

PROJECT REPORT



Title of Project: Predicting the House Rents

Name: Kalal Rohan Nitin

Year: 2020

Introduction

Determining the sale price and the rent of the house is very important, nowadays as the rent of the land and price of the house increases every year. So our future generation needs a simple technique to predict the house rents and the prices in a better and easy way. The price of house helps the buyer to know the rent of the house and also the right time to buy it. There are several factors that affect the price of the house such as the physical condition, location, landmark, area, insurance and etc.

Changes in the price of real estate can affect various investors in households, bankers, policy makers and many others. Investment in the real estate sector appears to be an attractive investment choice. Predicting the value of immovable property is therefore an important economic index.

Defining the Problem

Problem Statement:

Prices of real estate properties are sophisticatedly linked with our economy. Despite this, we do not have accurate measures of housing prices based on the vast amount of data available. Therefore, the goal of this project is to use machine learning to predict the selling prices of houses based on many economic factors.

Objective:

The aim is to predict the efficient house rents for real estate customers with respect to their budgets and priorities. By analysing previous market trends and rents ranging, and also upcoming developments future prices will be predicted.

Data Sources:

The data is provided by Cognifront. The Dataset contains of 10692 records and 13 features. This given dataset helps in predicting the house rents.

Tools and Techniques:

Here in this project different packages are used such as Sk Learn. It is an open source platform. It assists in regression. In this project two models are used: Linear Regression and Logistics Regression.

Limitations:

One of the major limitations of this project is that we can perform operations and predict the rents of the house from given dataset only.

Data Preparation and Pre-processing

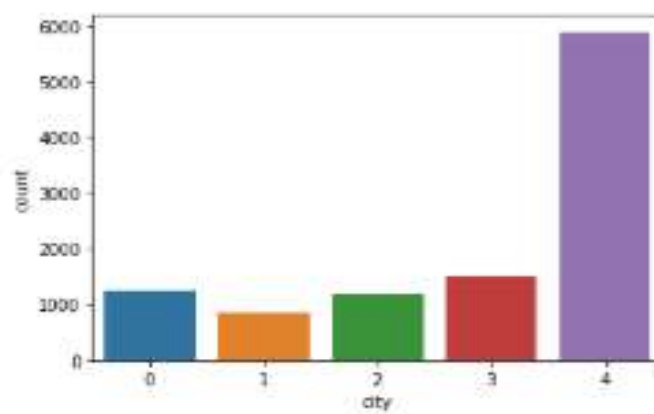
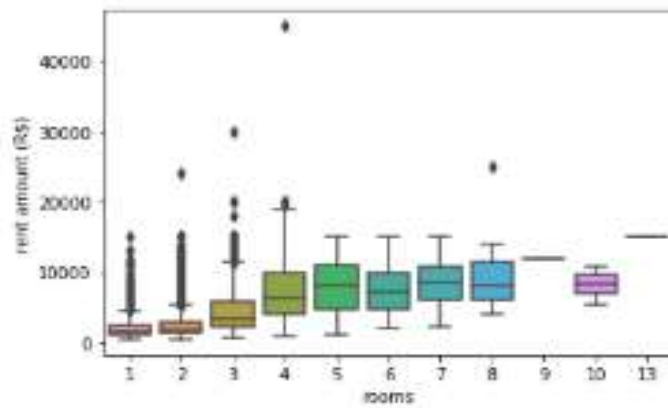
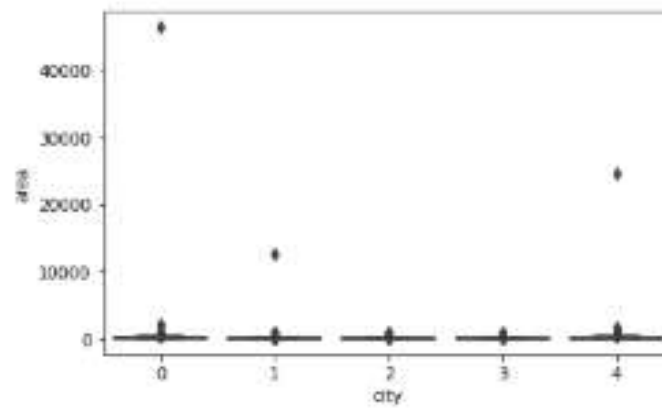
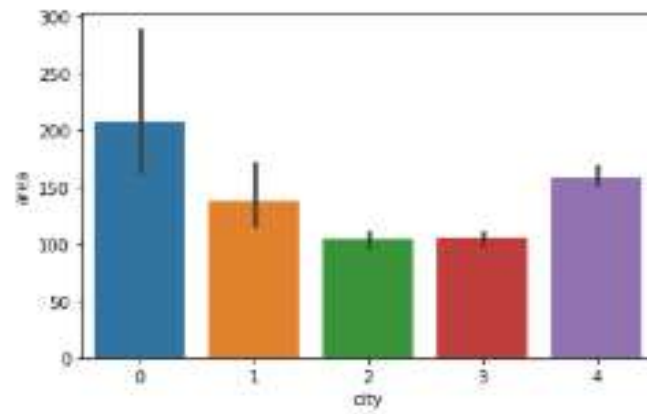
In this stage of project implementation, focus is put on data collection, data selection, data pre-processing, and data transformation.

