VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANA SANGAMA", BELAGAVI - 590 018

A MINI PROJECT REPORT

on

# "HOUSE PRICE PREDICTION"

*Submitted by*


| | |
|---|---|
| Mahesh | 4SF20CD022 |
| Sanskar S Khandelwal | 4SF20CD038 |


*In partial fulfillment of the requirements for the VII semester*

## VISUALIZATION & DS MINI PROJECT LABORATORY

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE & ENGINEERING(DATA SCIENCE)

*Under the Guidance of*

Mrs. Shwetha. S. Shetty

Assistant Professor, Department of ISE and CSE(Data Science)

at


SAHYADRI

College of Engineering & Management

An Autonomous Institution

MANGALURU


2023 - 24

# SAHYADRI

## College of Engineering & Management

### An Autonomous Institution

### MANGALURU

### Department of Computer Science & Engineering(Data Science)


# CERTIFICATE

This is to certify that the **Mini Project** entitled **"House Price Prediction"** has been carried out by **Mahesh (4SF20CD022)** and **Sanskar S Khandelwal (4SF20CD038)**, the bonafide students of Sahyadri College of Engineering & Management in partial fulfillment of the requirements for the VII semester **Visualization & DS Mini Project Laboratory (18ADL76)** of **Bachelor of Engineering in Computer Science & Engineering(Data Science)** of Visvesvaraya Technological University, Belagavi during the year 2023 - 24. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The mini project report has been approved as it satisfies the academic requirements in respect of mini project work.

| | |
|---|---|
| **Mrs. Shwetha. S. Shetty** | **Dr. Mustafa Basthikodi** |
| Assistant Professor | Professor & Head |
| Dept. of ISE & CSE(DS), SCEM | Dept. of ISE & CSE(DS), SCEM |

### External Practical Examination:

| Examiner's Name | Signature with Date |
|---|---|
| 1. ..... | ..... |
| 2. ..... | ..... |

# SAHYADRI

## College of Engineering & Management

### An Autonomous Institution

### MANGALURU

### Department of Computer Science & Engineering(Data Science)


# DECLARATION

We hereby declare that the entire work embodied in this Mini Project Report titled **"House Price Prediction"** has been carried out by us at Sahyadri College of Engineering and Management, Mangaluru under the supervision of **Mrs. Shwetha. S. Shetty** as the part of the VII semester **Visualization & DS Mini Project Laboratory(18ADL76)** of **Bachelor of Engineering** in **Computer Science & Engineering(Data Science)**. This report has not been submitted to this or any other University.

**Mahesh (4SF20CD022)**

**Sanskar S Khandelwal (4SF20CD038)**

SCEM, Mangaluru

# Abstract

The House Price Prediction and Visualization web application integrate machine learning and data visualization to assist users in estimating house prices in Bengaluru. Utilizing a Linear Regression model trained on a preprocessed dataset, the app enables real-time predictions based on user-input parameters such as total square feet, number of bathrooms, and location. The streamlined interface, developed using Streamlit, facilitates user-friendly interactions, offering insights into the intricate relationships within the dataset. The Visualization page enhances the user experience with scatter charts, histograms, box plots, and heatmaps, providing a comprehensive analysis of key factors influencing house prices. Through the seamless integration of scikit-learn, Streamlit, and a Linear Regression model, this web app bridges the gap between machine learning and user-friendly visualization, empowering users to make informed decisions in the real estate domain.

# Acknowledgement

It is with great satisfaction and euphoria that we are submitting the Mini Project Report on **"House Price Prediction"**. We have completed it as a part of the VII semester **Visualization & DS Mini Project Laboratory(18ADL76)** of **Bachelor of Engineering** in **Computer Science & Engineering(Data Science)** of Visvesvaraya Technological University, Belagavi.

We are profoundly indebted to our guide, **Mrs. Shwetha. S. Shetty**, Assistant Professor, Department of Information Science & Engineering and Computer Science & Engineering( Data Science) for innumerable acts of timely advice, encouragement and We sincerely express our gratitude.

We express our sincere gratitude to **Dr. Mustafa Basthikodi**, Professor &Head, Department of Information Science & Engineering and Computer Science & Engineering(Data Science) for his invaluable support and guidance.

We sincerely thank **Dr. S. S. Injaganeri**, Principal, Sahyadri College of Engineering & Management who have always been a great source of inspiration.

