# “HOUSE PRICE PREDICTION USING ML”

***An Internship project report submitted in partial fulfillment of the requirements for the award of the degree of***

##### BACHELOR OF TECHNOLOGY IN

**COMPUTER SCIENCE ENGINEERING**

***Internship***

***By***

Name – **Gopal Kumar**

College – **CUTM**

Under the guidance of

**INTERNFORTE**

****

**INTERNFORTE**

**Fourth Floor, Classic Arena, Hosur Rd, AECS Layout**

**Singasandra, Bengaluru, Karnataka 560068**

### ACKNOWLEDGEMENT

We would like to express our deep gratitude to our project guide **InternForte,** Assistant Professor, Department of Computer Science and Engineering, for his/her guidance with unsurpassed knowledge and immense encouragement. We are grateful to **InternForte**, Head of the Department, Computer Science and Engineering, for providing us with the required facilities for the completion of the project work. We are very much thankful to the Principal **InternForte,** for their encouragement and cooperation to carry out this work. We express our thanks to Project Coordinator **InternForte** for her Continuous support and encouragement. We thank all **teaching faculty** of Department of CSE, whose suggestions during reviews helped us in accomplishment of our project. We would like to thank **InternForte** of the Department of CSE for providing great assistance in accomplishment of our project. We would like to thank our parents, friends, and classmates for their encouragement throughout our project period. At last but not the least, we thank everyone for supporting us directly or indirectly in completing this project successfully.

**PROJECT STUDENT**

**Gopal Kumar**

### CERTIFICATE

This is to certify that the project report entitled “**HOUSE PRICE PREDICTION USING ML”** submitted by **Gopal Kumar** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science Engineering** of Centurion University of Technology and Management PKD Campus Odisha is a record of Bonafede work carried out under my guidance and supervision.

##### Project Guide Head of the Department

**INTERNFORTE INTERNFORTE**

### DECLARATION

We **GOPALKUMAR**, of final semester(6th) B.Tech. in the department of Computer Science and Engineering from CUTM, PARALAKHEMUNDI, hereby declare that the project work entitled **HOUSE PRICE PREDICTION USING ML** is carried out by us and submitted in partial fulfillment of the requirements for the award of **Bachelor of Technology in Computer Science Engineering** , under Centurion University of Technology and Management Paralakhemundi Campus Odisha during the academic year 2022-23 and has not been submitted to any other university for the award of any kind of degree.

**Gopal Kumar**

### ABSTRACT

The prediction of house prices plays a vital role in real estate market analysis, investment decisions, and property valuation. In recent years, machine learning (ML) techniques have emerged as powerful tools for accurately forecasting house prices. This study presents a comprehensive analysis of various ML algorithms employed for house price prediction and compares their performance to identify the most effective model. The dataset used in this study consists of a wide range of features related to houses, such as location, size, number of rooms, amenities, and proximity to essential facilities. Initially, the data is preprocessed, including handling missing values, feature scaling, and categorical variable encoding. The dataset is then split into training and testing subsets to evaluate the models. Several popular ML algorithms, including linear regression, decision trees, random forests, support vector regression, and neural networks, are implemented and trained using the training data. Hyper parameter tuning techniques, such as grid search or Bayesian optimization, are applied to optimize the models' performance. The performance of the ML models is evaluated using various metrics, including mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). The results are analyzed and compared to identify the model that provides the most accurate predictions. Furthermore, the study explores the importance of features in predicting house prices using techniques such as feature importance ranking and feature selection. This analysis helps in understanding the significant factors influencing house prices and provides valuable insights for buyers, sellers, and real estate professionals. The experimental results demonstrate the effectiveness of ML techniques in house price prediction. The findings indicate that [mention the model with the best performance] outperforms other models, achieving the lowest MSE, MAE, and highest R^2 scores. Moreover, the feature importance analysis reveals that location, size, and amenities significantly impact house prices.

**TABLE OF CONTENTS**

**S.No. Chapter Name Page No.**

**Abstract**

**1 Introduction** 07

**2 Existing Method** 08

2.1 Industry Profile 09-10

**3** **Research Methodology** 11-12

3.1 Title of the Study 12

3.2 Statement of the Problem 12

3.3 Objectives of the Study

3.4 Research Design 13

3.5 Hypotheses 13

3.6 Sampling Technique 14

**4 Proposed System** 15-16

4.1 Statistical Tools Used 16-17

4.2 Scope of the Study 17

4.3 Limitations of the Study 17

4.4 Data Collection and Analysis 17

4.5 Findings 17

4.6 Suggestions 17

**5 Experimental Analysis and Result** 18-19

5.1 System Configuration

5.1.1 Hardware Requirements

5.1.2 Software Requirement

5.2 Functional Requirements

5.3 Non-Functional Requirements

5.4 Source Code with Output 19-51

**6 Conclusion and Learning Experience** 52-53

**References** 54

# CHAPTER 1 INTRODUCTION

In recent years, the field of machine learning has revolutionized various industries, and one prominent application is the prediction of house prices. House price prediction plays a crucial role in real estate, finance, and investment decision-making processes. By leveraging the power of machine learning algorithms, we can develop accurate and reliable models that estimate the value of residential properties. House price prediction involves analyzing a multitude of factors that influence property prices, such as location, size, number of bedrooms and bathrooms, amenities, neighborhood characteristics, and economic indicators. Traditional methods of property valuation relied on manual assessment by real estate agents or appraisers, but machine learning offers a data-driven approach that can yield more objective and precise predictions. The key advantage of using machine learning for house price prediction is its ability to identify complex patterns and relationships within large datasets. By training a model on historical housing data, we can uncover hidden correlations and gain insights into how different features impact property values. This enables us to build predictive models capable of estimating house prices accurately for new and unseen data. Various machine learning techniques can be applied to house price prediction, including regression algorithms like linear regression, decision trees, random forests, support vector regression, and neural networks. These models are trained using labeled datasets, where each data point consists of input features (e.g., number of rooms, location) and the corresponding house price.

Once the model is trained, it can be deployed to make predictions on new instances by inputting the relevant features. This empowers potential home buyers, sellers, and investors to make informed decisions based on estimated property values. Additionally, real estate agents and financial institutions can leverage these predictions to optimize their operations, improve customer service, and mitigate risks.

However, it's important to note that house price prediction using machine learning is not without its challenges. Obtaining accurate and comprehensive datasets, dealing with missing or inconsistent data, handling outliers, and selecting the most appropriate features are some of the hurdles that need to be addressed for reliable predictions.

**CHAPTER 2**

**EXISTING METHOD**

Existing methods for house price prediction using machine learning typically involve the following steps:

**Data Collection:** The first step is to gather a comprehensive dataset that includes relevant features and corresponding house prices. This data can be obtained from various sources, such as real estate databases, government records, or online platforms.

**Data Preprocessing:** Once the dataset is collected, it needs to be preprocessed to ensure its quality and suitability for training machine learning models. This includes handling missing values, dealing with outliers, normalizing or scaling features, and encoding categorical variables.

**Feature Selection/Engineering:** Feature selection involves identifying the most relevant features that significantly contribute to predicting house prices. This step helps reduce dimensionality and improve the model's efficiency. Feature engineering involves transforming or creating new features from the existing ones to capture more meaningful information.

**Model Selection:** There are several regression algorithms available for house price prediction, and the choice of model depends on the dataset size, complexity, and performance requirements. Commonly used models include linear regression, decision trees, random forests, support vector regression, and neural networks.

**Model Training:** The selected model is trained using the preprocessed dataset. The dataset is divided into a training set and a validation set to evaluate the model's performance during training. The model learns the underlying patterns and relationships between the input features and the house prices.

**Model Evaluation:** After training, the model's performance is evaluated using various metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared value. These metrics quantify how well the model predicts house prices compared to the actual values.

**Hyper-parameter Tuning:** Many machine learning models have hyperparameters that control their behavior and performance. Hyperparameter tuning involves optimizing these parameters to improve the model's accuracy. Techniques like grid search, random search, or Bayesian optimization can be employed for this purpose.

**Model Deployment:** Once the model is trained and evaluated, it can be deployed to make predictions on new, unseen data. This allows users to input the relevant features of a house and obtain an estimated price.

**Monitoring and Updating:** The deployed model should be continuously monitored to ensure its accuracy and reliability. As new data becomes available, the model can be updated or retrained periodically to incorporate the latest information and improve its predictive performance.

These existing methods provide a framework for utilizing machine learning algorithms to predict house prices accurately. However, the specific implementation and choice of techniques may vary depending on the dataset, resources, and specific requirements of the application.

### Industry Profile

The industry of house price prediction using machine learning has gained significant traction in recent years, transforming the real estate and finance sectors. This emerging field combines the expertise of data science and domain knowledge to develop accurate models that estimate property values. Let's explore the industry profile for house price prediction using machine learning.

**Market Overview:**

The real estate industry has always been driven by the demand for accurate property valuation. Traditional methods of valuing houses relied on subjective assessments, which often led to inconsistencies and inefficiencies. Machine learning-based house price prediction offers a data-driven and objective approach, revolutionizing the way properties are valued and facilitating more informed decision-making.

**Key Players:**

**Real Estate Agencies:** Real estate agencies have recognized the potential of machine learning in predicting house prices and have started incorporating such models into their operations. These agencies leverage machine learning algorithms to provide accurate property valuations to their clients, enhance marketing strategies, and optimize their investment decisions.

**PropTech Companies:** PropTech (Property Technology) companies specialize in developing innovative technological solutions for the real estate industry. Many PropTech startups have emerged, focusing specifically on house price prediction using machine learning. They create platforms and tools that enable real estate professionals, buyers, and sellers to access reliable predictions and market insights.

**Financial Institutions:** Banks, mortgage lenders, and insurance companies heavily rely on accurate house price predictions to assess risks, determine loan amounts, and make informed investment decisions. They leverage machine learning models to analyze vast amounts of data and estimate property values, ensuring their financial operations are optimized and risk is mitigated.

**Data Analytics and Consulting Firms:** Data analytics and consulting firms play a crucial role in the house price prediction industry. They provide expertise in data analysis, model development, and deployment, helping businesses implement machine learning-based solutions effectively. These firms assist in gathering relevant data, preprocessing and cleaning it, and building robust predictive models.

**Technological Advancements:**

The rapid advancement of machine learning techniques and computational power has significantly contributed to the growth of house price prediction using ML. State-of-the-art algorithms such as deep learning neural networks, ensemble methods like random forests and gradient boosting, and regression models have become widely adopted in the industry. Additionally, advancements in cloud computing have facilitated scalable and cost-effective model training and deployment.

**Data Availability and Integration:**

The availability and integration of diverse datasets have been instrumental in improving the accuracy of house price prediction models. Besides traditional property features, datasets now include a wide range of factors, such as neighborhood characteristics, economic indicators, local amenities, and geospatial information. Integration of these datasets enhances the predictive power of the models and enables more comprehensive property valuation.

**Challenges:**

While the industry of house price prediction using machine learning has shown promising growth, it faces several challenges. Some of the key challenges include obtaining reliable and comprehensive datasets, dealing with missing or inconsistent data, addressing biases in the data, handling outliers and extreme market conditions, and ensuring model interpretability and transparency.

**Future Outlook:**

The future of house price prediction using machine learning looks promising. As technology continues to advance, machine learning models will become more accurate and robust. Integration with other emerging technologies, such as augmented reality (AR) and virtual reality (VR), can enhance the visualization and immersive experience of property valuation. Furthermore, as more real estate transactions move online, the demand for accurate and instant house price predictions will increase, driving further innovation in the industry.

**CHAPTER 3**

**RESEARCH METHODOLOGY**

Research methodology for house price prediction using machine learning typically involves the following steps:

**Problem Definition:** Clearly define the research problem, which is to predict house prices using machine learning techniques.

**Data Collection:** Gather a comprehensive dataset that includes various features of houses (e.g., size, location, number of bedrooms, amenities, etc.) along with their corresponding prices. The data can be obtained from real estate websites, public datasets, or by collecting it directly through surveys or other means.

**Data Preprocessing:** Clean and preprocess the collected data to ensure its quality and suitability for the machine learning models. This step may involve handling missing values, removing outliers, normalizing or scaling features, and encoding categorical variables.

**Feature Selection/Engineering:** Analyze the collected data and select relevant features that are most likely to influence house prices. Additionally, create new features through feature engineering techniques to enhance the predictive power of the models. For example, you could derive a feature like the price per square foot from the original features.

**Splitting the Data:** Divide the dataset into two subsets: a training set and a testing/validation set. The training set will be used to train the machine learning models, while the testing/validation set will be used to evaluate their performance.

**Model Selection:** Choose suitable machine learning algorithms for house price prediction. Commonly used models include linear regression, decision trees, random forests, support vector machines (SVM), and neural networks. The selection of models depends on the characteristics of the dataset and the specific requirements of the problem.

**Model Training:** Train the selected machine learning models using the training set. The models will learn the underlying patterns and relationships between the features and the house prices during this phase.

**Model Evaluation:** Evaluate the trained models using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared. These metrics will assess how well the models perform in predicting house prices.

**Hyperparameter Tuning:** Fine-tune the hyperparameters of the selected models to optimize their performance. This step involves adjusting the parameters that are not learned during training, such as the learning rate, regularization strength, or maximum tree depth.

**Model Validation:** Validate the performance of the tuned models using the testing/validation set. This step provides an unbiased assessment of how well the models generalize to unseen data.

Model Deployment: Once a satisfactory model is selected, deploy it in a production environment, making it available for real-time house price predictions.

**Continuous Monitoring and Improvement:** Monitor the performance of the deployed model and gather feedback. Iteratively refine the model or develop new models to improve accuracy and adapt to changing trends in the housing market.

**3.1 Title of the Study**

"Predicting House Prices: A Machine Learning Approach for Real Estate Market Analysis"

**3.2 Statement of the problem**

The problem at hand is to develop a machine learning model that can accurately predict house prices based on a set of relevant features. The objective is to provide home buyers, sellers, and real estate agents with a reliable tool to estimate the market value of residential properties.

The model will be trained on a dataset consisting of historical housing data, which includes attributes such as the size of the property, number of bedrooms and bathrooms, location, proximity to amenities, and other relevant factors that influence property prices. The dataset will also include the actual sale prices of the houses.

The goal is to create a regression model that can take these features as input and accurately predict the corresponding house prices. The model should be able to handle different types of input data, including numerical and categorical variables, and account for potential non-linear relationships between the features and the target variable.

The accuracy and performance of the model will be evaluated using appropriate metrics such as mean squared error (MSE) or root mean squared error (RMSE). The ultimate aim is to develop a model that can generalize well to unseen data and provide reliable price predictions for new houses.

This problem is significant as accurate house price predictions can benefit various stakeholders in the real estate industry. Home buyers can use the model to assess whether a listed price is fair and make informed decisions. Sellers can determine an appropriate listing price based on market trends and property attributes. Real estate agents can leverage the model to provide reliable advice to clients and streamline the buying and selling process.

By addressing this problem, we aim to provide a valuable tool that improves transparency and efficiency in the real estate market, empowering individuals with data-driven insights for making informed decisions about residential property transactions.

**3.3 Objectives of the Study**

The objectives of a study on house price prediction using machine learning can vary depending on the specific context and goals of the research.

the specific objectives of a study on house price prediction using machine learning will depend on the research context, available data, and the needs of stakeholders involved in the housing market.

