**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **4-5** |
| Chapter 2 | Literature Survey | **6-8** |
| Chapter 3 | Methodology | **9-11** |
| Chapter 4 | Result and Discussion | **11-13** |
| Chapter 5 | Conclusion and Future Work | **14-15** |
|  | References | **16** |

**Chapter 1**

**Introduction**

* 1. **Introduction**

Accurately estimating the value of real estate is an important problem for many stakeholders including house owners, house buyers, agents, creditors, and investors. It is also a difficult one. Though it is common knowledge that factors such as the size, number of rooms and location affect the price, there are many other things at play. Additionally, prices are sensitive to changes in market demand and the peculiarities of each situation, such as when a property needs to be urgently sold. The sales price of a property can be predicted in various ways, but is often based on regression techniques. All regression techniques essentially involve one or more predictor variables as input and a single target variable as output. In this paper, we compare different machine learning methods performance in predicting the selling price of houses based on a number of features such as the area, the number of bed- and bathrooms and the geographical position.

* 1. **Why we use Data Science ?**

Data Science solves real business problems by utilizing data to construct algorithms and create programs that help in proving optimal solutions real business problems by using hybrid models of math and computer science to get actionable insights.

**1.2 Role of Machine Learning in predictive analytics ?**

We use Machine Learning for predictive analytics .Predictive analytics involves certain manipulations on data from existing data sets with the goal of identifying some new trends and patterns. These trends and patterns are then used to predict future outcomes and trends. By performing predictive analysis, we can predict future trends and performance. It is also defined as the prognostic analysis, the word prognostic means prediction. Predictive analytics uses the data, statistical algorithms and machine learning techniques to identify the probability of future outcomes based on historical data.

* 1. **Why Python?**

Python is a general-purpose, and high-level programming language which is best known for its efficiency and powerful functions. Python is loved by data scientists because of its ease of use, which makes it more accessible. Python provides data scientists with an extensive amount of tools and packages to build machine learning models. One of its special features is that we can build various machine learning with less-code.

* 1. **What is Price prediction?**

Price prediction uses an algorithm to analyze a product or service based on its characteristics, demand, and current market trends. Then the software sets a price at a level it predicts will both attract customers and maximize sales.

**Chapter 2**

**Literature Survey**

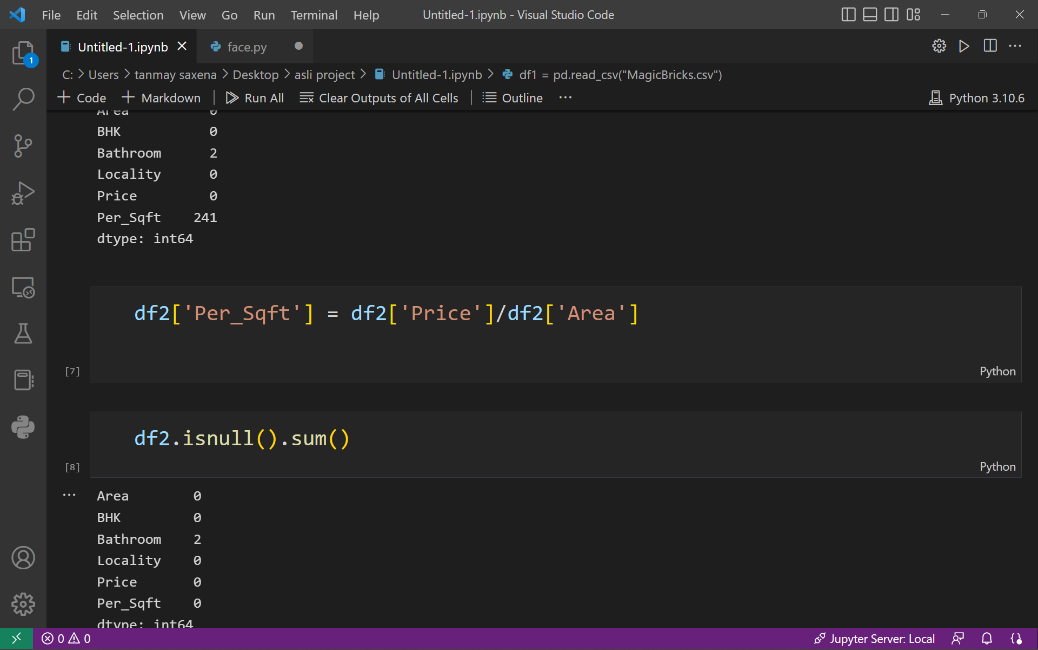
* Literature survey Manasa and Gupta have taken Bengaluru as city for case study. The property size in square feet, location, and its facilities are all key aspects affecting cost 9 different attributes are used. The Multiple linear regression (Least Squares), Lasso/Ridge regression, SVM, and XG Boost are used for experimental work.
* Panjali and Vani state that forecasting the resale price of a house on a long-term is vital, especially for those who will be residing there for a considerable duration while selling it again later. It also applies for those who want no risks while the dwelling is being constructed. Authors utilize various classification methods such as Logistic regression, Decision tree, Naive Bayes, and Random Forest to work out the house’s resale value. It also applies AdaBoost technique to assist weak learners to be strong ones. The physical characteristics, location, as well as numerous economic aspects persuading at the time decide the resale price of a house. Accuracy is used to measure performance for different datasets and unleash the optimal way for sellers while expecting the resale price.
* Varma et al. worked upon regression techniques. They used weighted average of the multiple techniques. The selling price of a house is forecasted using fuzzy, ANN, and KNN.
* Mukhlishin et al. assess the predictive performance to the real-world selling price of a house using MAPE. The fuzzy technique beats neural networks and k-nearest neighbour for house price prediction in data training using part of dataset, as per the findings of the experiments. Madhuri et al. focused to anticipate house prices based upon their financial capacity and objectives in continuous manner, for people looking for their first potential house. Prospective prices will be derived by evaluating the prior merchandise, rental ranges, and forthcoming developments. Multiple linear, Ridge, LASSO, Elastic Net, Gradient boosting, and Ada Boost Regression are among the regression techniques employed during work. Physical situations, concept, and locality were properly considered while estimating. The approach of authors in for predicting a house sale price blends common ML techniques with their original ideas like the residual regressor, logit transform, and neural network machine. In this paper , the objective of this essay is to look into a few models for predicting property prices. Three ML algorithms, including Random Forest, XGBoost, and LightGBM, as well as two ML methodologies, Stacked Generalization Regression and Hybrid Regression, are evaluated. On the training set, the RandomForest approach has the lowest deviation. However, it tends to overfit. LightGBM gives the best accuracy. Hybrid Regression method performs much better than the three prior methods. Dharwadkar and Arage took a different approach. They used scheduled rates of construction projects of last 12 years to predict project cost using OLSR and MLP techniques. MLP techniques proved to avail best accuracy ranging between 91-98%.