Finally, yet importantly, We express our heartfelt thanks to our family & friends for their wishes and encouragement throughout the work.

**Mahesh**
4SF20CD022
VII Sem, B.E., CSE(DS)
SCEM, Mangaluru

**Sanskar S Khandelwal**
4SF20CD038
VII Sem, B.E., CSE(DS)
SCEM, Mangaluru

# Table of Contents

# List of Figures

# Chapter 1

# Introduction

The housing price prediction data visualization project represents a sophisticated exploration into the multifaceted landscape of real estate dynamics. Leveraging cutting-edge tools such as Streamlit and potentially incorporating a linear regression model, the initiative aspires to provide an immersive and visually captivating platform. At its core, the project endeavors to demystify the intricate web of factors influencing housing prices, translating complex datasets into dynamic charts and graphs. Through the intuitive user interface, both potential homebuyers and sellers are invited to navigate the correlations between pivotal variables like square footage, bedrooms, bathrooms, and location. This interactive experience serves as a conduit for fostering a deeper comprehension of the anticipated trends in house prices, empowering users to make more informed decisions. With a user-friendly design, the platform aims to enhance transparency in the real estate realm, contributing to the broader goal of facilitating well-informed decision-making in the dynamic and often opaque domain of property transactions. By bridging the gap between raw data and actionable insights, the project aligns with the overarching mission of instilling transparency and promoting a more enlightened approach to navigating the challenges of the real estate market.

## 1.1 Overview

The housing price prediction data visualization project offers a comprehensive overview of real estate market dynamics. Leveraging tools like Streamlit and possibly incorporating a linear regression model, the project aims to create an interactive interface for users. Through dynamic charts and graphs, it visualizes relationships between key factors such as square footage, bedrooms, bathrooms, and location, empowering users to explore and

understand predicted house prices. This user-friendly application serves both potential buyers and sellers, aiding in informed decision-making. With a focus on transparency and accessibility, the project contributes to a more data-driven and efficient real estate ecosystem, enhancing understanding and decision-making in the housing market.

## 1.2 Purpose

The purpose of a housing price prediction data visualization project is to provide a user-friendly and interactive platform for understanding and exploring the factors influencing property values. By leveraging data visualization techniques, the project aims to offer insights into complex market trends, empowering users, including potential homebuyers, sellers, and real estate professionals, to make informed decisions. The visualization enhances transparency, allowing users to grasp the relationships between various features and predicted house prices. Additionally, the project may contribute to a more data-driven real estate ecosystem, enabling stakeholders to navigate market complexities and fostering a greater understanding of the dynamics shaping property values.

## 1.3 Scope

The housing price prediction data visualization project holds immense scope in empowering both consumers and industry professionals. By presenting intricate market trends through interactive charts and graphs, it facilitates informed decision-making for potential buyers and sellers. Additionally, it serves as a valuable tool for real estate agents, allowing them to better understand and communicate property values. The project's scalability enables integration with advanced machine learning models and expansion into broader real estate analytics. In an evolving market, this visualization project is pivotal for navigating complexities, enhancing transparency, and contributing to a more data-driven and efficient real estate ecosystem.

# Chapter 2

# Requirements Specification

## 2.1 Hardware Specification

- Processor : Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz
- RAM : 8GB
- Hard Disk : 500GB
- Input Device : Standard keyboard and Mouse
- Output Device : Monitor

## 2.2 Software Specification

- Programming Language :Python
- IDE : Visual Studio Code
- Framework : Streamlit

# Chapter 3

# System Process

## 3.1 Framework

The framework of the web application is shown in below figure:

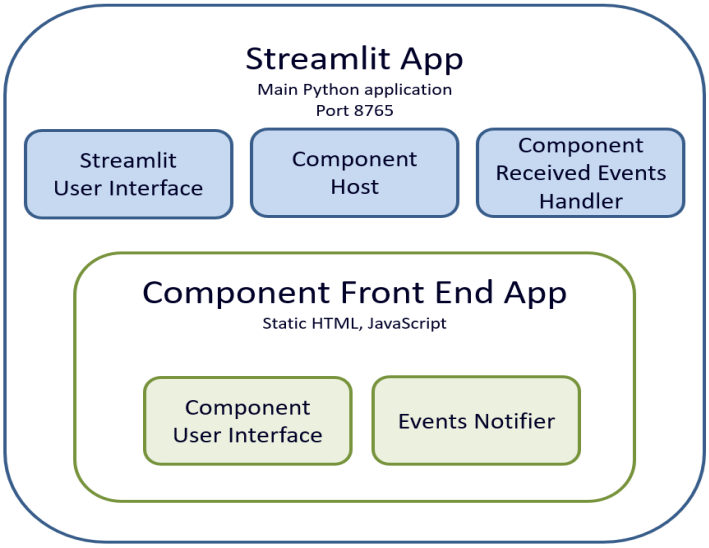


Figure 3.1: Framework of data visualization web application

Streamlit is a Python framework instrumental in crafting a seamless and interactive user interface for the house price prediction data visualization project. With its intuitive syntax and real-time updates, Streamlit transforms data scripts into dynamic web applications effortlessly. This framework facilitates the integration of charts, graphs, and user input elements, allowing users to explore predicted house prices based on various factors. Its simplicity and versatility make it an ideal choice for developers, providing a visually engaging and accessible platform for users to gain insights into the intricate relationships between different variables influencing property values in the real estate market.

## 3.2 Plotting

### 3.2.1 Heat Map

The heatmap visually represents the correlation matrix of numeric columns in the dataset. Each cell displays the correlation coefficient between two variables, with colors ranging from cool (negative correlation) to warm (positive correlation). Annotations provide the precise correlation values, aiding in the interpretation of relationships between different features.

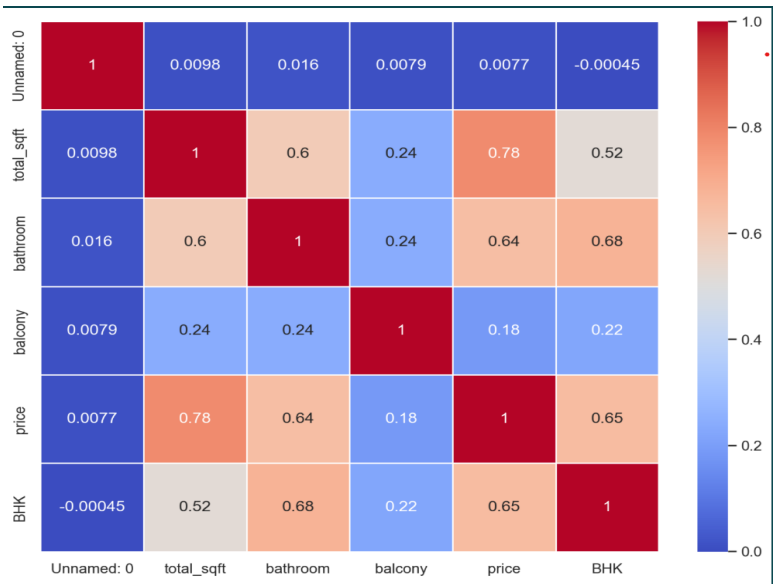


Figure 3.2: Heat map analysis for dataset

### 3.2.2 Regression plot

The regression plot of Actual vs Predicted values visually compares the model's predictions to the actual observed values. The diagonal line represents perfect predictions. Data points above or below the line indicate overestimation or underestimation. The plot helps assess the model's accuracy and identify potential patterns or discrepancies.