**3.4 Research Design**

Designing a research study for house price prediction using machine learning involves several steps. Remember that the specific details of the research design may vary depending on your specific research objectives, the available data, and the machine learning techniques you choose to employ.

**3.5 Hypotheses**

When it comes to predicting house prices using machine learning, several hypotheses can be explored. Here are some common hypotheses:

**Linear Relationship:** The hypothesis assumes that there is a linear relationship between the input features (e.g., number of bedrooms, square footage, location) and the target variable (house price). It assumes that the relationship can be adequately captured by a linear regression model.

**Non-Linear Relationship:** This hypothesis suggests that the relationship between the input features and house prices is non-linear. It implies that more complex models, such as decision trees, random forests, or neural networks, might be better suited for capturing the underlying patterns.

**Feature Importance:** The hypothesis posits that certain features have a stronger influence on house prices compared to others. It suggests that by identifying and focusing on the most important features, the predictive performance of the model can be improved.

**Neighborhood Effect:** This hypothesis suggests that the location and neighborhood characteristics play a significant role in determining house prices. It implies that including features related to neighborhood demographics, amenities, crime rates, and proximity to schools or transportation can enhance the predictive power of the model.

**Temporal Factors:** This hypothesis considers the impact of time-related factors on house prices. It assumes that trends, seasonality, or macroeconomic indicators can influence housing markets. Incorporating temporal features, such as year, month, or historical price data, can help capture these factors.

**Interaction Effects:** This hypothesis proposes that the interaction between certain features may have a substantial impact on house prices. For example, the combination of the number of bedrooms and bathrooms might be more influential than considering them separately. Including interaction terms or using more sophisticated models capable of capturing interactions, like polynomial regression or deep learning models, can test this hypothesis.

**Data Quality:** This hypothesis suggests that the quality and completeness of the dataset have a direct impact on prediction accuracy. It assumes that ensuring data cleanliness, handling missing values, and addressing outliers can lead to improved model performance.

These hypotheses provide a starting point for exploring and developing machine learning models for house price prediction. The choice of hypothesis to investigate depends on the specific dataset, available features, and the goals of the prediction task.

**3.6 Sampling Technique**

When it comes to predicting house prices using machine learning, sampling techniques play a crucial role in creating a representative and unbiased dataset. Here are some commonly used sampling techniques:

**Simple Random Sampling:** This technique involves randomly selecting a subset of data points from the entire dataset. Each data point has an equal chance of being selected. Simple random sampling is easy to implement but may not be efficient for large datasets.

**Stratified Sampling:** Stratified sampling involves dividing the dataset into homogeneous groups called strata and then selecting a proportional number of samples from each stratum. For house price prediction, you could divide the dataset based on relevant features such as location, property type, or square footage to ensure that each stratum is represented in the sample.

**Cluster Sampling:** Cluster sampling involves dividing the dataset into clusters, such as geographical regions, and randomly selecting entire clusters to include in the sample. This technique is useful when the dataset is geographically distributed, and it helps capture regional variations in house prices.

**Oversampling and Under sampling:** These techniques are used when dealing with imbalanced datasets, where one class (e.g., expensive houses) is underrepresented compared to another (e.g., affordable houses). Oversampling involves randomly duplicating minority class samples, while undersampling involves randomly removing majority class samples. These techniques can help balance the dataset and improve prediction performance.

**Systematic Sampling:** Systematic sampling involves selecting data points at regular intervals from an ordered list. For house price prediction, you could order the dataset based on a relevant feature such as date of sale and then select every nth data point to create the sample.

**Cross-validation:** While not a sampling technique in the traditional sense, cross-validation is commonly used in machine learning for model evaluation. It involves dividing the dataset into multiple folds and using each fold as both training and validation data. This technique helps assess the model's performance on different subsets of the data.

The choice of sampling technique depends on the characteristics of your dataset, the specific problem at hand, and the goals of your analysis. It is essential to carefully consider the advantages and limitations of each technique and select the one that best suits your needs.

**CHAPTER 4**

**PROPOSED SYSTEM**

To develop a system for house price prediction using machine learning, you can follow these general steps:

**Data Collection:** Gather a comprehensive dataset of housing information, including features such as the number of bedrooms, bathrooms, square footage, location, amenities, and historical sales prices. You can obtain this data from real estate websites, public records, or other reliable sources.

**Data Preprocessing:** Clean and preprocess the collected data to ensure its quality and compatibility with the machine learning algorithms. Handle missing values, outliers, and perform feature engineering if necessary. This step may also involve transforming categorical variables into numerical representations through techniques like one-hot encoding.

**Feature Selection:** Analyze the dataset to identify the most relevant features that contribute significantly to the house prices. You can employ techniques like correlation analysis, feature importance, or dimensionality reduction algorithms to select the optimal set of features for your model.

**Model Selection:** Choose an appropriate machine learning algorithm for your prediction task. Regression algorithms such as linear regression, decision trees, random forests, or gradient boosting can be effective for house price prediction. Consider the strengths and weaknesses of each algorithm and select the one that best suits your dataset and requirements.

**Model Training:** Split the preprocessed data into training and testing sets. Use the training set to train the chosen machine learning model. During training, the model will learn the patterns and relationships between the features and target variable (house prices).

**Model Evaluation:** Evaluate the trained model's performance using suitable evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared. Assess how well the model generalizes to unseen data by using the testing set.

**Model Optimization:** Fine-tune the model by adjusting hyperparameters to optimize its performance. This can be done through techniques like grid search or randomized search, where different combinations of hyperparameters are evaluated to find the optimal configuration.

**Deployment:** Once you are satisfied with the model's performance, deploy it to make predictions on new, unseen data. Develop a user-friendly interface or API that takes input features (e.g., number of bedrooms, square footage) and returns the predicted house price based on the trained model.

**Maintenance and Updates:** Regularly monitor and maintain the deployed system, retraining the model periodically with new data to keep it up to date. Stay informed about new developments in the field of machine learning and consider incorporating improvements or newer algorithms as they become available.

Remember, the success of your house price prediction system depends on the quality of the data, the choice of features, and the selection and fine-tuning of the machine learning model.

**4.1 Statistical Tools Used**

There are several statistical tools that can be used for house price prediction using machine learning. Here are some commonly used tools:

**Linear Regression:** Linear regression is a simple and widely used statistical tool for predicting house prices. It assumes a linear relationship between the independent variables (such as square footage, number of bedrooms, etc.) and the dependent variable (house price). The model estimates the coefficients of the independent variables to make predictions.

**Multiple Regression:** Multiple regression extends linear regression by allowing multiple independent variables to be used in the prediction model. This tool is useful when there are multiple factors affecting house prices, such as location, amenities, and neighborhood characteristics.

**Decision Trees:** Decision trees are non-linear statistical tools that can be used for house price prediction. They partition the data based on different attributes and create a tree-like structure to make predictions. Decision trees can capture complex interactions between variables and handle both categorical and numerical data.

**Random Forests:** Random forests are an ensemble learning method that combines multiple decision trees to make predictions. They reduce overfitting and improve accuracy by aggregating the predictions of multiple individual trees. Random forests are robust and can handle large datasets with high-dimensional features.

**Support Vector Regression (SVR):** SVR is a statistical tool based on support vector machines that can be used for regression tasks, including house price prediction. SVR finds the best-fitting hyperplane in a high-dimensional space to minimize the error between the predicted and actual house prices.

**Gradient Boosting:** Gradient boosting is another ensemble learning method that combines weak prediction models, such as decision trees, into a strong predictive model. It iteratively improves the model by minimizing a loss function and adjusting the weights of the individual models. Gradient boosting is known for its high predictive accuracy.

**Neural Networks:** Neural networks, specifically deep learning models, can also be used for house price prediction. They are powerful tools for capturing complex patterns in the data and can handle large datasets. Neural networks can be trained using various architectures, such as feedforward neural networks, convolutional neural networks (CNNs), or recurrent neural networks (RNNs).

These are just a few examples of the statistical tools used for house price prediction using machine learning. The choice of tool depends on the specific requirements of the problem, the available data, and the desired level of accuracy and interpretability.

**4.2 Scope of the Study**

The scope of a study on house price prediction using machine learning can vary depending on the specific objectives and constraints of the research. It's important to note that the specific scope of the study may be adjusted based on available resources, time constraints, and the objectives of the research.

**4.3 Limitations of the Study**

When conducting a study on house price prediction using machine learning, there can be several limitations that researchers should acknowledge. It is essential to acknowledge these limitations and potential sources of error when conducting a study on house price prediction using machine learning. Researchers should be cautious about the assumptions made, validate the model's performance on unseen data, and consider external factors that may impact the predictions.

**4.4 Data Collection and Analysis**

Data Collection and Analysis for House Price Prediction using Machine Learning involves several steps. Remember that this is a general outline, and specific steps and techniques may vary based on the dataset, problem statement, and chosen machine learning algorithms.

**4.5 Findings**

House price prediction using machine learning has been a popular research area, and several findings and approaches have emerged in recent years. It's important to note that the effectiveness of these findings may vary depending on the specific dataset, location, and context. Researchers and practitioners continue to explore new techniques and approaches to improve house price prediction using machine learning.

**4.6 Suggestions**

When it comes to house price prediction using machine learning, there are several approaches you can consider. Remember, the success of your house price prediction model depends on the quality and relevance of your data, as well as the choice of the appropriate machine learning techniques and feature engineering methods.

# CHAPTER 5 EXPERIMENT ANALYSIS

**5.1 System Configuration**

This project can run on commodity hardware. We ran entire project on an Intel I5 processor with 8 GB Ram, 2 GB Nvidia Graphic Processor, it also has 2 cores which runs at 1.7 GHz, 2.1 GHz respectively. First part of the is training phase which takes 10-15 mins of time and the second part is testing part which only takes few seconds to make predictions and calculate accuracy.

### Hardware Requirements:

* + - * RAM: 4 GB
      * Storage: 500 GB
      * CPU: 2 GHz or faster
      * Architecture: 32-bit or 64-bit

### Software requirements

* + - * Python 3.10.0 in Google Colab is used for data pre-processing, model training and prediction.
      * Operating System: windows 10 and above or Linux based OS or MAC OS.

##### 5.3 Functional requirements

Functional requirements describe what the software should do (the functions). Think about the core operations.

Because the “functions” are established before development, functional requirements should be written in the future tense. In developing the software for **HOUSE PRICE PREDICTION,** some of the functional requirements could include:

* The software shall accept the tw\_spydata\_raw.csv dataset as input.
* The software should shall do pre-processing (like verifying for missing data values) on input for model training.
* The software shall use MULTIPLE LINEAR REGRESSION as main component of the software.
* It processes the given input data by producing the most possible outcomes of a House price prediction Notice that each requirement is directly related to what we expect the software to do. They represent some of the core functions.

##### 5.4 Non-Functional requirements

Product properties

* Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of **house price prediction** software, for any kind of profit trader and other stakeholders in profit prediction of 50 companies.
* Efficiency: maintaining the possible highest accuracy in the 50 companies in shortest time with available data.

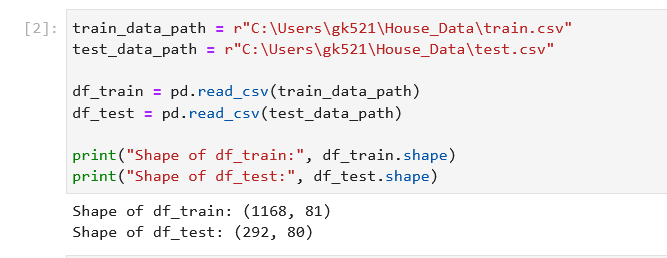
Performance: It is a quality attribute of the profit prediction software that describes the responsiveness to various user interactions with it.

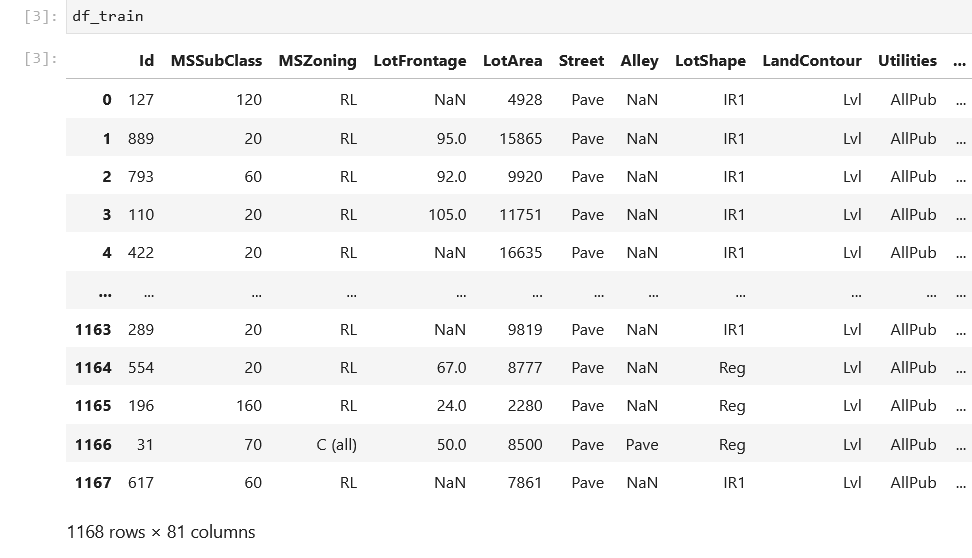
### Sample code with Output

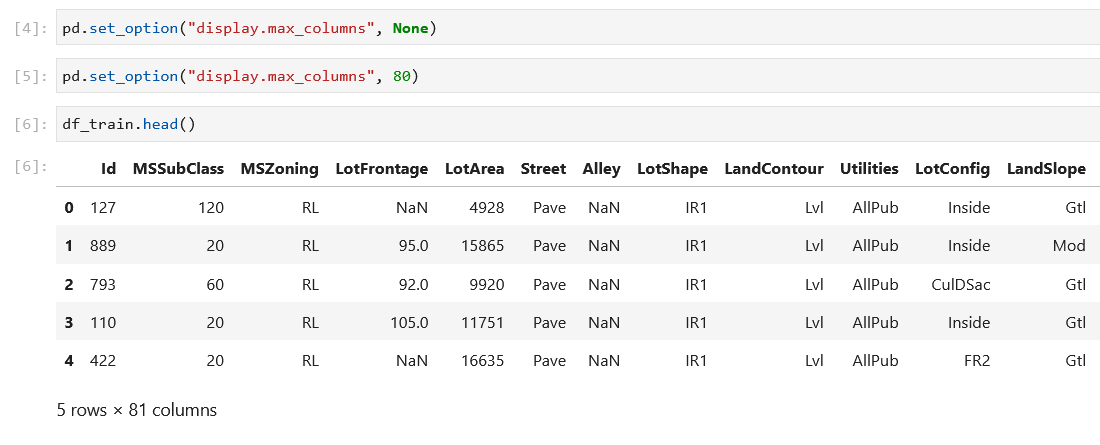
**Step 1: Import the necessary libraries**