**Chapter 3**

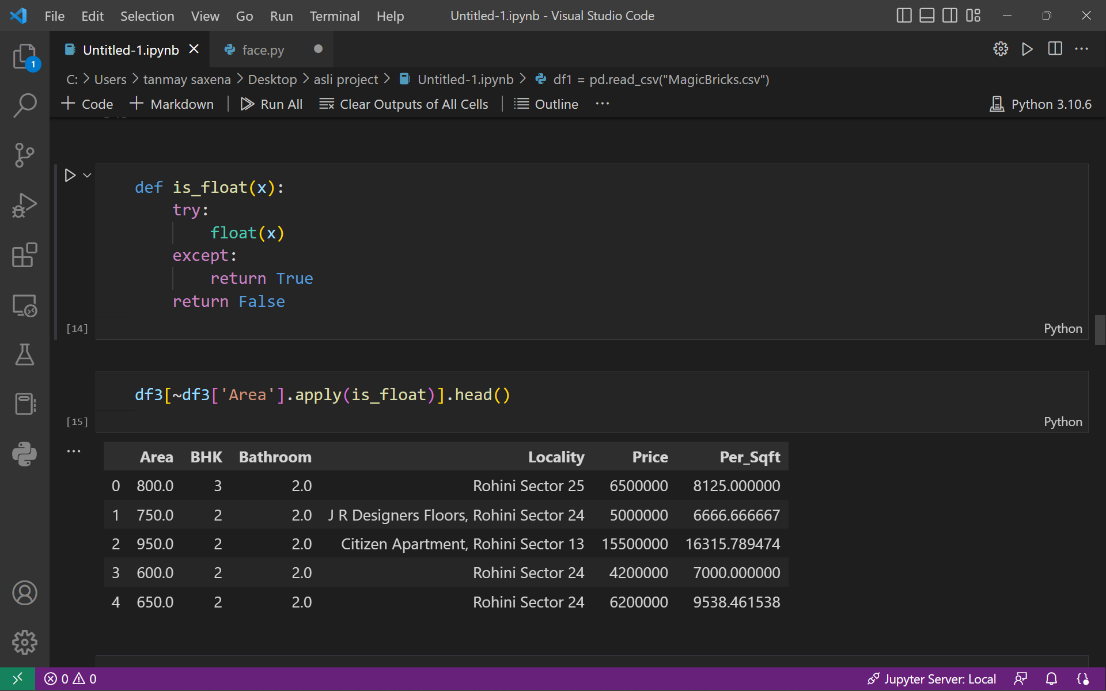
**Methodology**