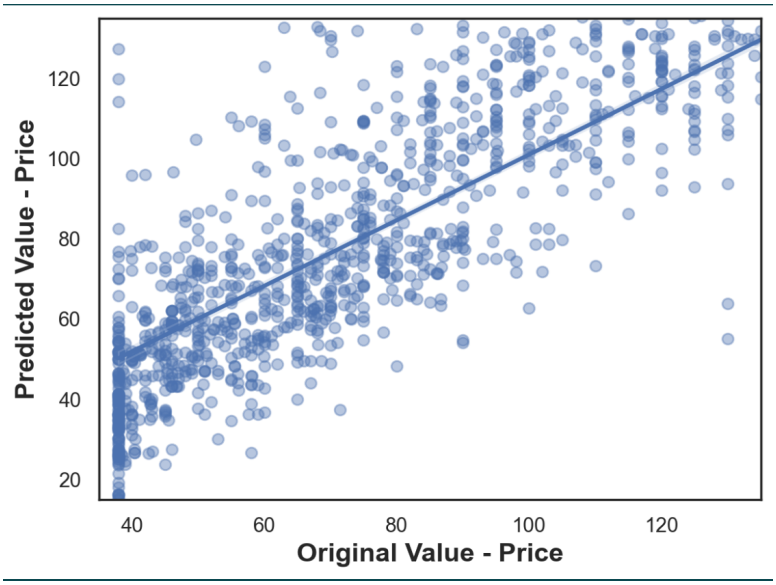


Figure 3.3: Regression plot for Actual vs Predicted value

### 3.2.3 Distribution plot

The distribution plot of Actual vs Predicted values illustrates the alignment and spread of the model predictions compared to the actual values. It provides insights into the consistency and accuracy of predictions, showcasing any tendencies for overestimation or underestimation. The plot aids in assessing the model's overall performance and reliability.

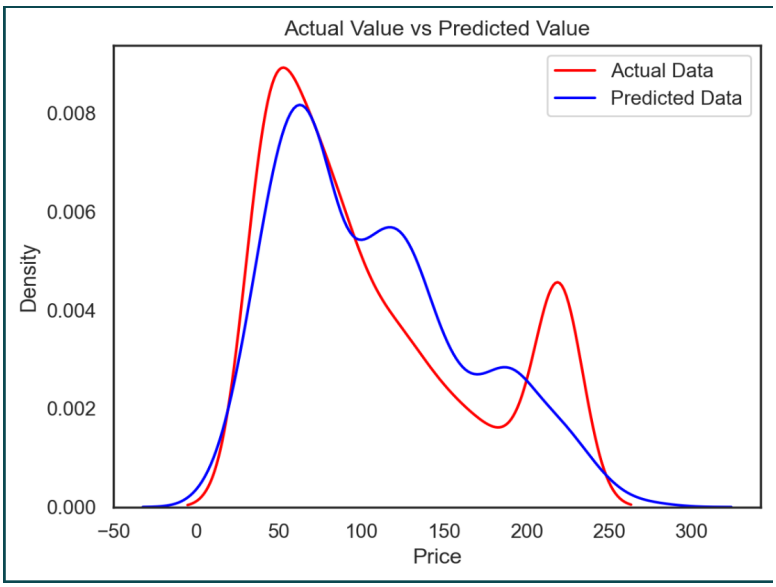


Figure 3.4: Distribution plot for Actual vs Predicted value

### 3.2.4 Bar plot

A bar plot comparing Bathroom and BHK typically displays the frequency or count of each category for these variables. Each bar represents the count of bathrooms or BHK values, allowing for a visual comparison of their distribution. This type of plot helps identify patterns, trends, or disparities in the data related to the number of bathrooms and bedrooms in a given context, such as in real estate or housing analysis.

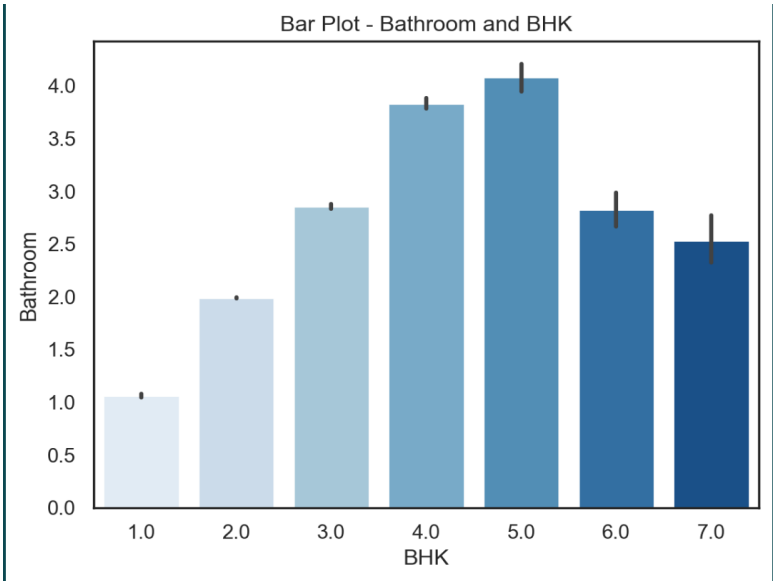


Figure 3.5: Bar plot used for comparing Bathroom and BHK

### 3.2.5 Histogram plot

A histogram plot for Bathroom count distribution visualizes the frequency or count of different ranges of bathroom counts in a dataset. The x-axis represents the ranges of bathroom counts (e.g., 1-2, 3-4), and the y-axis shows the frequency of occurrences within each range. This type of plot helps to understand the distribution of bathroom counts in a dataset, highlighting common ranges and potential outliers