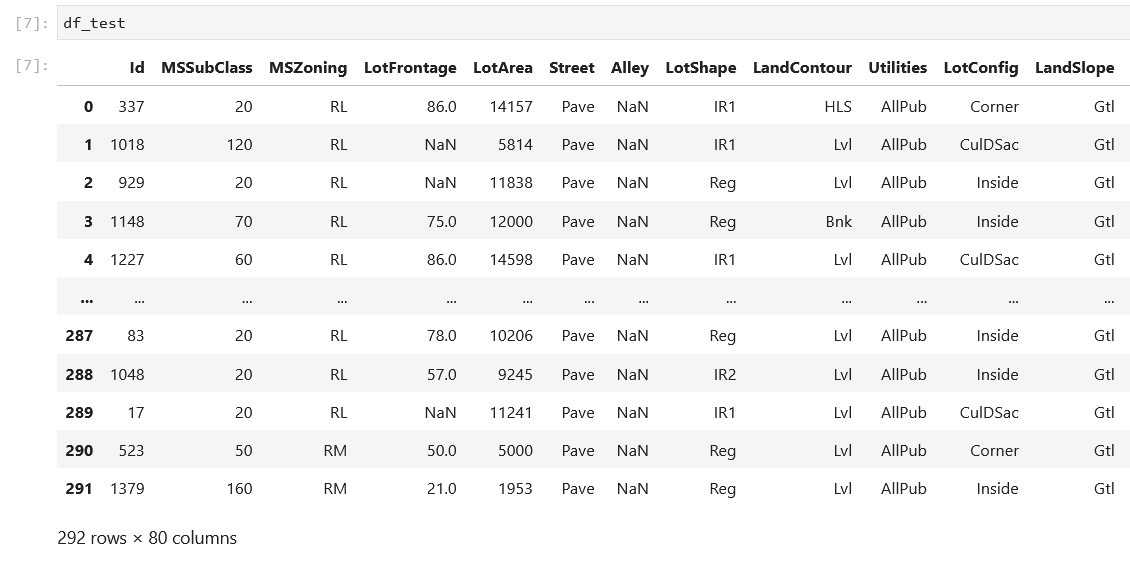


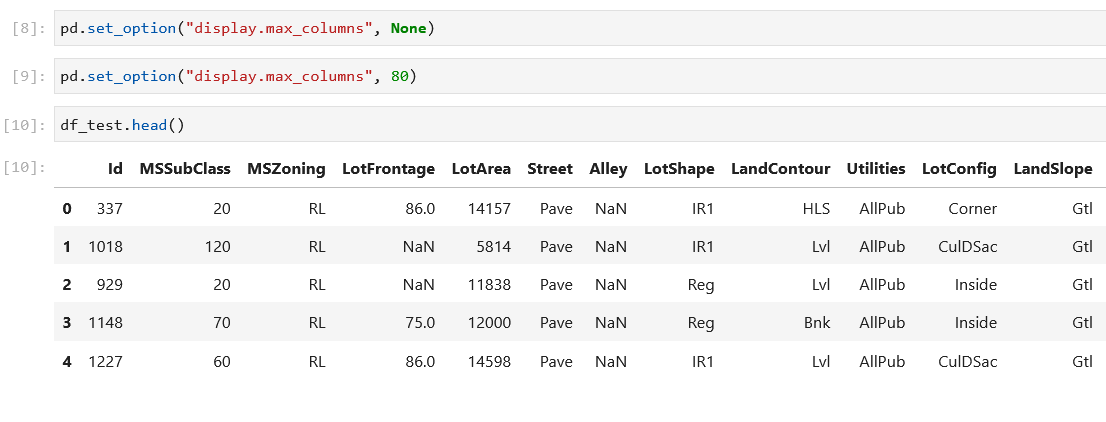
**Step 2: Load and explore the dataset.**

