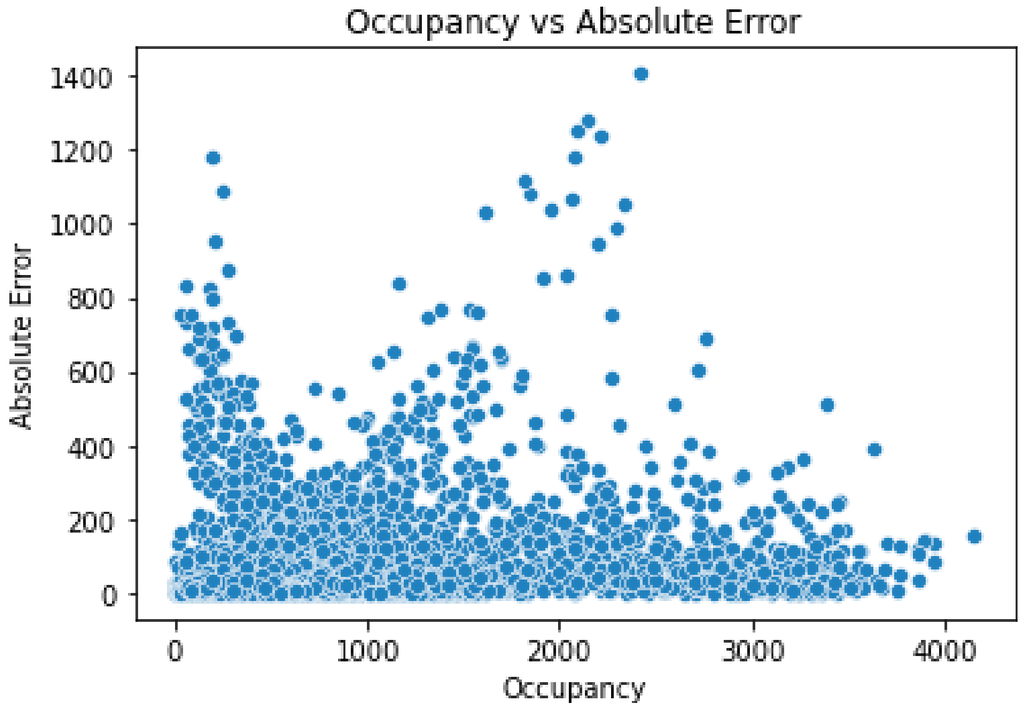


Y-Val vs Y-Pred

- Y-Val and Y-Pred are fitting correctly. From the chart, we can infer that our predictions are correct on the validation set.
- However, it could be the case that our model is strongly overfitting the dataset.