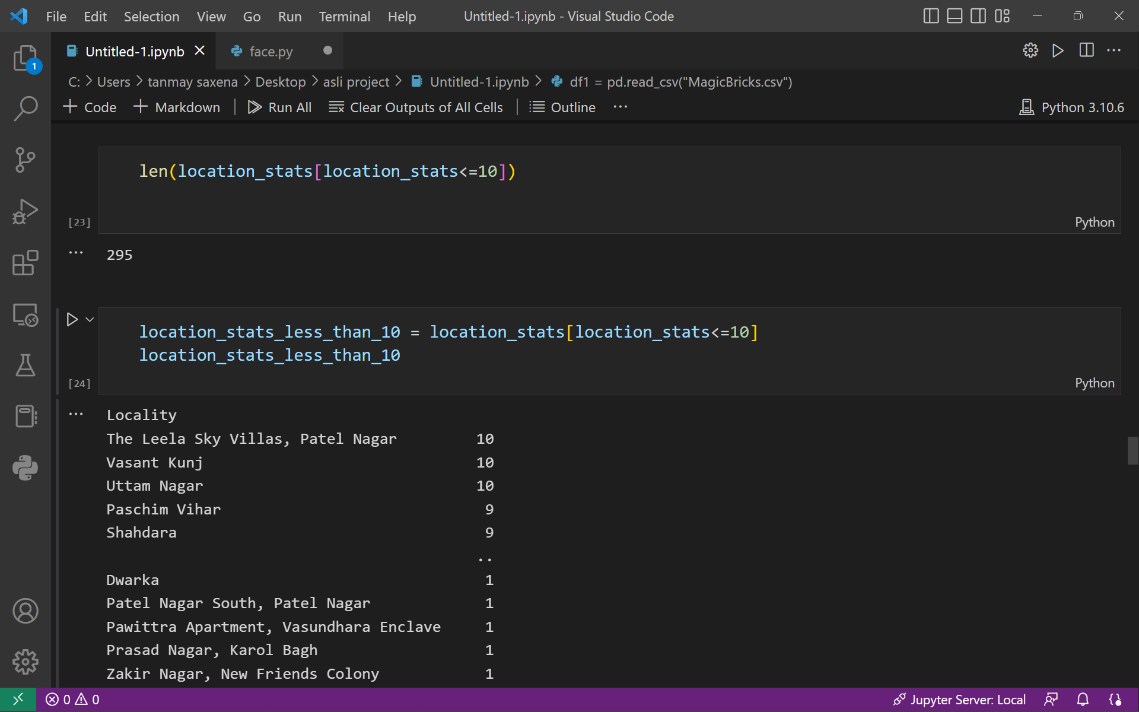
**3.1 Data Cleaning**

**3.1.1 Removing Null**