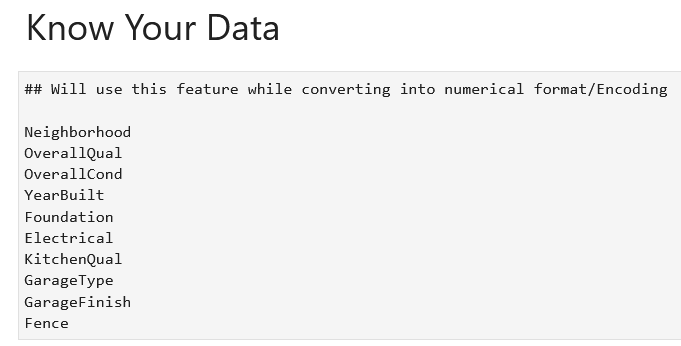
****

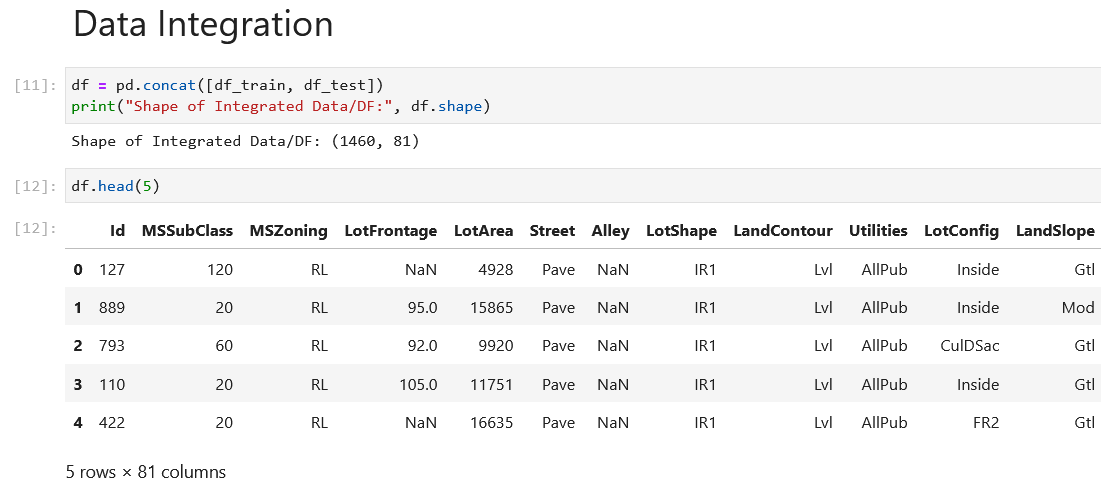
****

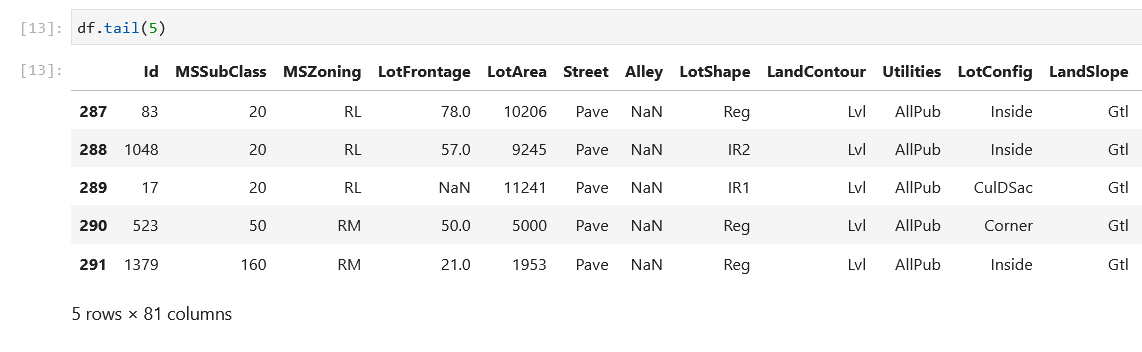
****

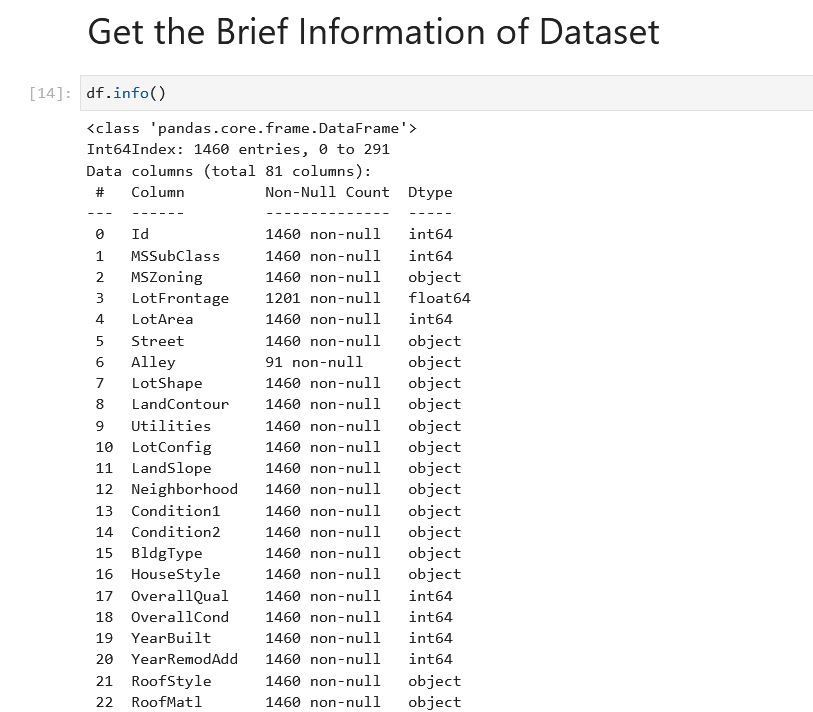
****

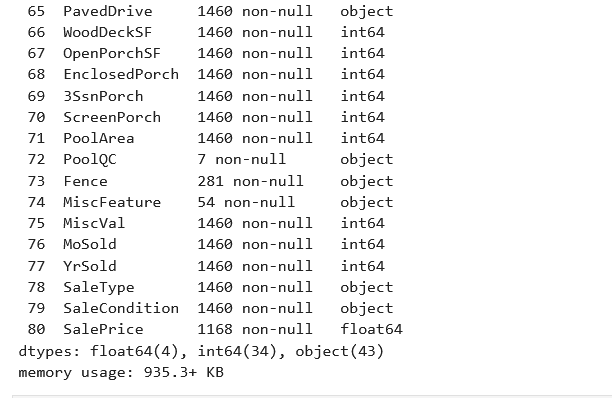
****

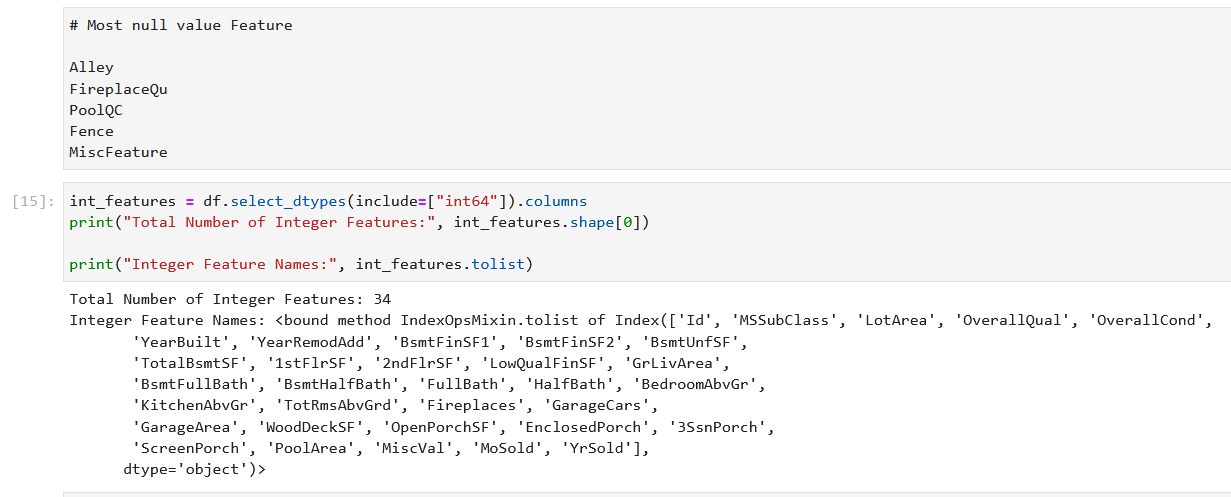
****

****

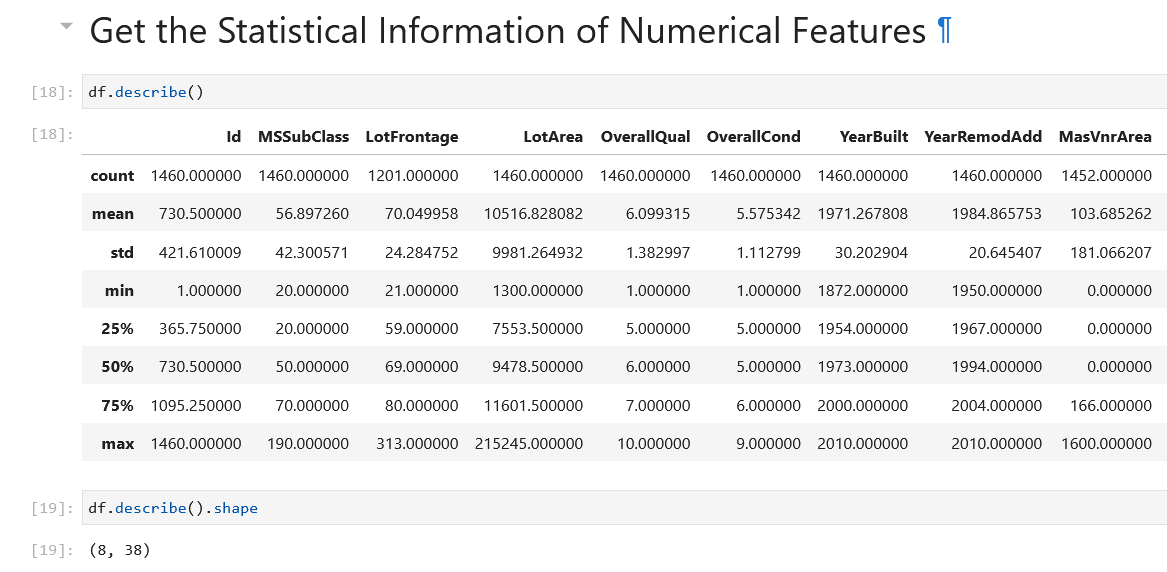
****

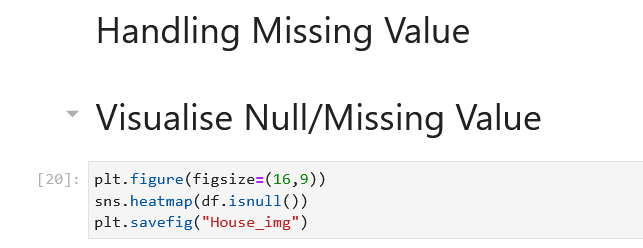
****

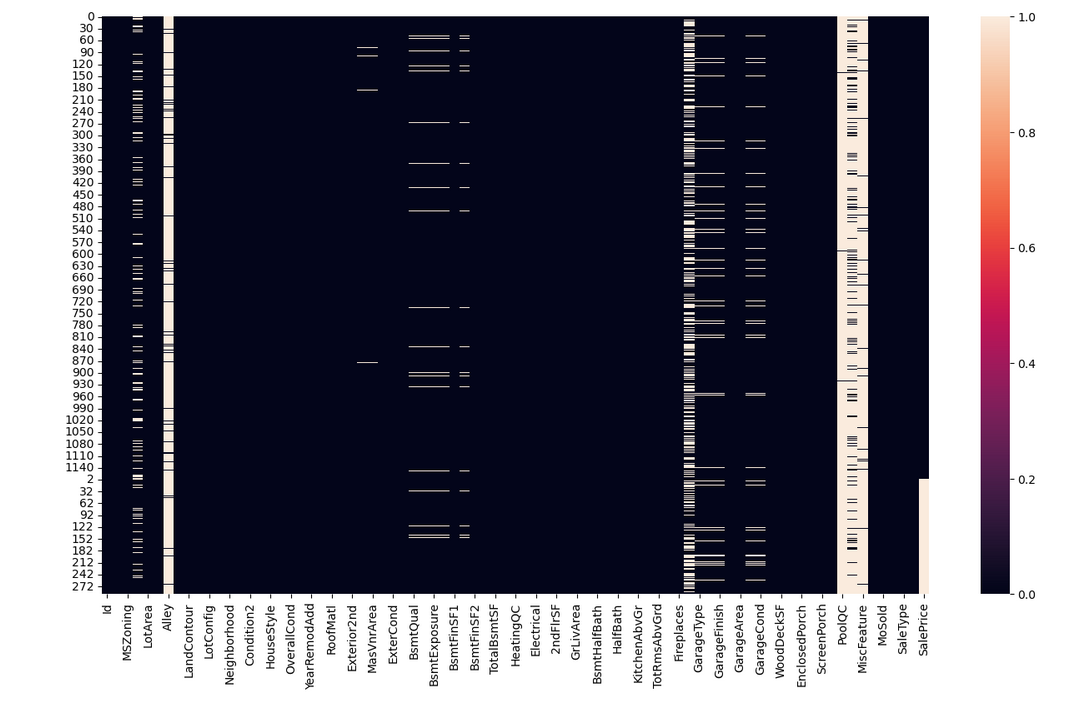
****

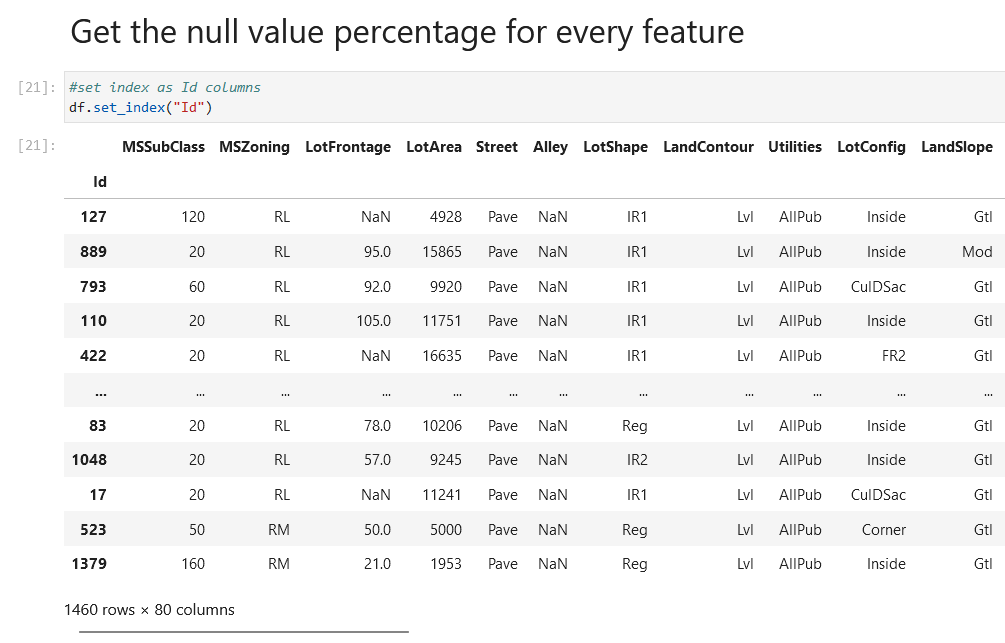
****

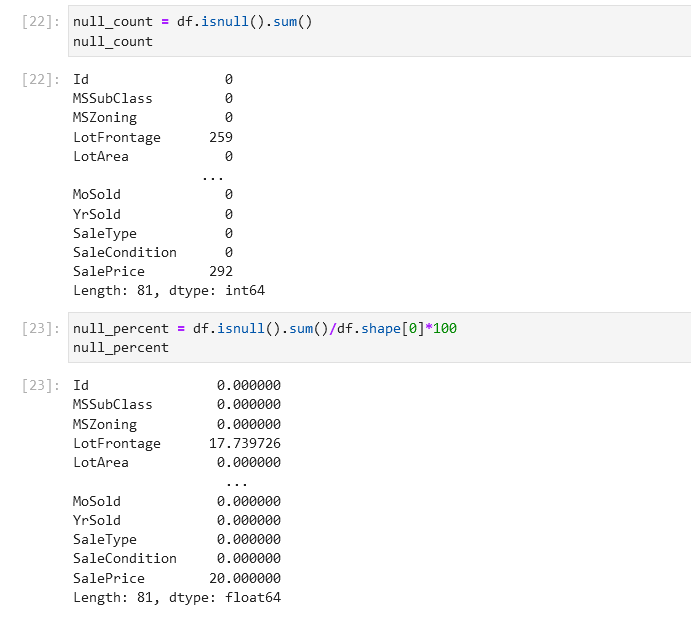
****

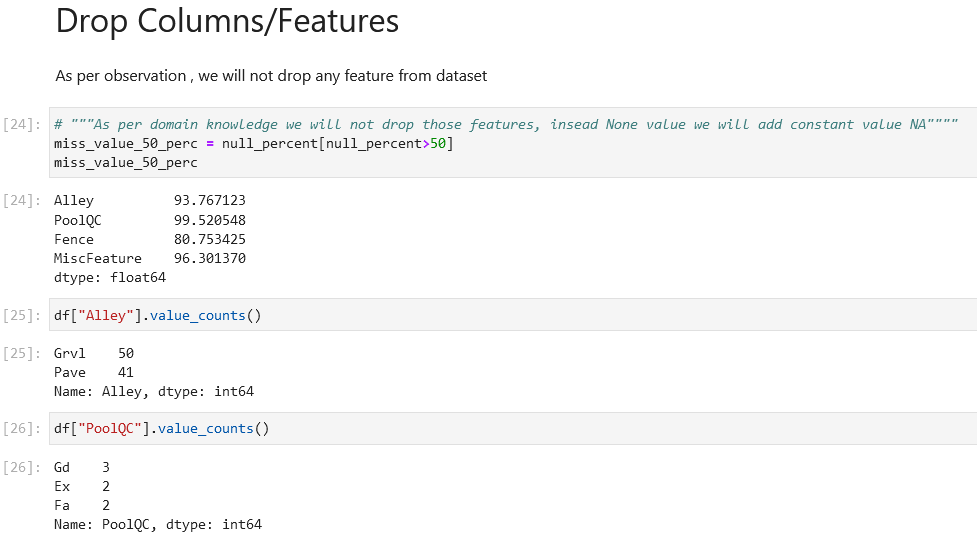


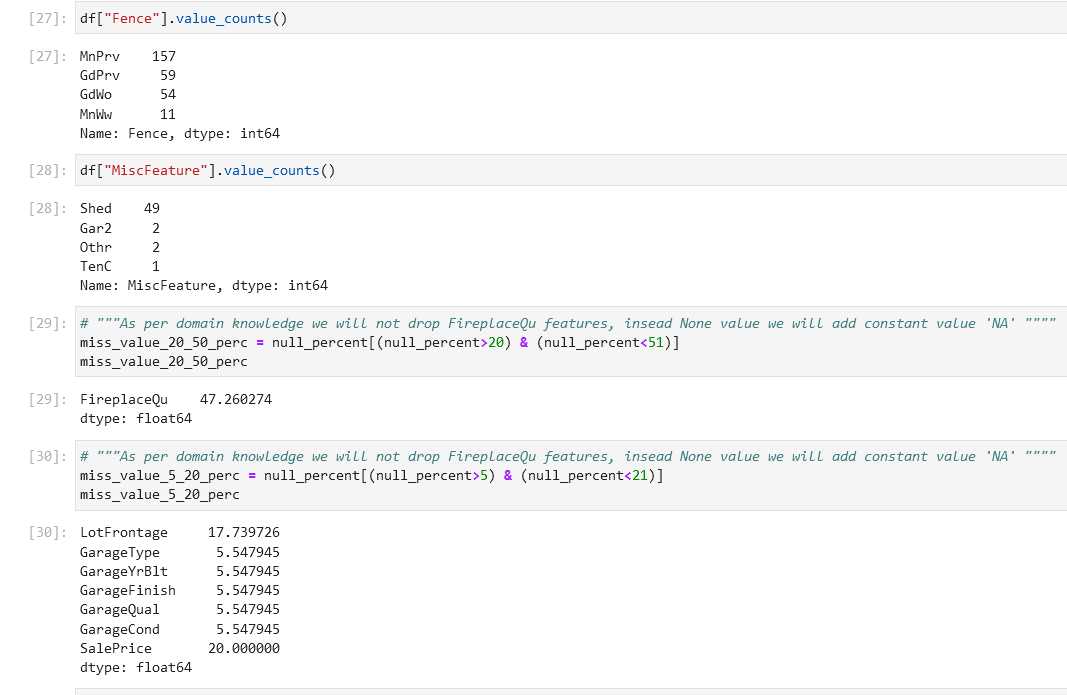


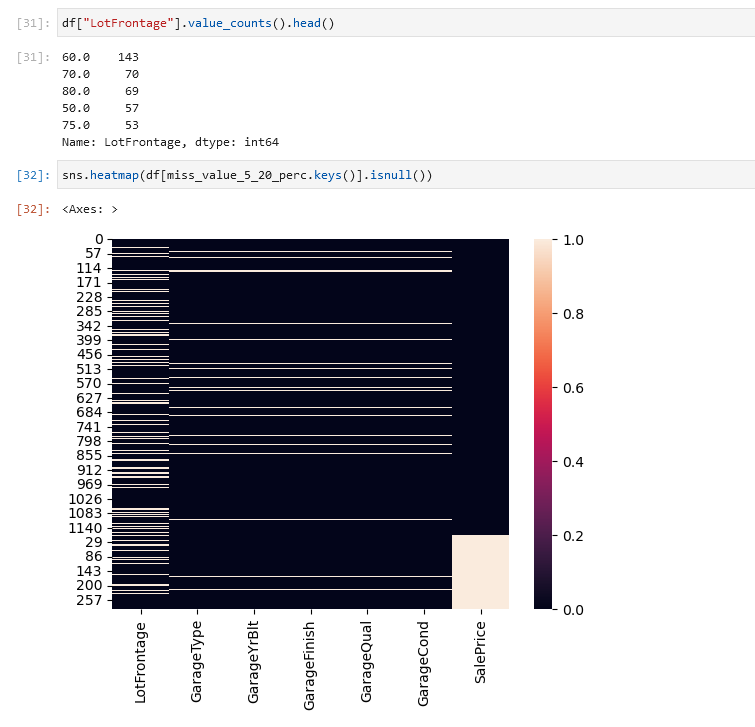


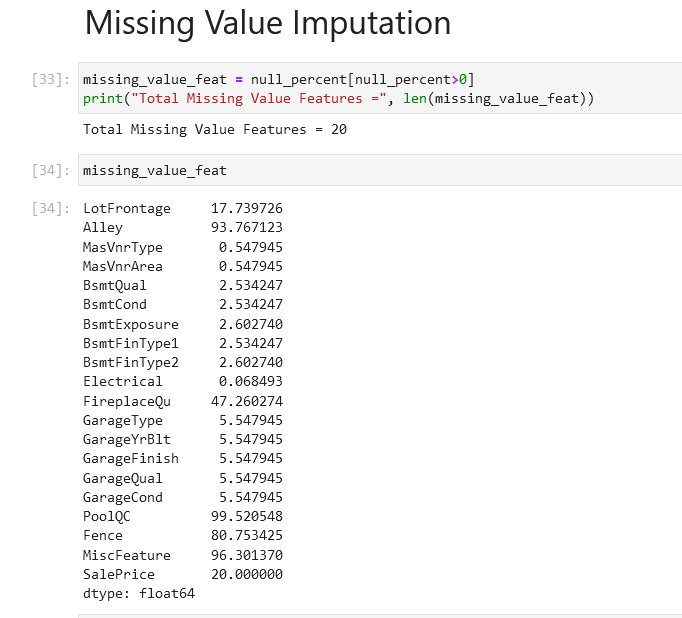


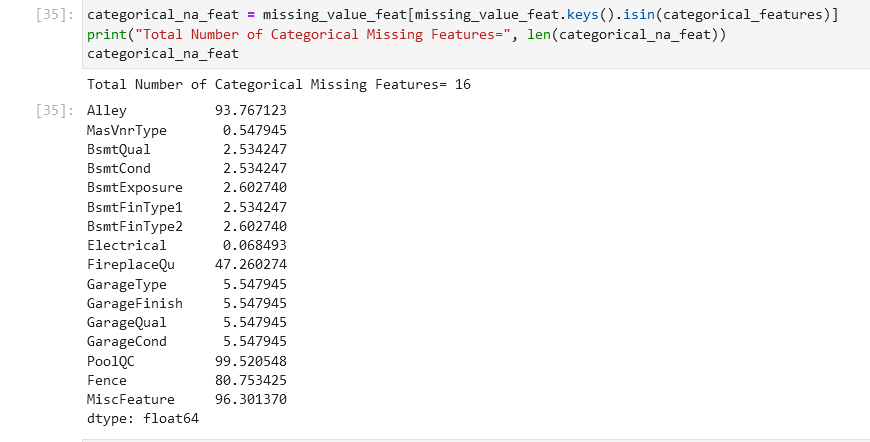


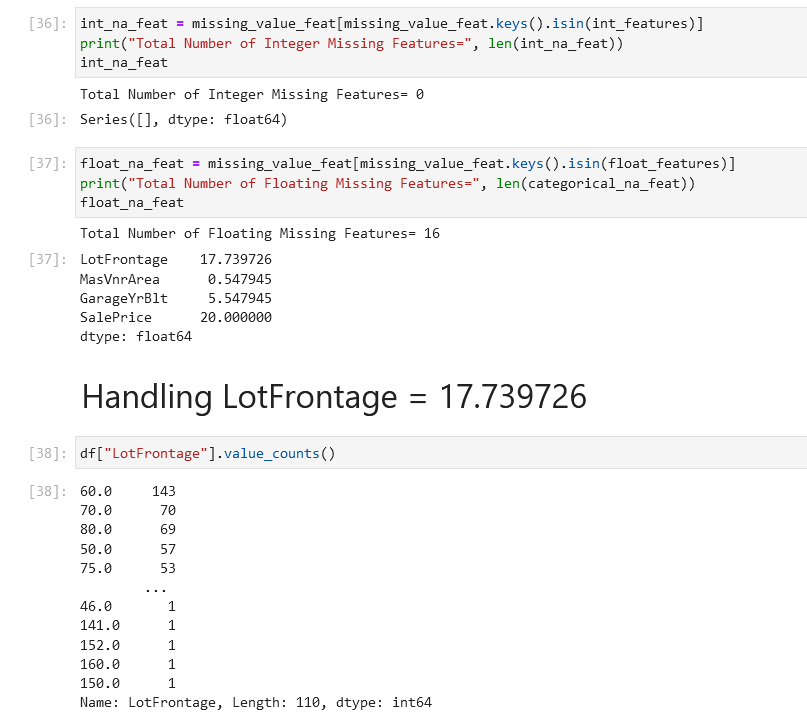


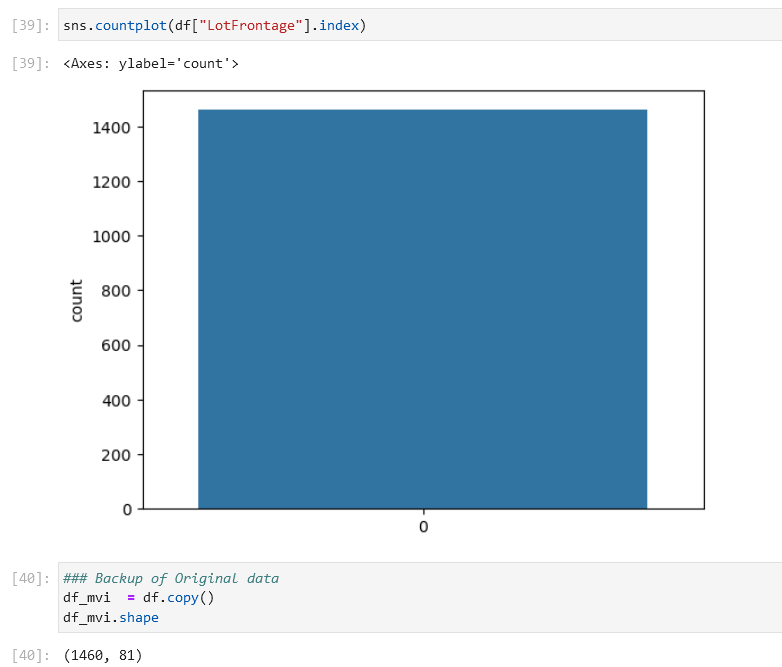


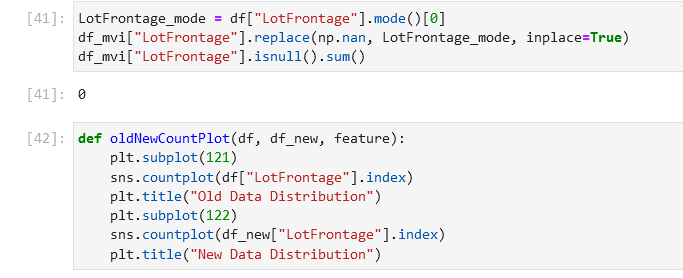


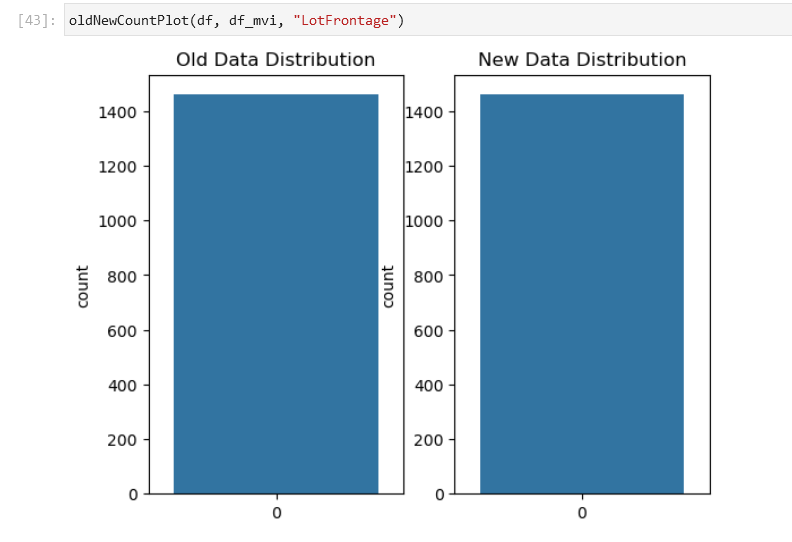




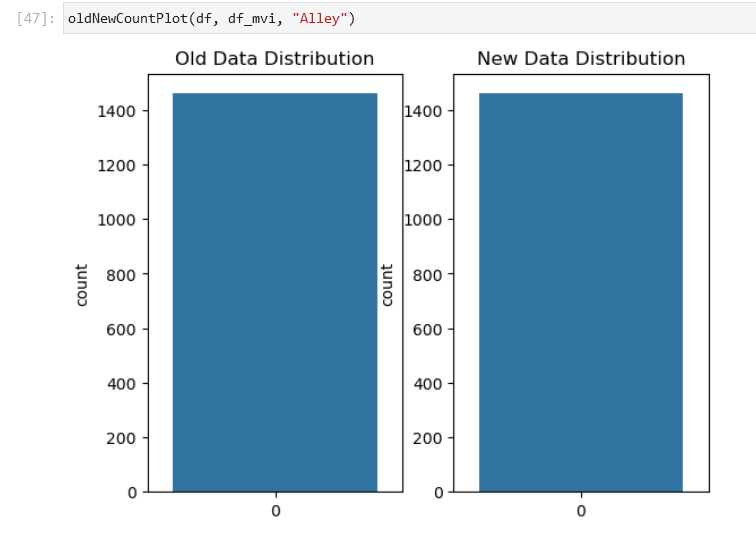




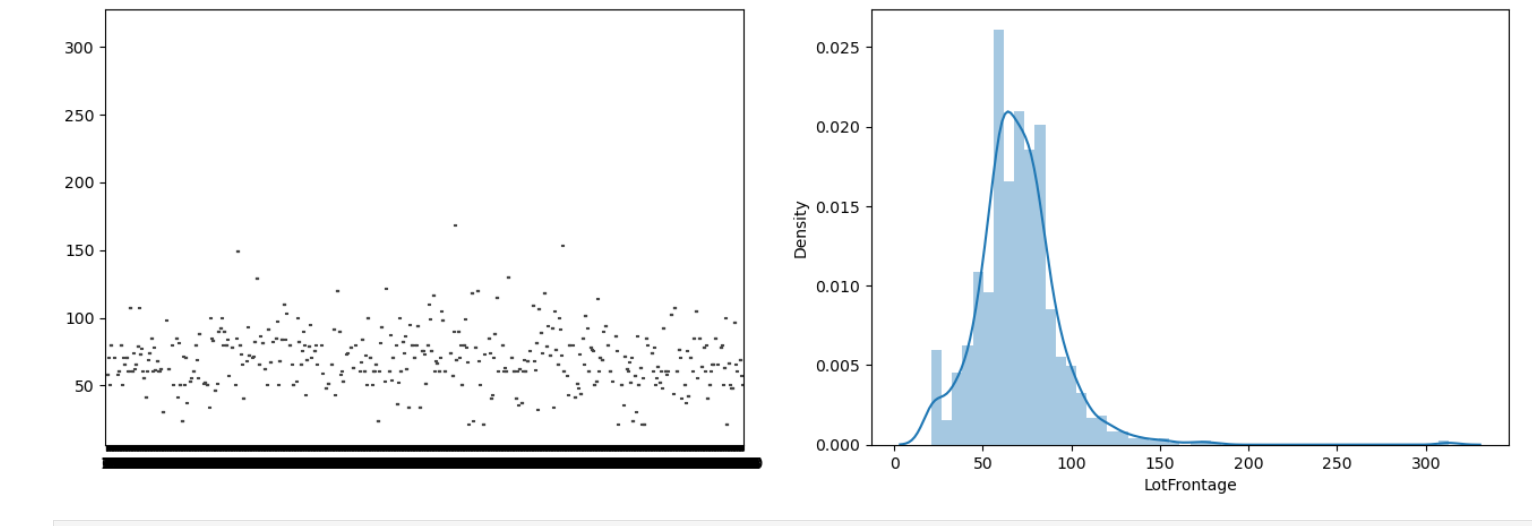






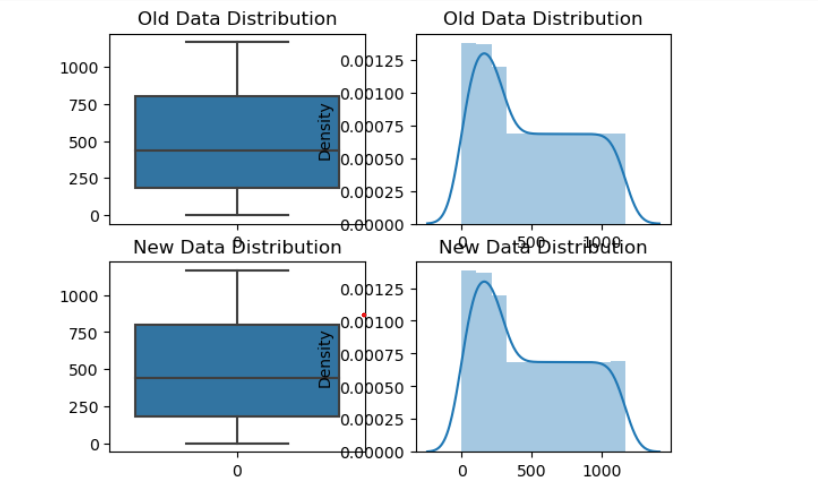


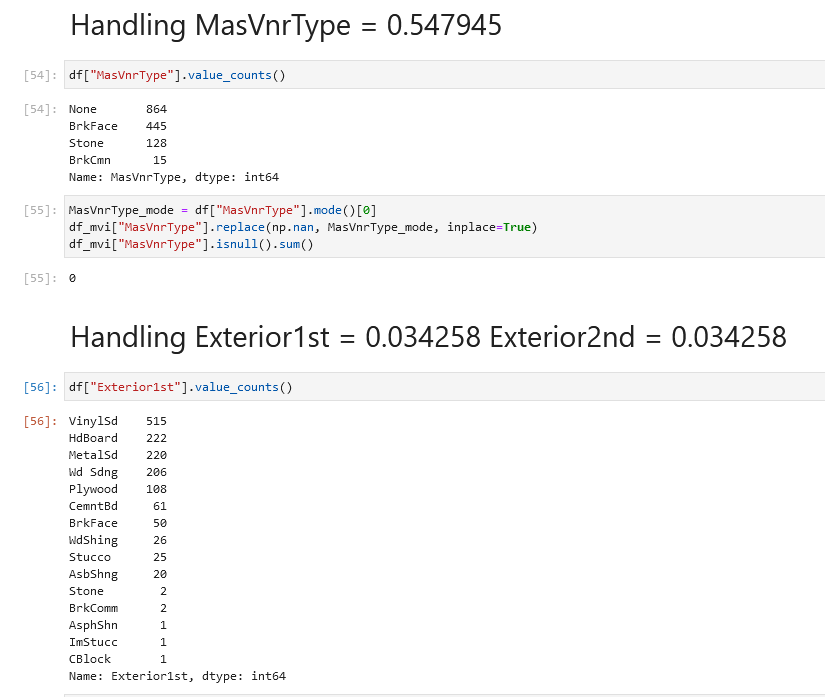


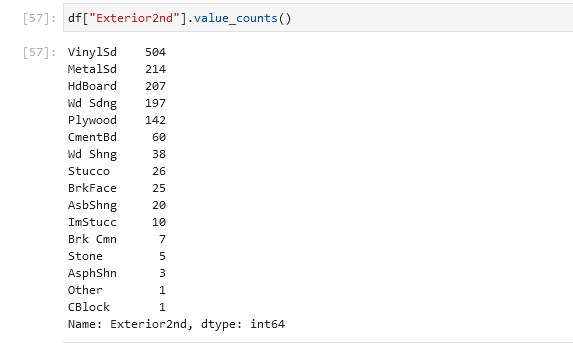




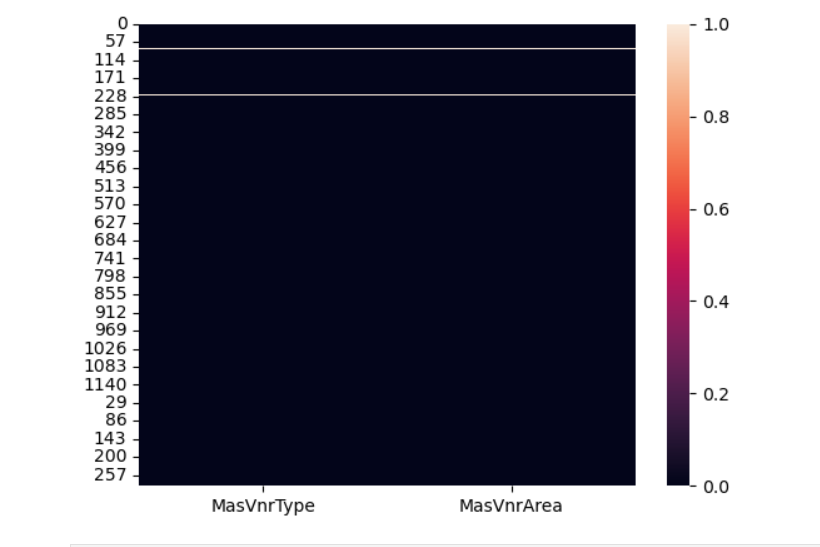


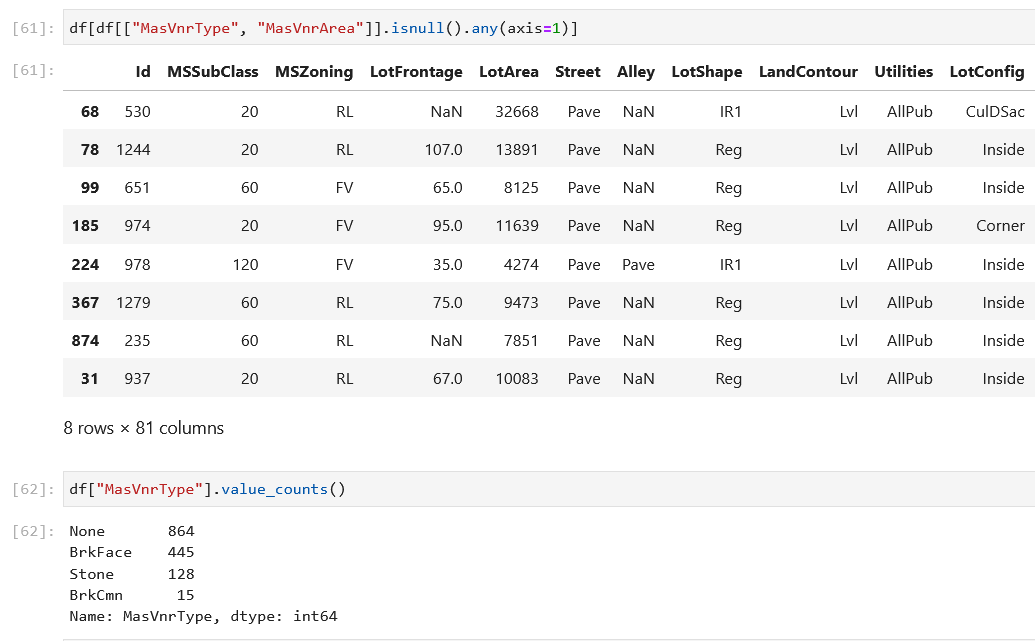


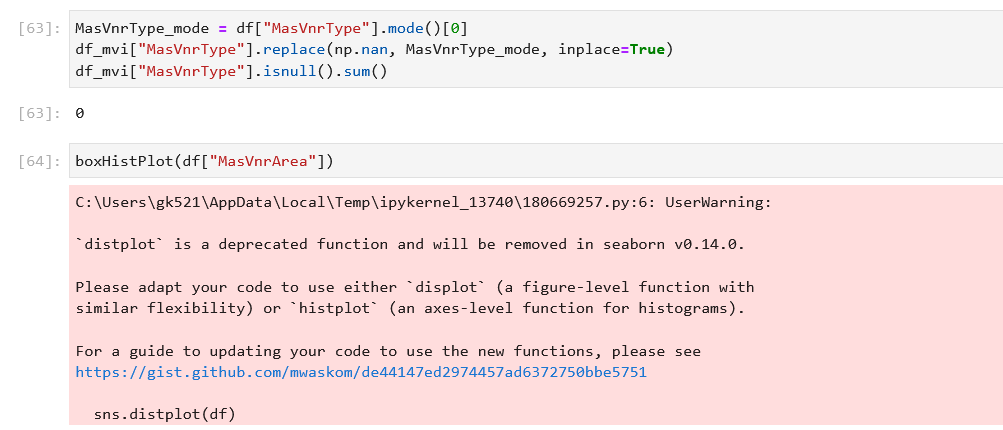


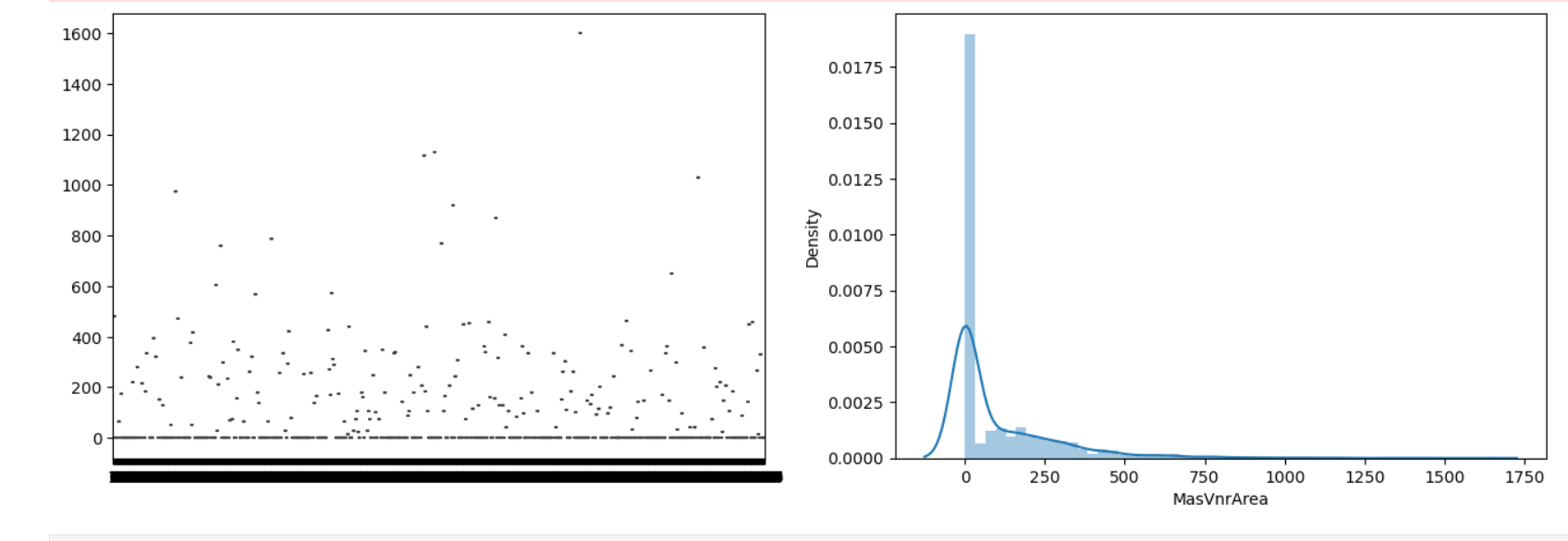


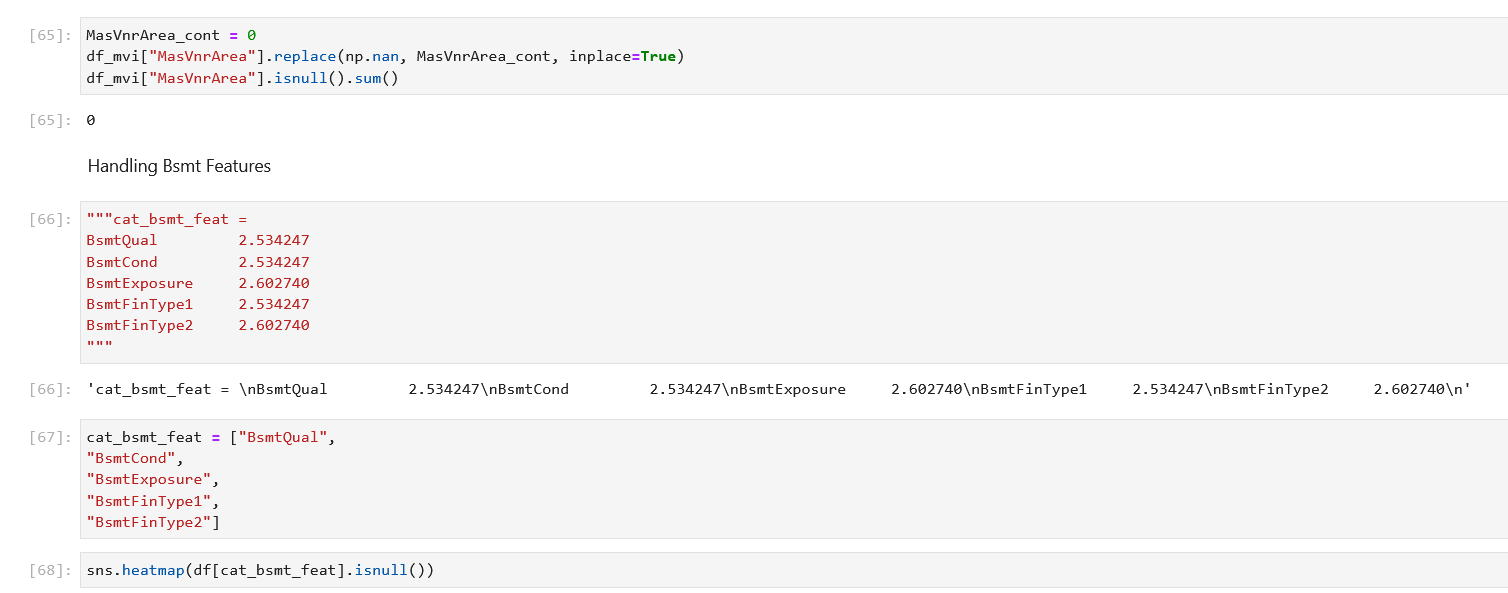


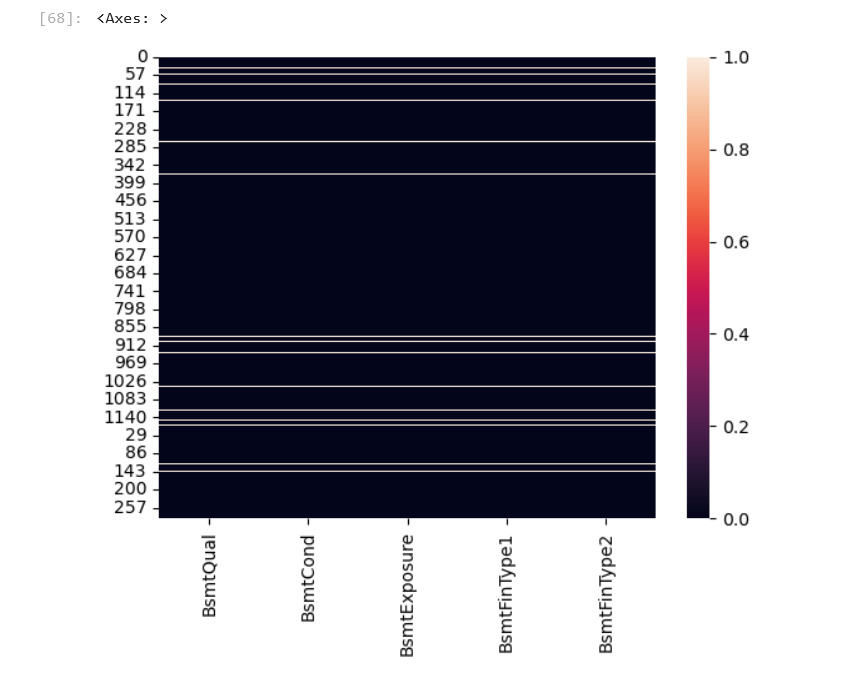


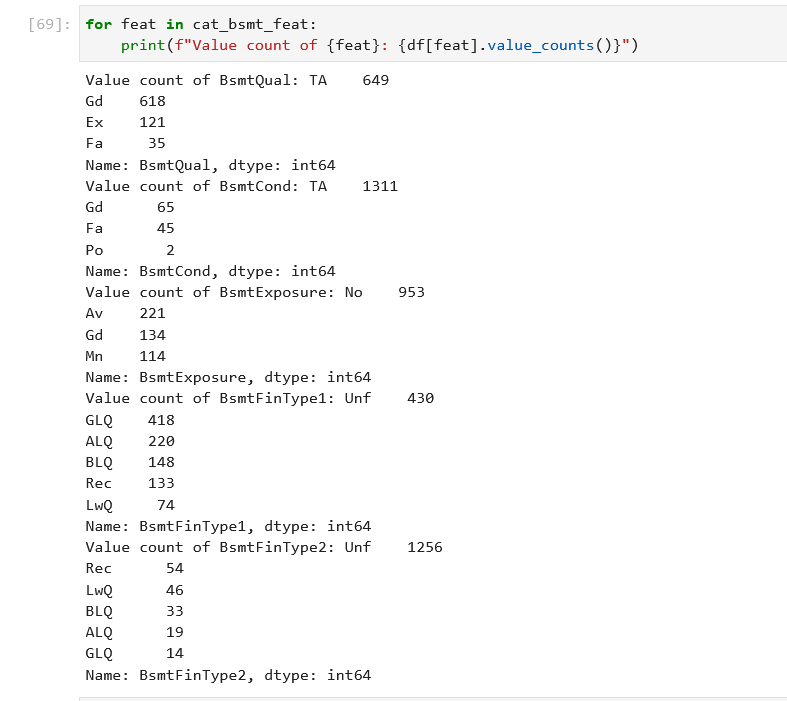