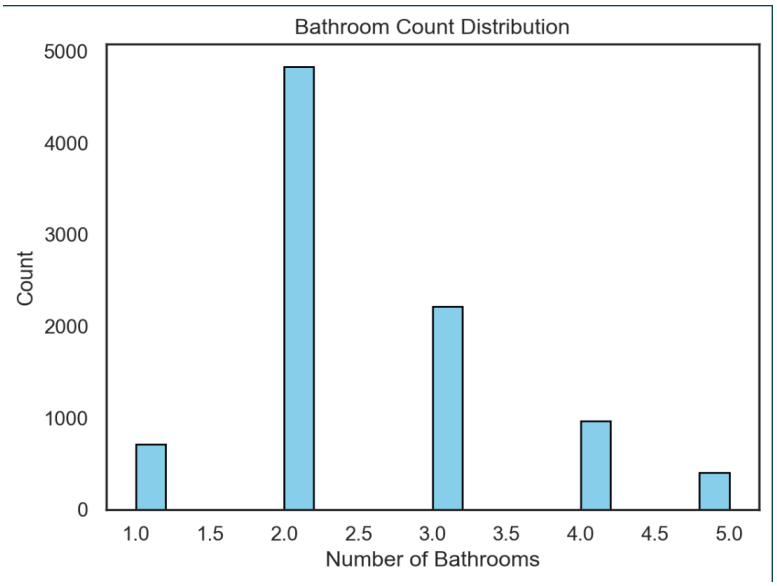


Figure 3.6: Histogram plot for Bathroom count Distribution

### 3.2.6 Histogram Plot

A histogram plot for Total Square Feet distribution illustrates the frequency or count of different ranges of total square footage in a dataset. The x-axis represents the ranges of total square feet (e.g., 0-500, 500-1000), and the y-axis shows the frequency of occurrences within each range. This type of plot provides insights into the distribution of property sizes, helping to identify common size ranges and potential patterns in the dataset.

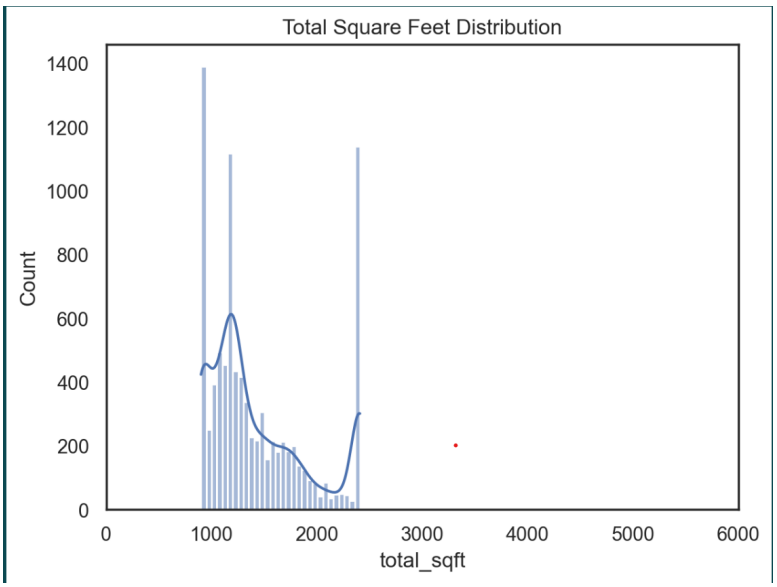


Figure 3.7: Histogram plot for Total Sqaure Feet

### 3.2.7 Box Plot

A box plot for price analysis provides a visual summary of the distribution of prices in a dataset. The box represents the interquartile range (IQR), with the median depicted as a line inside the box. Whiskers extend to the minimum and maximum values within a certain range. Outliers may be shown as individual points beyond the whiskers. This plot helps assess the central tendency, spread, and presence of outliers in the price distribution.