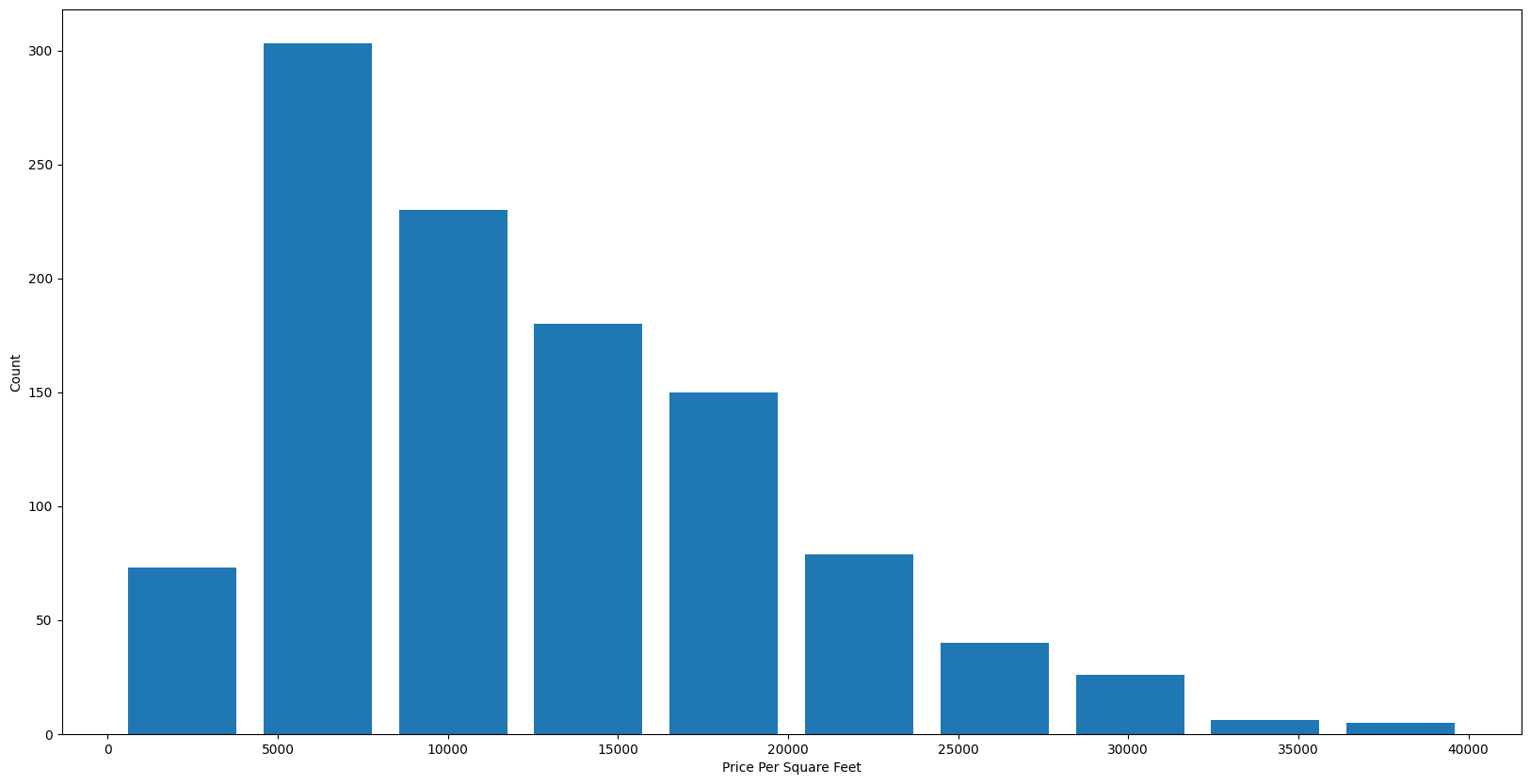
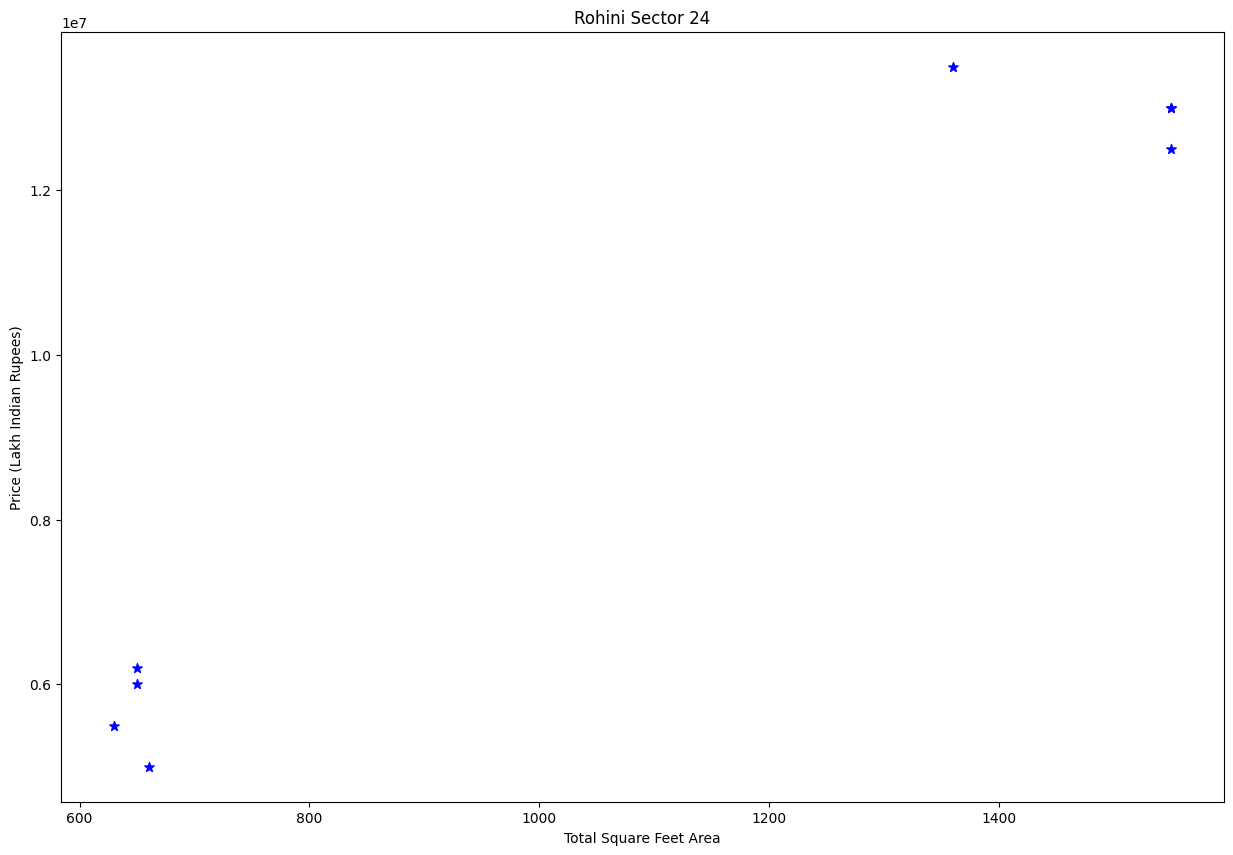