PARKING LOT MODEL

Decision Tree Regressor



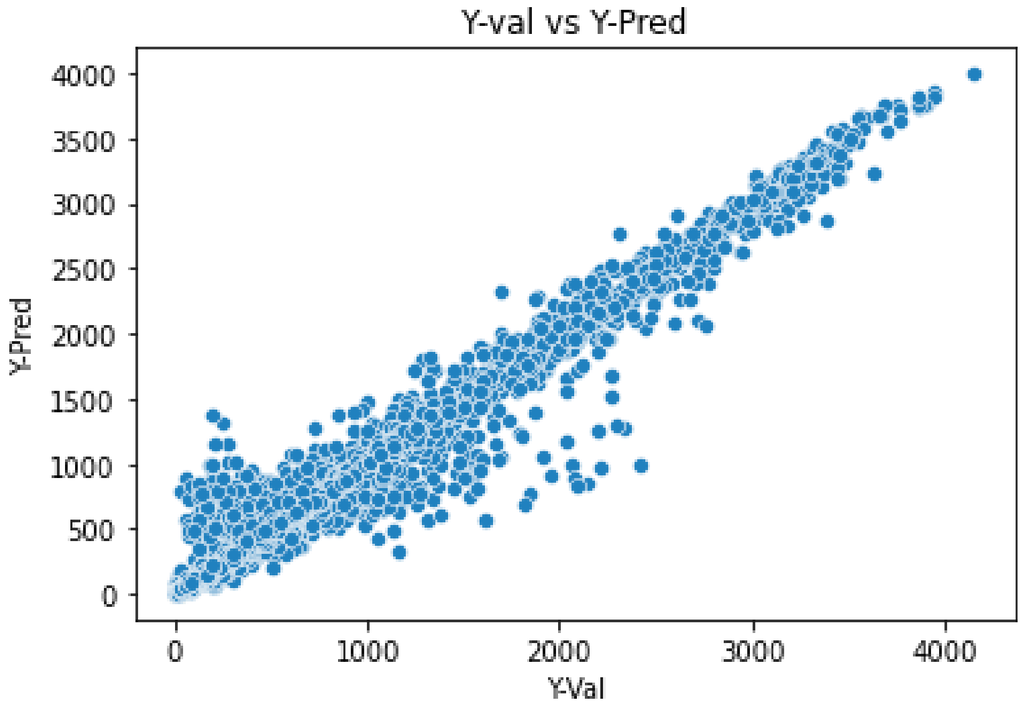
Occupancy vs Absolute Error

- We see that our model is correctly predicting Occupancy across different levels as the Absolute error is clustered in a tight range
- However, we need to look in the case for predictions wherein the Absolute Error is higher than average for Low Occupancy instances

Validation Set R2 Score is 0.96

PARKING LOT MODEL

Random Forest Regressor

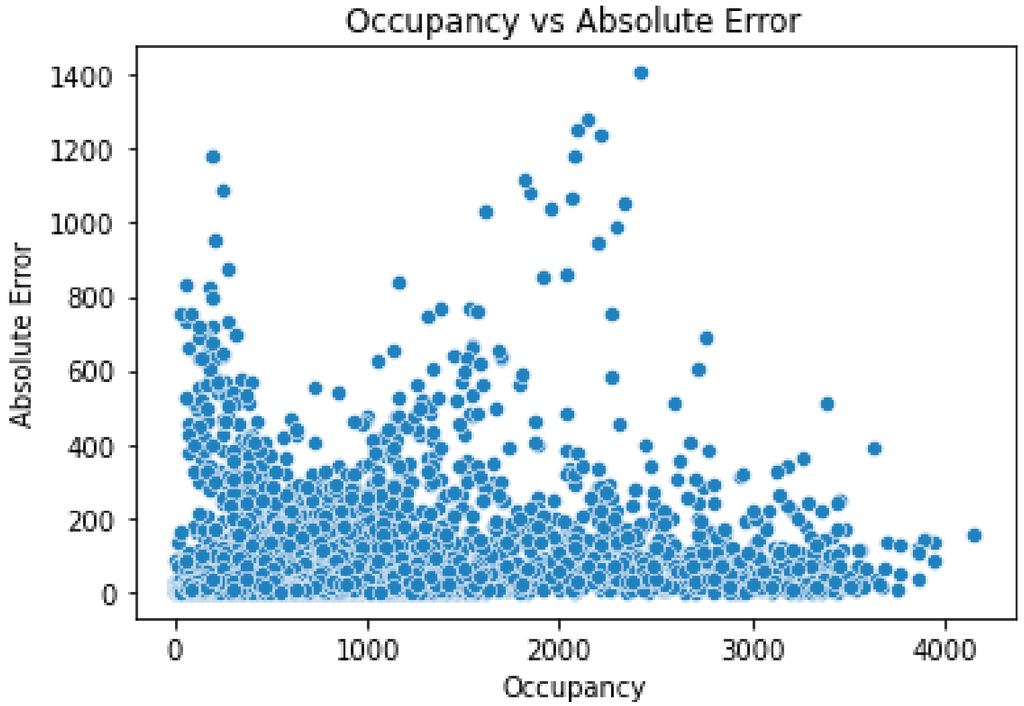


Y-Val vs Y-Pred

- Y-Val and Y-Pred are fitting correctly. From the chart, we can infer that our predictions are correct on the validation set.
- However, it could be the case that our model is strongly overfitting the dataset.