Data Collection:

The Dataset contains of 10692 records and 13 features. This dataset contains information regarding Brazil houses. The dataset contains of name of the city, parameter such as area, number of rooms, whether the home is furnished or not, the rent of the house and total costing of the house.

Data Visualisation:

Here EDA is performed on the dataset using Visualisation such as count plot, bar plot and boxplot. For performing Visualization we have used packages like Matplotlib and Seaborn. Follow given below are the visual actions through EDA.



Labelling:

It is the process of detecting and tagging data samples. The process can be manual but is usually performed or assisted by software. We Label the Targeted Data.

Data Selection:

The data which is provided in dataset not all of it is useful, the data is to be selected relevantly and also on the base of priority.

Data Pre-processing:

It is a process of transforming the raw, complex data into systematic understandable knowledge. Here the whole dataset is checked and is seen whether the data which is given is completely useful or not.

If not operation like data cleaning is performed and in case if values are missing they are also filled.

Dataset Splitting:

The given dataset is split into three parts: training, testing, and validation sets. The ration of training and testing sets is typically 80 to 20 percent. The 20 percent of the training set is further split as a validation set.

Model Training:

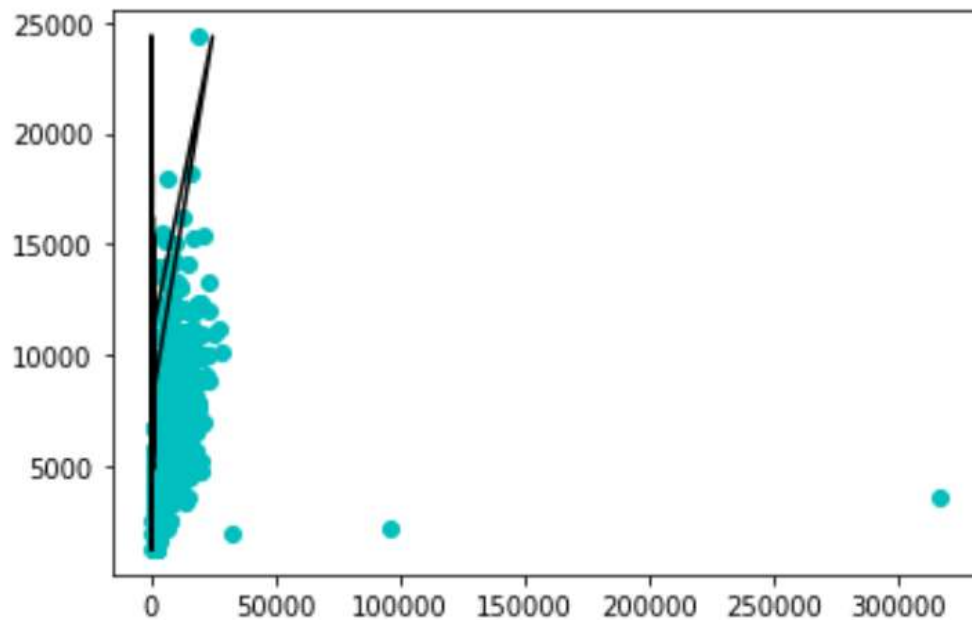
In this stage of Machine Learning we build a training model that is: the data is partitioned in training and testing model. Here we generally take 80 percent as training model. The basic purpose of this training is to develop a proper model.

Model Testing:

As discussed above the data is partitioned into training and testing model, out of which 20 percent is testing model. We use two different Machine algorithms to test our model and also we predict its output.

1. Linear Regression:

For finding a relationship between two continuous variables, linear regression is useful. One variable is predictor or independent, and the other variable is variable response or dependent. Given below is the Scatter plot for Linear Regression.



2. Logistic Regression:

It is appropriate regression model when variables are dependant. It is a predictive analysis. It is use to describe the data and relationship between one dependant variable with one or more nominal, ordinal, and interval.

Improvement in predictions:

In the given project, we can also use methods like Random Forest Label Encoder and Standard Scaler to improve the predictions of output on testing model.

Model Deployment

After a model is selected and validated, it is sent into production means it is put to some actual use. Usually the deployment is done by translating the model into another language which in our case is not done and thus the model is in Python Language.

There are different ways of Deployment depending on the size of team and available infrastructure.

There are number of ways of deploying a model in this project Batch Based deployment is done.

The prediction is done in batch of observations rather than on continuous basis.

Conclusion

Throughout this article we made a machine learning regression project from end-to-end and we learned and obtained several insights about regression models and how they are developed.

This was the first of the machine learning projects that will be developed on this series.