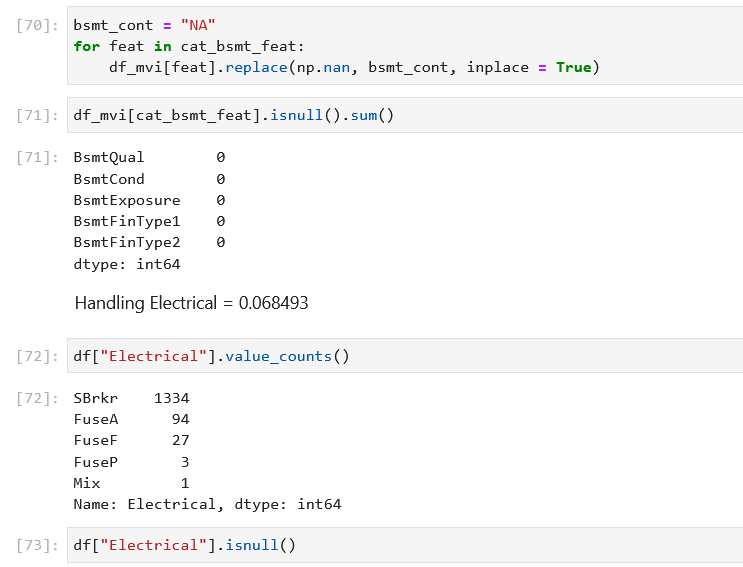


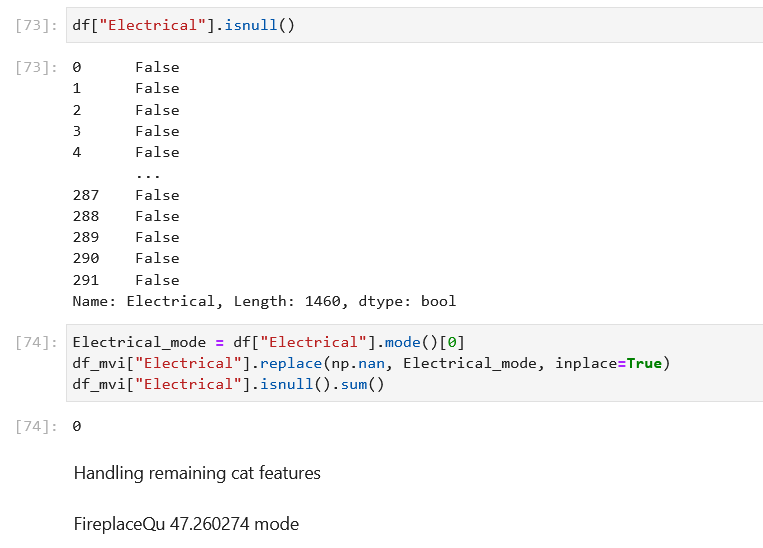


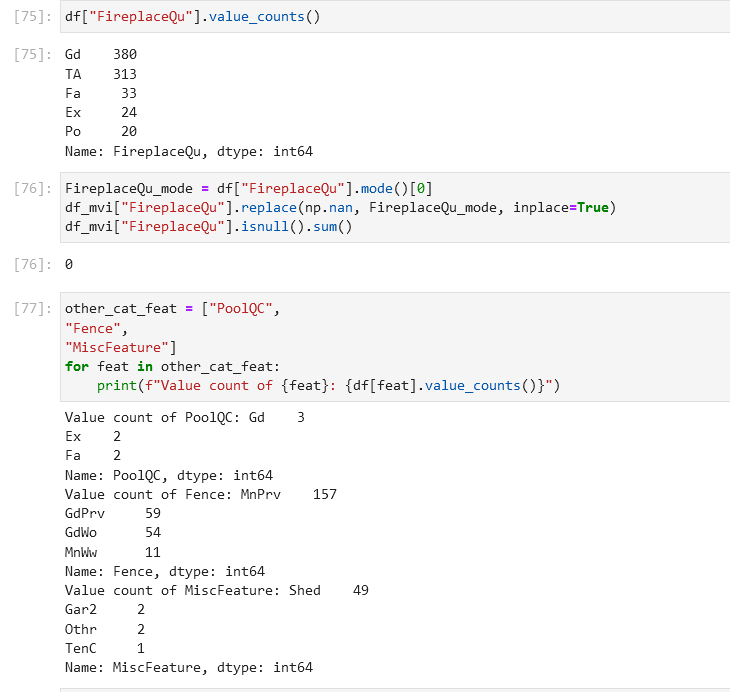


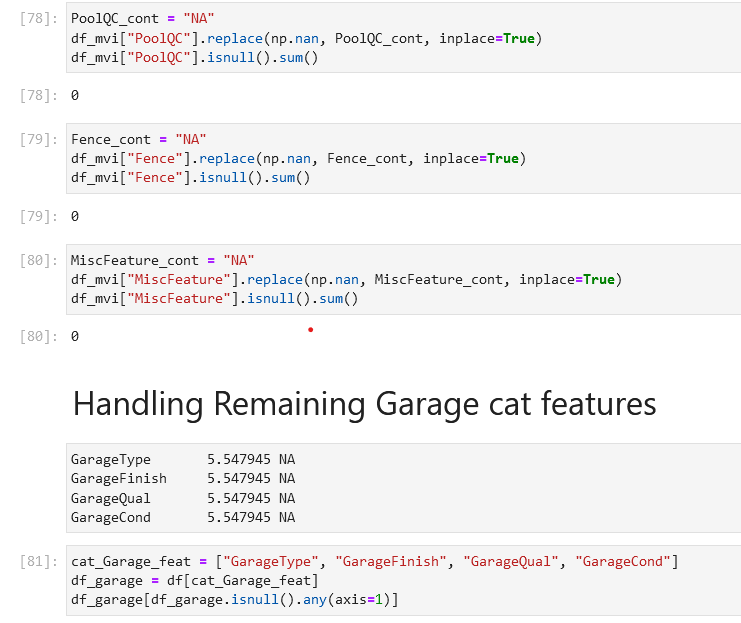
****

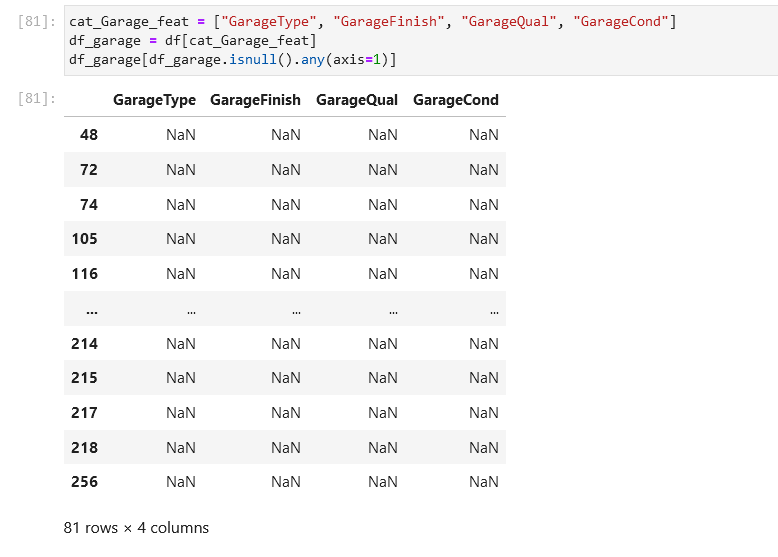


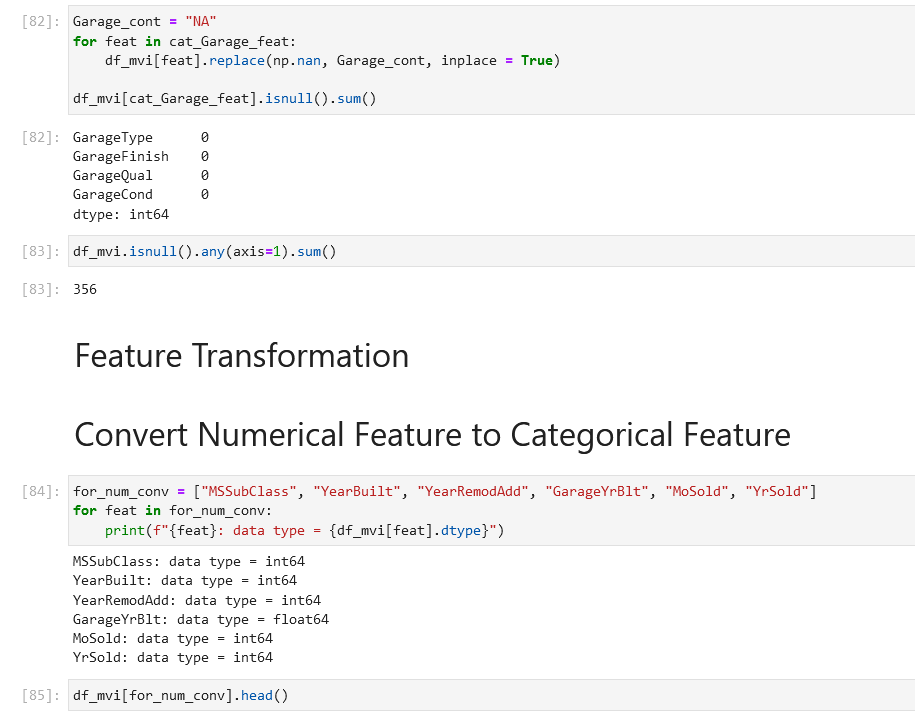


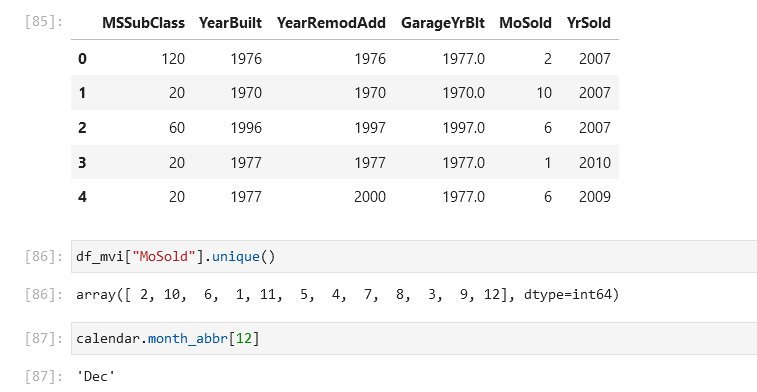


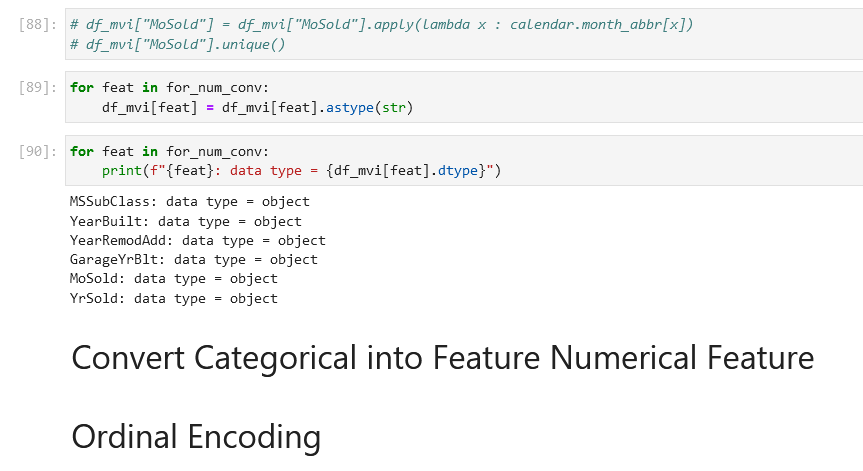




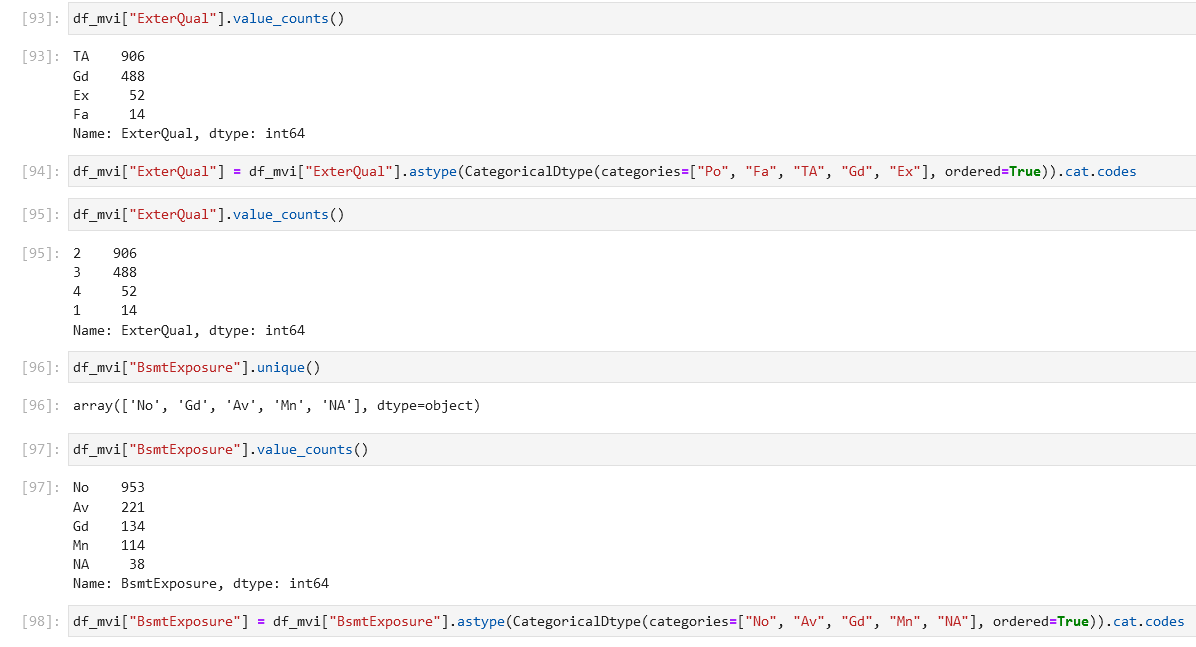


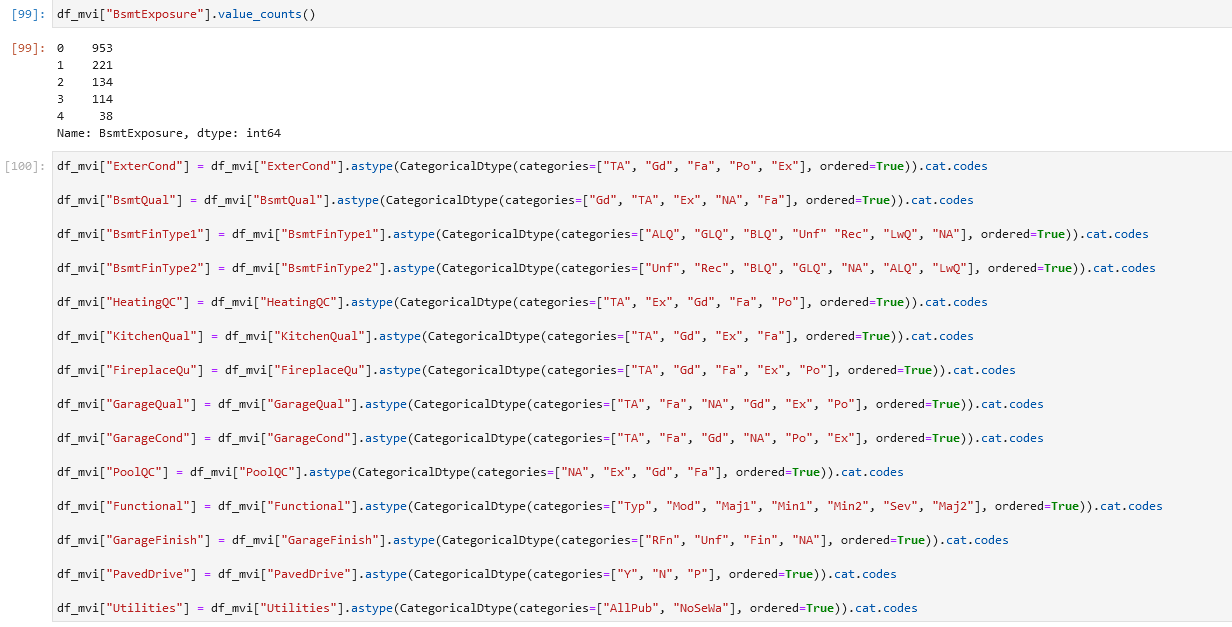


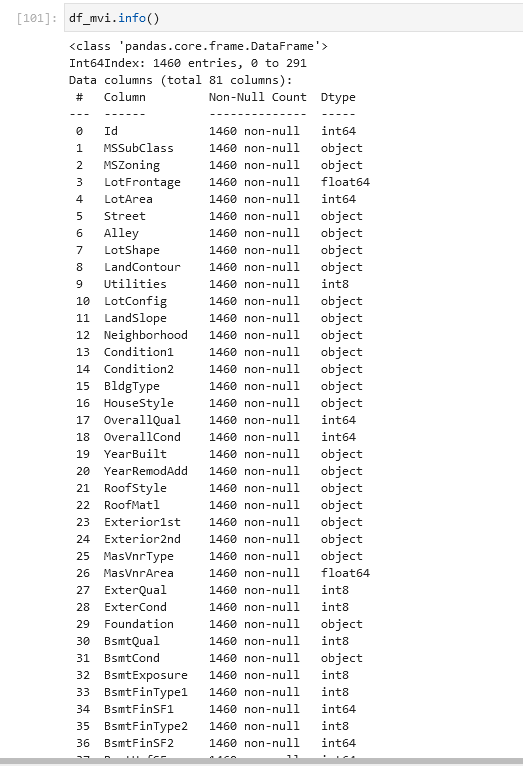


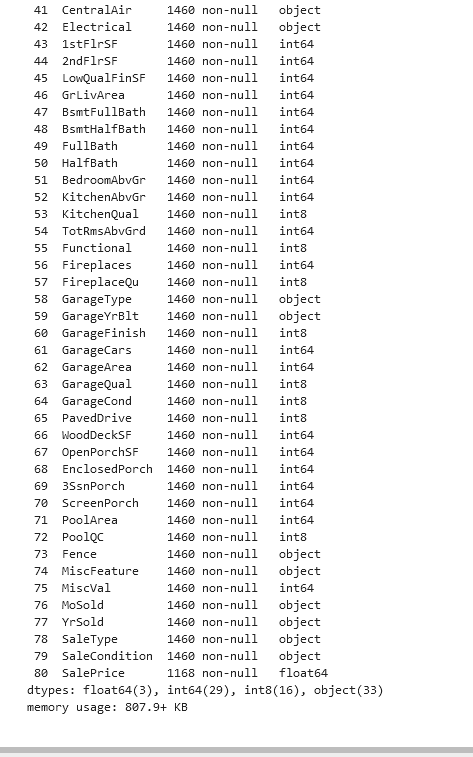


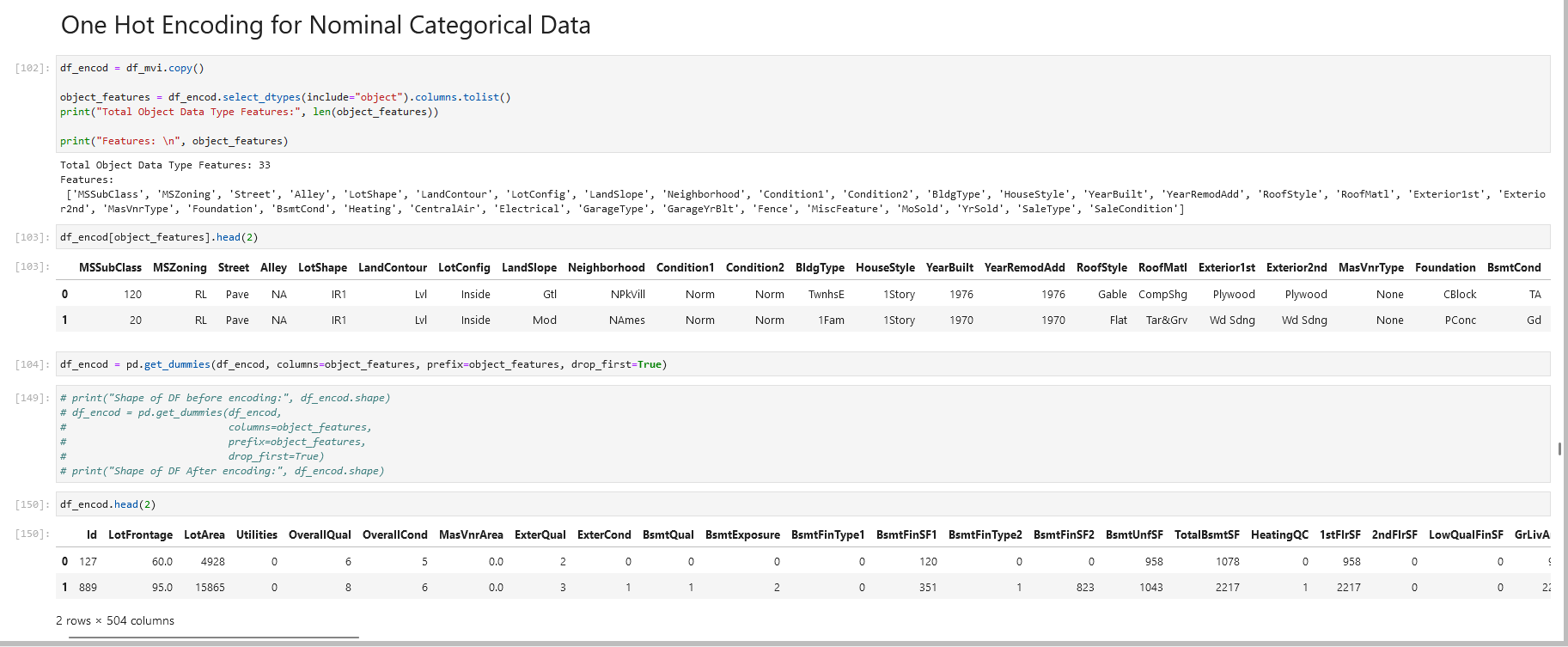




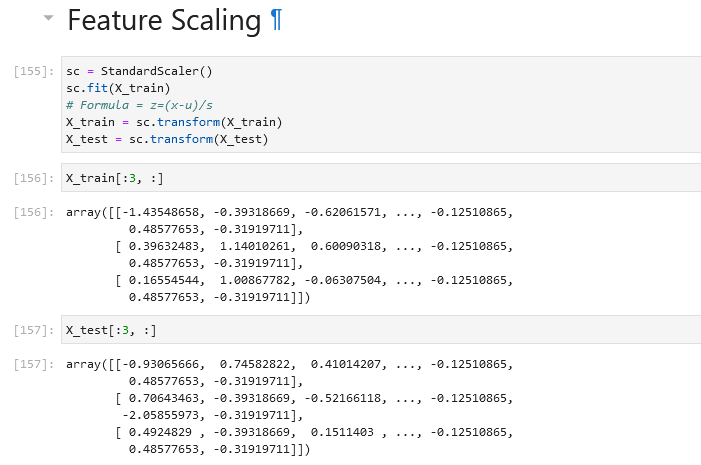


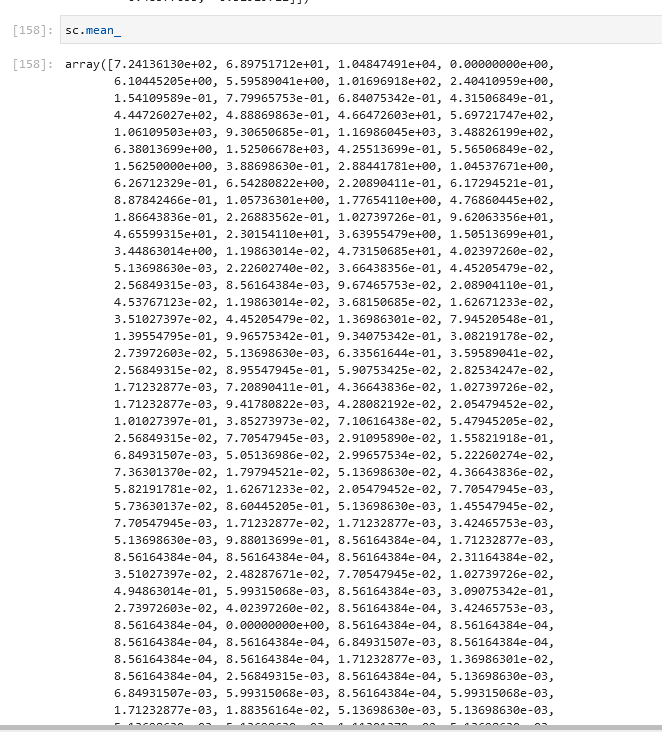


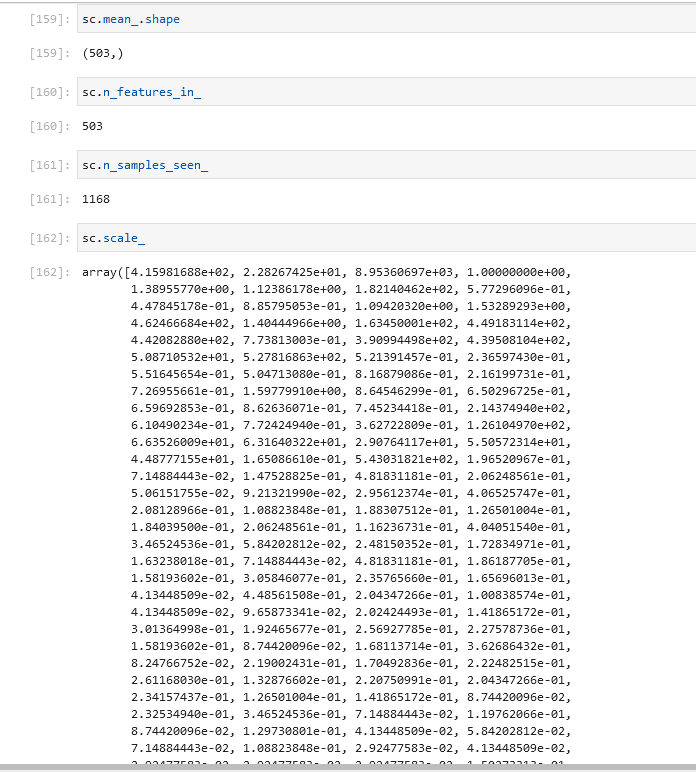


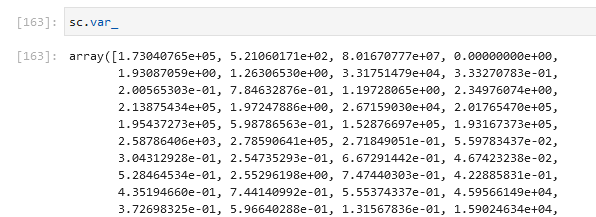


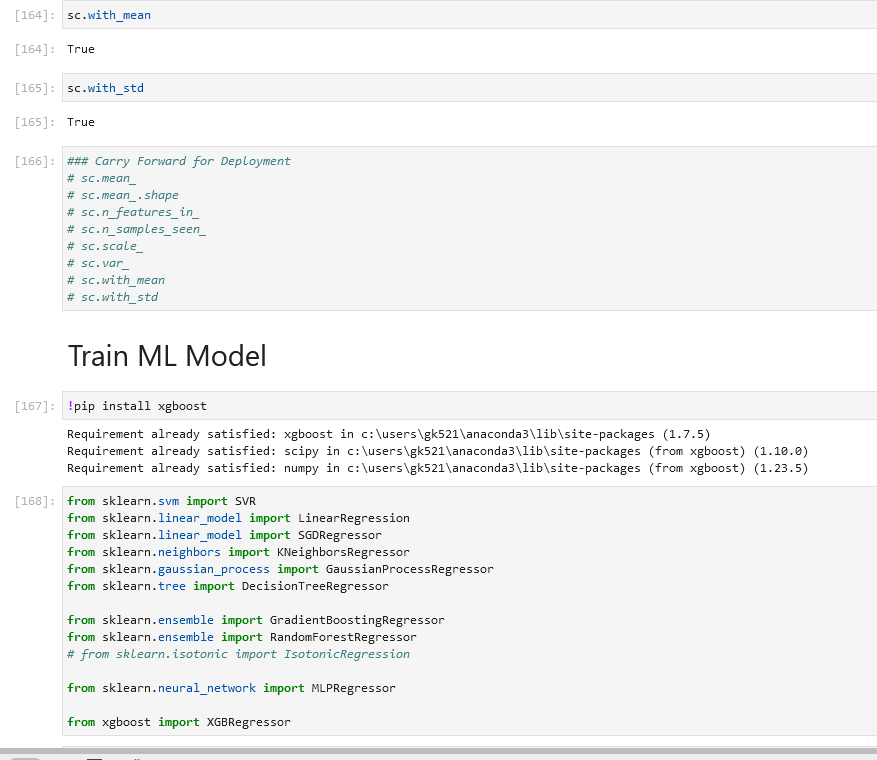




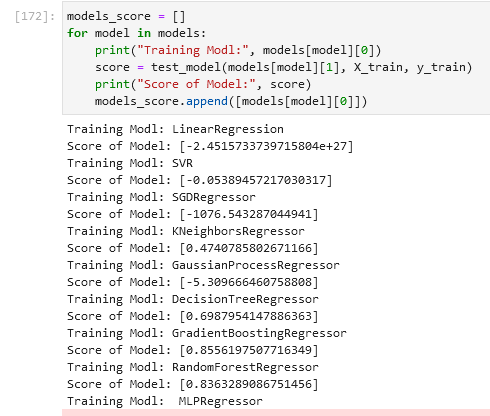


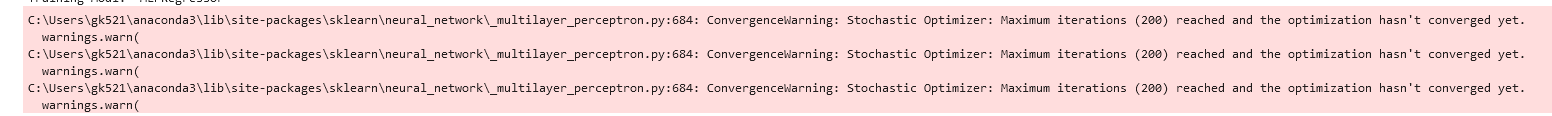


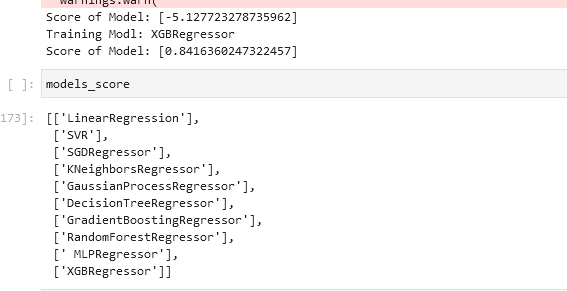












**CHAPTER 6**

# CONCLUSION AND LEARNING EXPERIENCE

### Conclusion

To get more accuracy, we trained all top supervised regression algorithms but you can try out a few of them which are always popular. After training all algorithms, we found that GradientBoostingRegressor and XGBoost regressor have given high accuracy than remain but we have chosen XGBoost.

As ML Engineer, we always retrain the deployed model after some period of time to sustain the accuracy of the model. We hope our efforts will help to predict the price of a house for the buyer and seller.