PARKING LOT MODEL

Random Forest Regressor



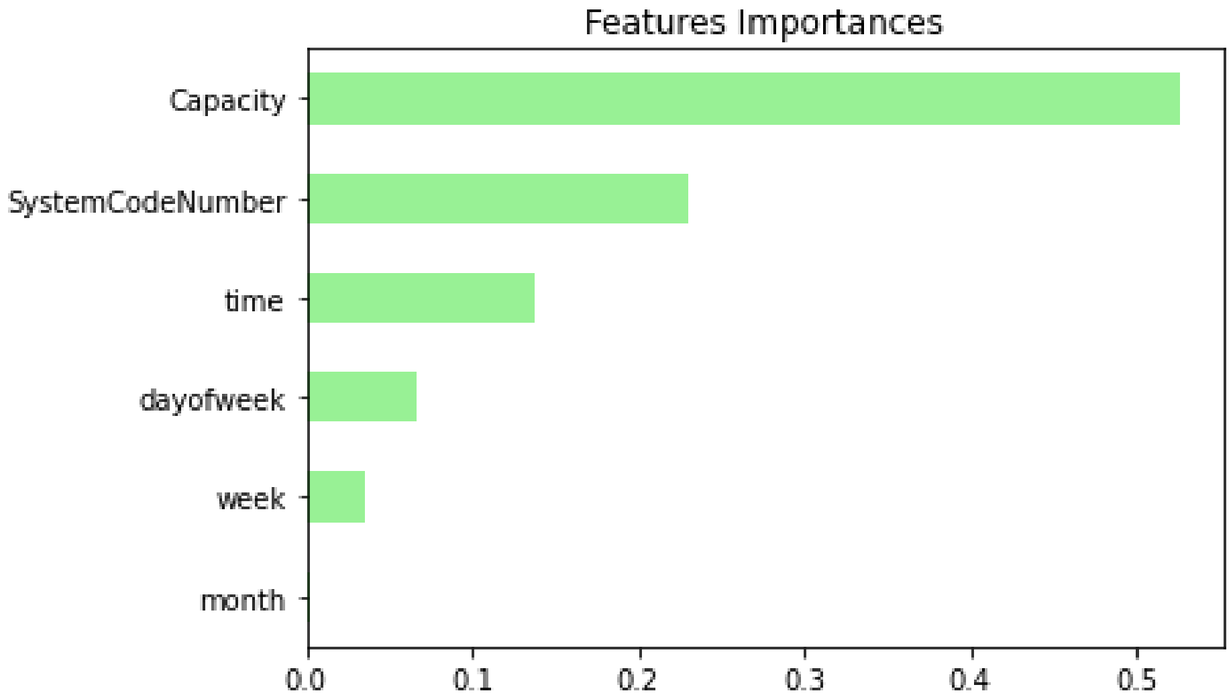
Occupancy vs Absolute Error

- We see that our model is correctly predicting Occupancy across different levels as the Absolute error is clustered in a tight range
- However, we need to look in the case for predictions wherein the Absolute Error is higher than average for Low Occupancy instances

Validation Set R2 Score is 0.975

PARKING LOT MODEL

Feature Importance



Capacity, Parking Lot ID and Time are the key features for our model

- It could be the case that there is little variation in the occupancy levels of parking lots and capacity serves as a good predictor of the occupancy levels.
- This makes sense as the parking patterns don't change much on a regular basis. If it is an office area, on average there won't be much variation in the parking patterns of office goers.

IMPROVEMENT AREAS

There is always scope for improvement!

Formatting / Presentation

- Better formatting of charts and Jupyter notebook code
- Some more detail in comments

Code

- Use pipelines so that we don't have to repeat the code process for test case data
- Don't repeat code for similar functionality
- Detailed Exploratory Data Analysis

Model Tuning and Validation

- Use GridSearchCV and other approaches to tune hyperparameters of the model
- Find the balance between bias and variance in the model
- Select the best performing model
- Analysis of errors

THANK YOU!