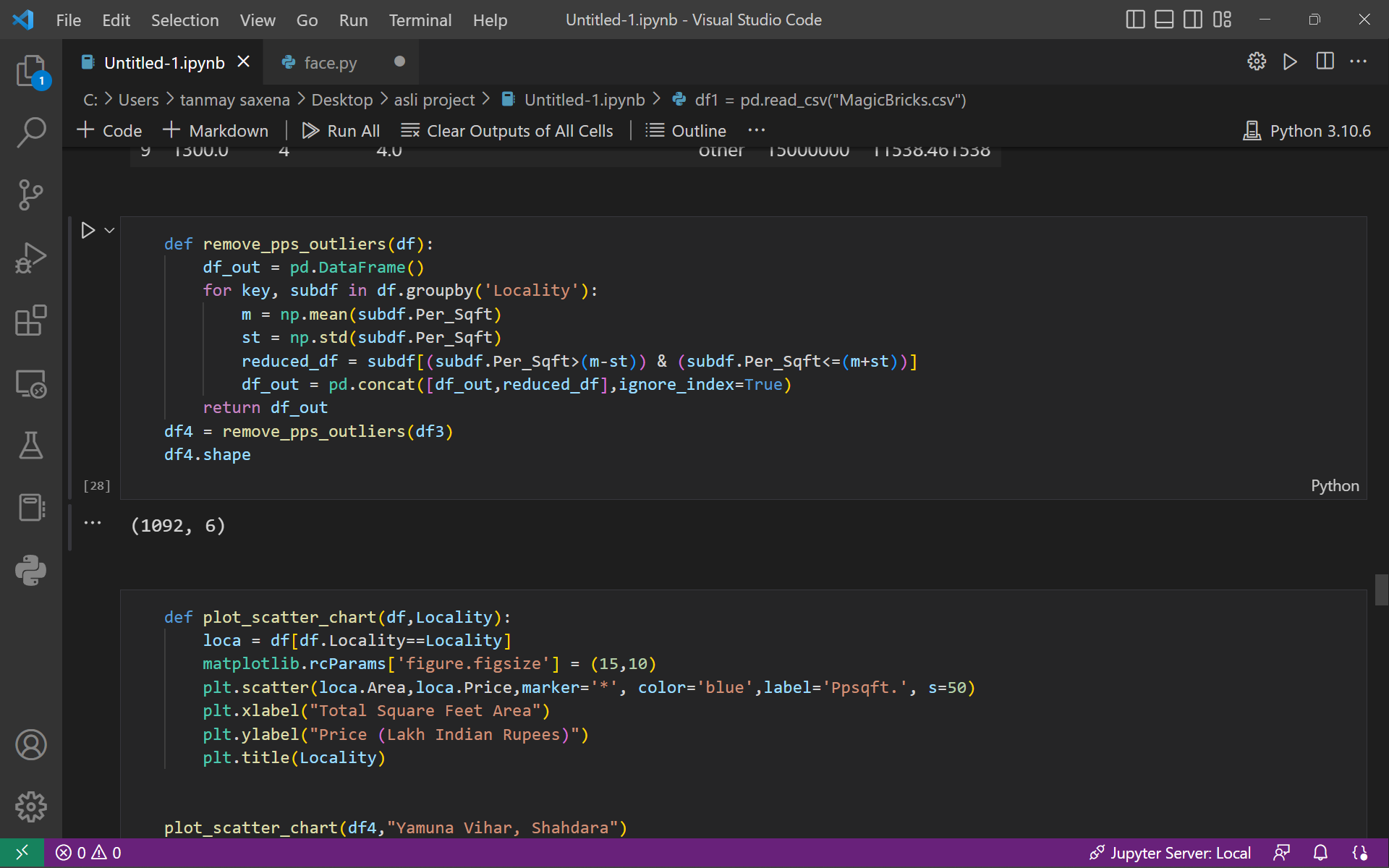
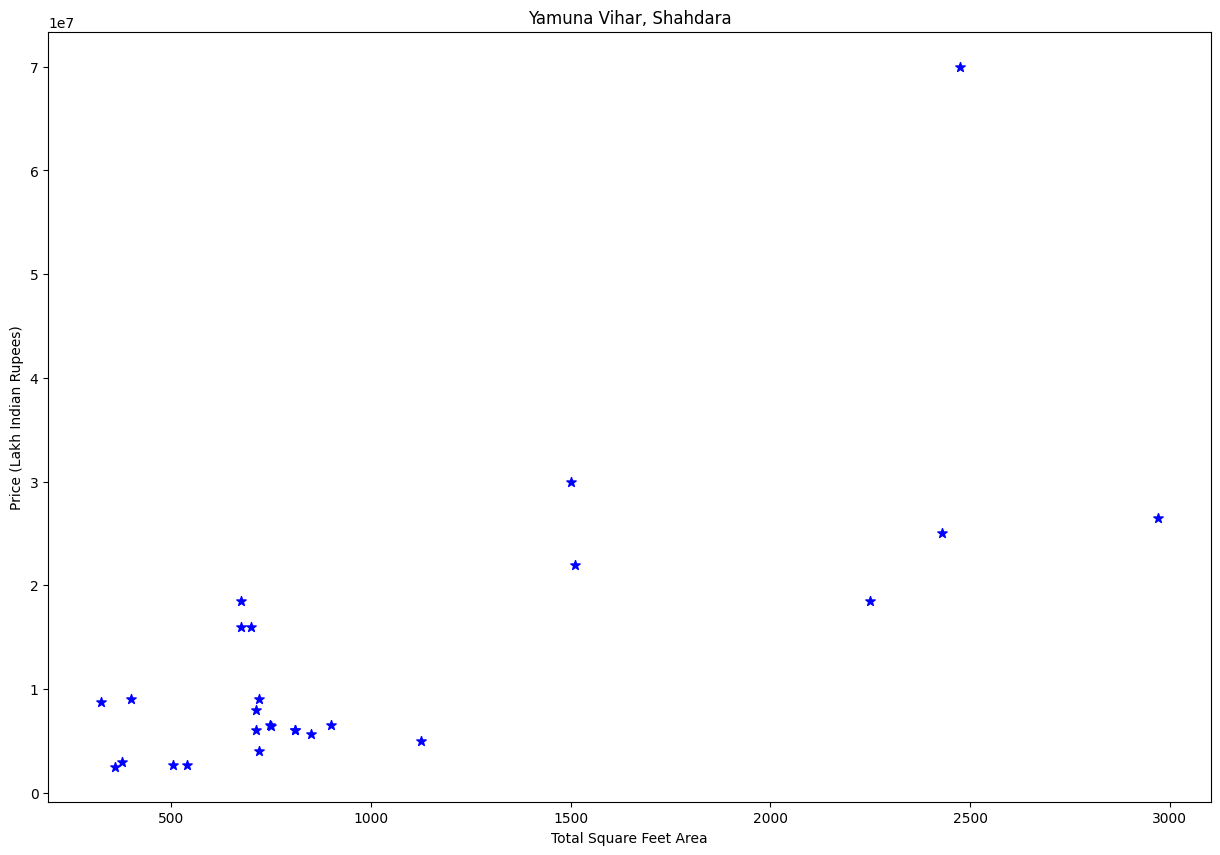