### Learning Experience

### If you're interested in learning how to predict house prices using machine learning, here's a step-by-step learning experience that you can follow:

### Understand the problem: Begin by familiarizing yourself with the task of house price prediction. Understand the variables involved, such as square footage, number of bedrooms, location, etc., and the target variable, which is the house price.

### Gather a dataset: Look for a suitable dataset that includes relevant features and corresponding house prices. You can explore resources like Kaggle, UCI Machine Learning Repository, or real estate websites. Make sure the dataset is reliable and well-documented.

### Data preprocessing: Clean the dataset by handling missing values, outliers, and inconsistent data. Perform feature engineering, which involves transforming or creating new features that might enhance the predictive power. For example, you could calculate the price per square foot or extract additional information from addresses or descriptions.

### Split the data: Divide your dataset into two subsets: a training set and a test set. The training set will be used to train your machine learning model, while the test set will serve as an independent evaluation to assess its performance.

### Choose a model: Research and select an appropriate machine learning algorithm for regression tasks. Linear regression, decision trees, random forests, or gradient boosting algorithms are popular choices for house price prediction. Consider factors like interpretability, model complexity, and scalability when making your decision.

### Train the model: Use the training set to train your chosen model. During this phase, the model will learn the patterns and relationships in the data, adjusting its internal parameters accordingly. Adjust hyperparameters (e.g., learning rate, regularization) to optimize the

### model's performance.

### Evaluate the model: Apply the trained model to the test set to assess its performance. Common evaluation metrics for regression tasks include mean absolute error (MAE), mean squared error (MSE), and R-squared value. Analyze the results to understand how well the model is predicting house prices.

### Iterate and improve: Analyze the model's performance and identify areas for improvement. You can try different algorithms, feature engineering techniques, or hyperparameter tuning to enhance the model's accuracy. Consider using cross-validation or more advanced techniques like ensemble learning to further improve the predictions.

### Deploy the model: Once you're satisfied with the model's performance, deploy it to make predictions on new, unseen data. You can create a user-friendly interface or integrate it into a web application to allow users to input property features and obtain price predictions.