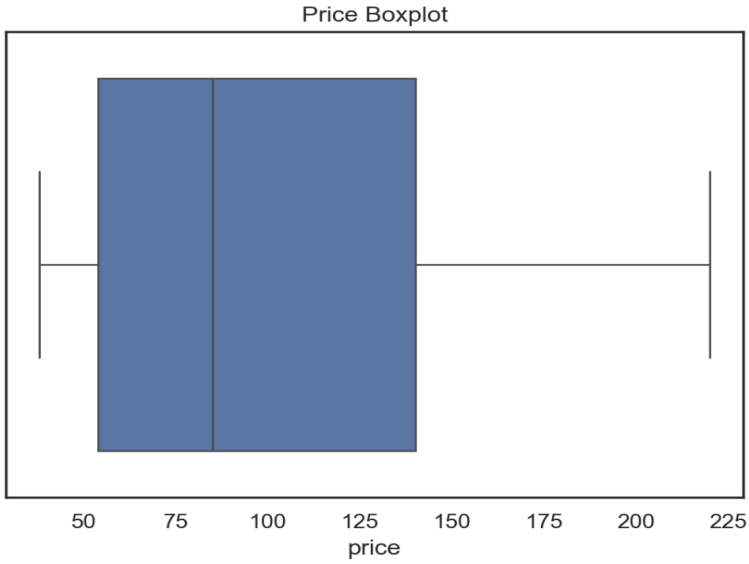


Figure 3.8: Box plot for Price analysis

### 3.2.8 Bar Plot

A bar plot of BHK (Bedrooms, Hall, Kitchen) distribution illustrates the frequency or count of different BHK categories in a dataset. Each bar represents a specific BHK category and the height of the bar corresponds to the frequency or count of occurrences for each category.

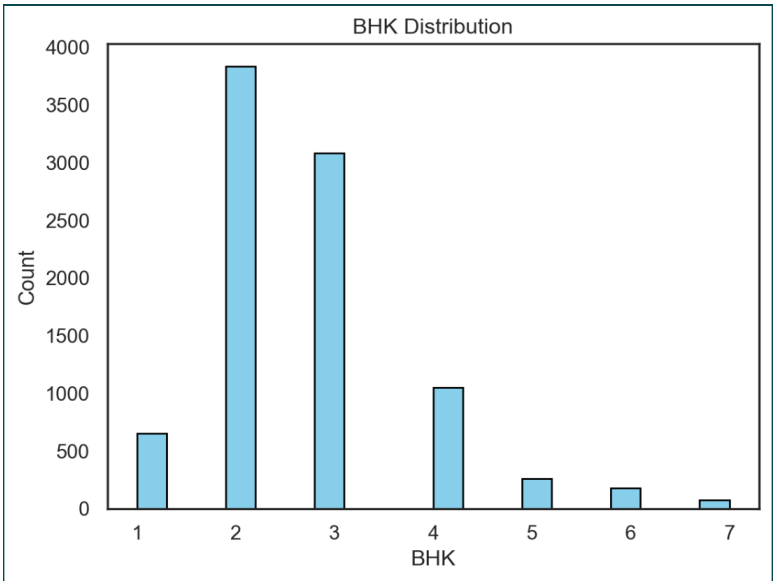


Figure 3.9: Bar plot for BHK Distribution

# Chapter 4

# Implementation

## 4.1 Overview

The house price prediction data visualization project aims to provide a comprehensive overview of factors influencing property values through the lens of visual representation. Utilizing tools like Python's Streamlit library for the front end and possibly employing a linear regression model for prediction, this project offers an interactive interface for users to explore and understand housing market trends. The visualization components may include dynamic charts, graphs, and maps showcasing relationships between features such as square footage, bedrooms, bathrooms, and geographical location. Users can input specific parameters and observe the predicted house prices in real-time, gaining valuable insights into the impact of different variables. The project not only enhances user experience but also serves as a practical tool for potential home buyers and sellers, providing them with a visual understanding of how various factors contribute to property valuations. The integration of data visualization with predictive modeling creates a user-friendly application, facilitating informed decision-making in the real estate domain. Continuous refinement and feedback loops ensure the project remains adaptable to the ever-changing dynamics of the housing market.

## 4.2 Languages Used

### 4.2.1 Python

Python is a popular general-purpose programming language that is interpreted, interactive, object-oriented, and high-level. Python's philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than might be possible in languages such as C++ or Java. Python serves as a robust language for developing house price prediction data visualization projects, providing a versatile ecosystem for data analysis and visualization. Leveraging libraries such as Pandas, NumPy, and Matplotlib, Python enables seamless data manipulation, statistical analysis, and the creation of insightful visualizations. Additionally, with the integration of machine learning libraries like Scikit-learn, Python empowers developers to implement predictive models for accurate house price estimation. Its readability, extensive community support, and vast array of packages make Python an ideal choice for efficiently crafting interactive and informative visualizations that enhance the understanding of factors influencing house prices in a user-friendly manner.