**3.1.2 Making all Area into Float(for easy accessing)**

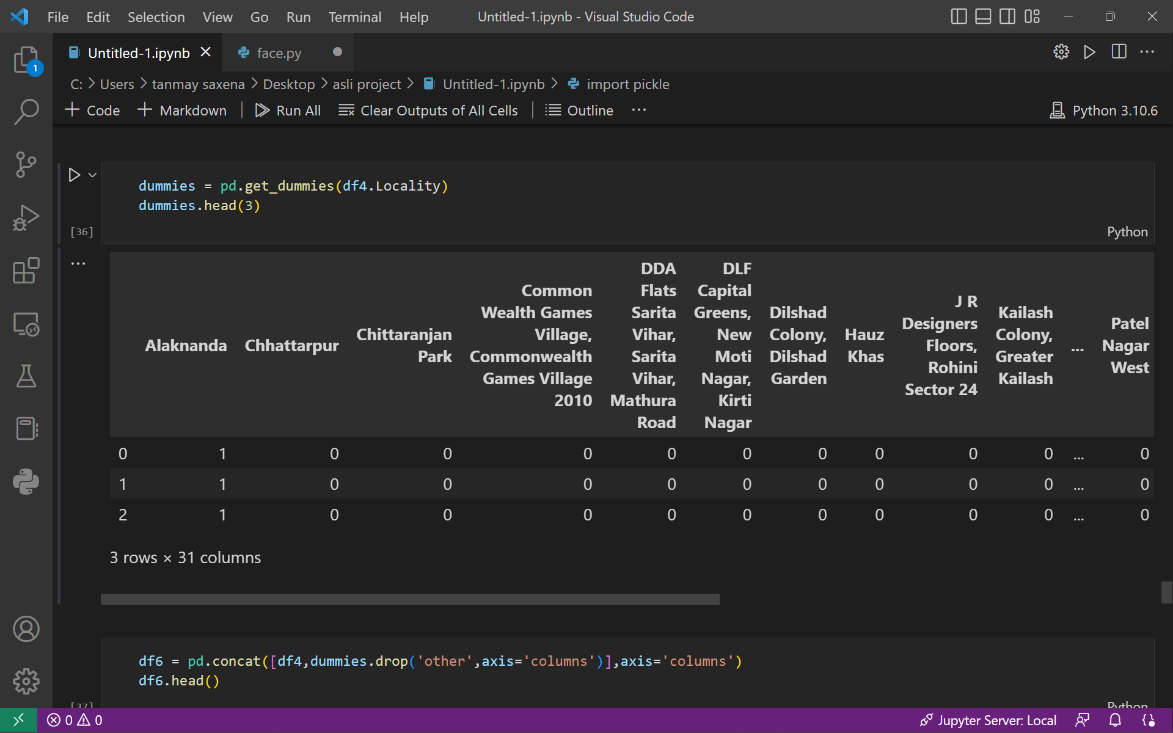


**3.1.3 Dimensionality Reduction**

**3.1.4 Outlier Removal**



**3.2 One Hot encoding for Locality**



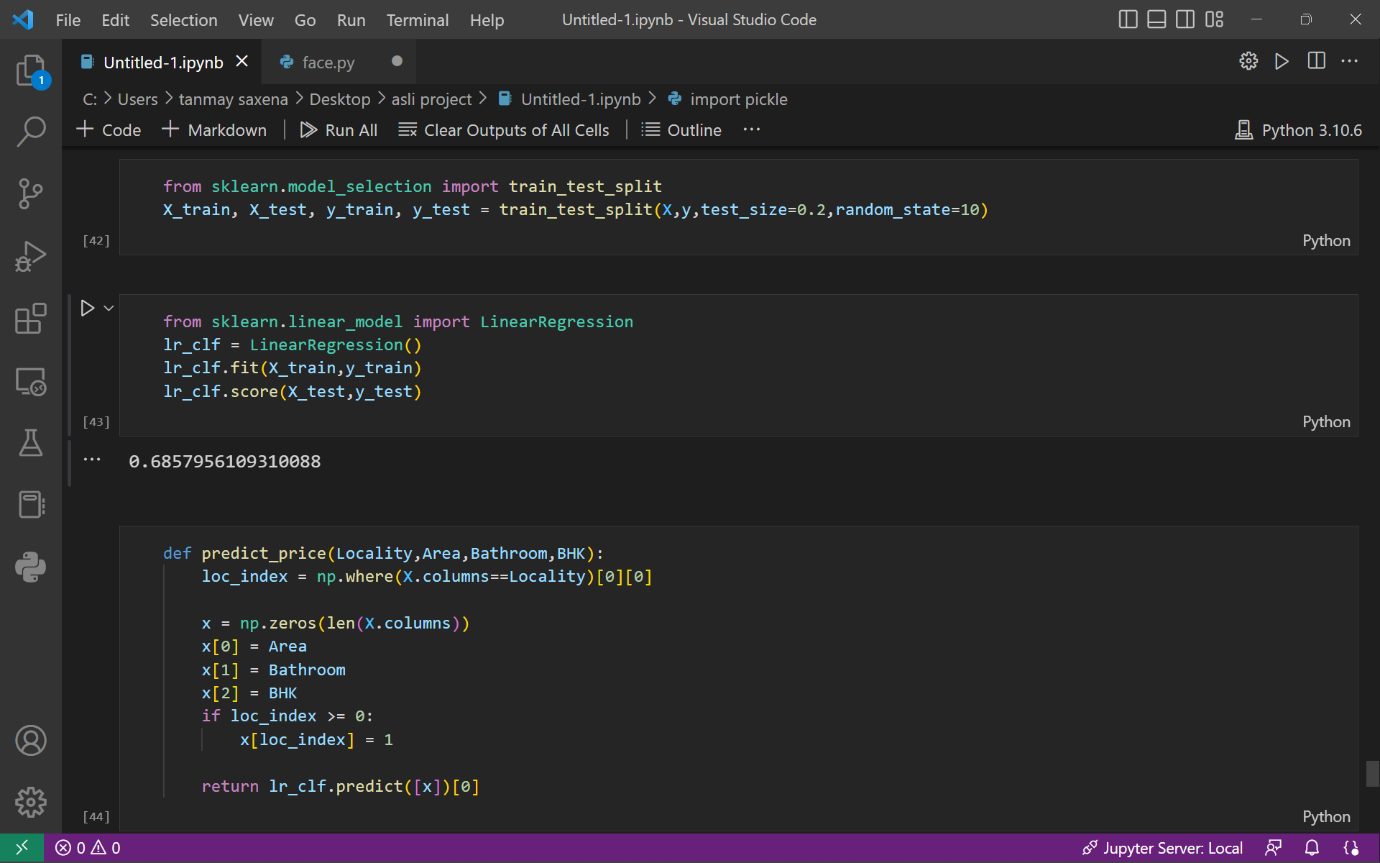
**Chapter 4**

**Result and Discussion**

**4.1 By Using linear regression technique(Build Model)**

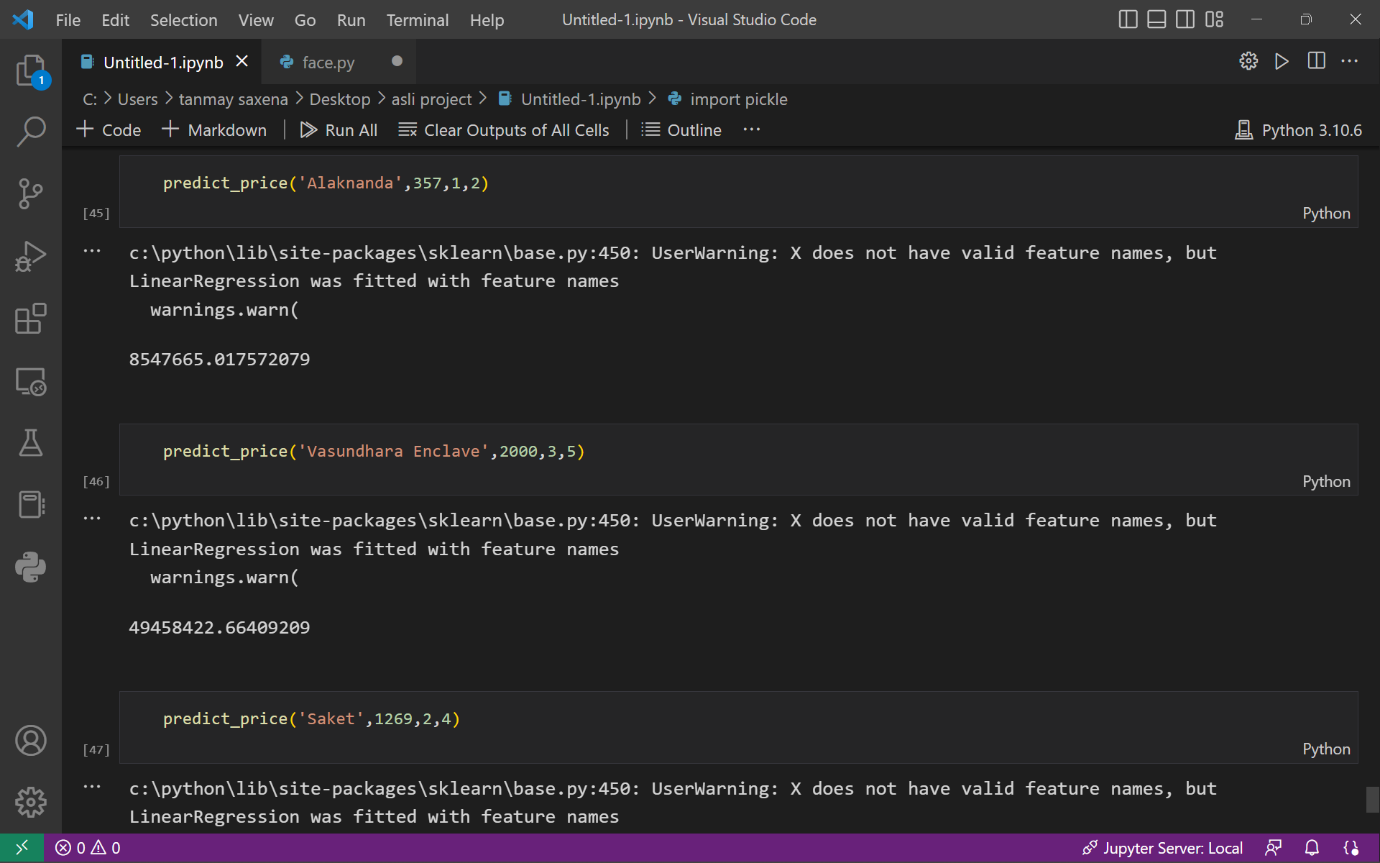
**Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used. There are many names for a regression’s dependent variable.  It may be called an outcome variable, criterion variable, endogenous variable, or regressand.  The independent variables can be called exogenous variables, predictor variables, or regressors.

Linear regression is used in many different fields, including finance, economics, and psychology, to understand and predict the behaviour of a particular variable. For example, in finance, linear regression might be used to understand the relationship between a company’s stock price and its earnings, or to predict the future value of a currency based on its past performance.



**4.2 Result**

**We predict some house prices as shown below:-**



**Chapter 5**

**Conclusion and Future Work**

Buying your own house is what every human wish for. Using this proposed model, we want people to buy houses and real estate at their rightful prices and want to ensure that they don't get tricked by sketchy agents who just are after their money. Additionally, this model