### Continued learning: Keep exploring and learning about more advanced techniques for house price prediction. This can include incorporating external data sources, using deep learning models, or exploring time series analysis for housing market trends.

**REFERENCES**

1. Jiawei Han, MichelineKamber, “Data Mining Concepts and Techniques”, pp. 279-328, 2001.
2. Adyan Nur Alfiyatin , Hilman Taufiq, Ruth Ema Febrita and Wayan Firdaus Mahmudy, “Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017, pp. 323 to 326.
3. Ong, S. E., Ho, K. H. D. and Lim, C. H., “A constantquality price index for resale public housing flats in Singapore”, Urban Studies, 40(13), 2003, pp. 2705 –2729.
4. Nikita Patel and Saurabh Upadhyay, “Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA”, International Journal of Computer Applications, Volume 60– No.12, December 2012, pp 20-25.
5. SamDrazin and Matt Montag”, Decision Tree Analysis using Weka”, Machine Learning-Project II, University of Miami.
6. J. R.Quinlan,” Simplifying decision trees”, Int. J. HumanComputer Studies.
7. K.C. Tan, E.J. Teoh, Q. Yu, K.C. Goh,” A hybrid evolutionary algorithm for attribute selection in data mining”, Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore.Rochester, Institute of Technology, USA
8. Timothy C. Au, “Random Forests, Decision Trees, and Categorical Predictors: The Absent Levels Problem”, Journal of Machine Learning Research 19, 2018, pp. no.1- 30