will also help Big companies by giving accurate predictions for them to set the pricing and save them from a lot of hassle and save a lot of precious time and money. Correct real estate prices are the essence of the market and we want to ensure that by using this model.

Housing price prediction has received conceivable attention by researchers. There are routine set of factors like no of rooms, carpet area, locality, floor number which affect price of a residential property. I present survey of different ML and DL techniques that have been employed by researchers. Random forest and gradient boosting techniques have proven to produce more accurate results. As part of this work, I propose to increase and hybridize the dataset by combining data of residential properties in Delhi city in predefined localities from more than one real estate websites e.g., 99acres.com, nobroker.com, magicbricks.com. It’s proposed to develop a model using complex ML and DL techniques that helps find out predicted price for a flat of given configuration more accurately. Similarly, I will collect data of rental properties of similar configuration listed in same date range. I will prepare a model that helps predict rental income for a flat of given configuration. If a user wants to buy a flat in given locality in Delhi, the model will be able to provide a yes/no recommendation based upon financial ratio of such investment.

The system is apt enough in training itself and in predicting the prices from the raw data provided to it. After going through several research papers and numerous blogs and articles, a set of algorithms were selected which were suitable in applying on both the datasets of the model. After multiple testing and training sessions, it was determined that the XGBoost Algorithm showed the best result amongst the rest of the algorithms. The system was potent enough for Predicting the prices of different houses with various features and was able to handle large sums of data. The system is quite user-friendly and time-saving.

The supplementary feature that can be added to our proposed system is to avail users of a full-fledged user interface so there can be multiple functionalities for users to use with the ML model for numerous locations. Also, an Amazon EC2 connection will take the system even further and increase the ease of use. Lastly, developing a well-integrated web application that can predict prices whenever users want it to will complete the project.

**References**

* Youtube
* <https://youtube.com/playlist?list=PLeo1K3hjS3uu7clOTtwsp94PcHbzqpAdg>
* <https://youtube.com/playlist?list=PLfP3JxW-T70GM0TzYaDz4LtnADjXv1jFK>
* Data set from Kaggle
* Delhi House Price data Set
* kaggle datasets download -d neelkamal692/delhi-house-price-prediction
* <https://www.kaggle.com/datasets/neelkamal692/delhi-house-price-prediction>
* MagickBricks.com
* <https://www.geeksforgeeks.org>
* https://www.sciencedirect.com