### 4.2.2 Streamlit

Streamlit is a Python library designed for creating web applications with minimal effort, particularly well-suited for data-driven projects and machine learning applications. It simplifies the development of interactive and shareable web apps directly from Python scripts. With its user-friendly API, developers can transform data scripts into functional applications with just a few lines of code. Streamlit handles the complexities of front-end development, allowing users to focus on data visualization and analysis. Its built-in widgets enable easy interaction with data, making it accessible to both technical and non-technical users. The ability to integrate seamlessly with popular data science libraries like Pandas and Matplotlib further enhances its appeal for rapid prototyping and deployment of data-centric projects. Overall, Streamlit facilitates a smooth transition from data analysis to user-friendly applications, making it an excellent choice for those who prioritize simplicity and efficiency in front-end development.

## 4.3 Visual Studio Code

Visual Studio Code (VS Code) is a lightweight yet powerful source-code editor developed by Microsoft. Widely embraced in the programming community, it supports various programming languages and provides a platform for extensions and customization. In the Diabetes Prediction project, we harnessed the capabilities of Python libraries within the Visual Studio Code integrated development environment (IDE) to streamline our development process. Utilizing popular libraries like Matplotlib, Seaborn, and Plotly, we leveraged VS Code's versatility to write and execute Python code seamlessly. The IDE's robust features, such as integrated Git support and interactive debugging tools, enhanced our collaboration and code efficiency. VS Code's extensibility further allowed us to integrate machine learning algorithms and create diverse data visualizations effortlessly, ensuring a smooth and productive development experience throughout the project's lifecycle.

## 4.4 Pseudo Code Used

### 4.4.1 Pseudocode for Heat map

The code segment is for the "Heatmap Analysis" page in a Streamlit web app. It selects numeric columns from the dataset, calculates the correlation matrix, and generates a heatmap using Seaborn and Matplotlib. The resulting heatmap visualizes the correlations between numeric features, providing insights into relationships within the data. The plot is displayed using Streamlit.

```
elif page == "Heatmap Analysis":
    st.title('Heatmap Analysis')

    # Select only numeric columns for correlation matrix
    numeric_columns = Data_frame.select_dtypes(include=['float64', 'int64']).columns
    corr_matrix = Data_frame[numeric_columns].corr()

    # Plot the heatmap
    fig, ax = plt.subplots(figsize=(10, 8))
    sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', linewidths=.5)
    st.pyplot(fig)
```

Figure 4.1: Pseudo code for the heat map

### 4.4.2 Pseudo code for Regression plot

This code segment corresponds to the "Regression Plot" page in a Streamlit web app. It utilizes Seaborn and Matplotlib to create a regression plot comparing actual house prices against predicted prices. The plot visualizes the relationship between the original and predicted values, with specified axis labels, limits, and styles. The resulting plot is displayed in the Streamlit web app.

```
# Regression Plot
elif page == "Regression Plot":
    st.title('Regression Plot - Actual vs Predicted')
    sns.set(color_codes=True)
    sns.set_style("white")
    fig, ax = plt.subplots()
    ax = sns.regplot(x=st.session_state['Y_test'], y=st.session_state['Y_pred'], scatter_kws={'alpha': 0.4})
    ax.set_xlabel('Original Value - Price', fontsize='large', fontweight='bold')
    ax.set_ylabel('Predicted Value - Price', fontsize='large', fontweight='bold')
    ax.set_xlim(35, 135)
    ax.set_ylim(15, 135)
    st.pyplot(fig)
```

Figure 4.2: Pseudo code for regression plot

### 4.4.3 Pseudo code for bar plot

This code segment corresponds to the "Bar Plot Analysis" page in a Streamlit web app. It uses Seaborn and Matplotlib to create a bar plot comparing the number of bathrooms for different numbers of bedrooms. The plot is customized with a blue color palette and appropriate axis labels. The resulting plot is displayed in the Streamlit web app.

```
# Bar Plot - Bathroom and BHK
elif page == "Bar Plot Analysis":
    st.title('Bar Plot - Bathroom and BHK')
    fig, ax = plt.subplots()
    sns.barplot(x='BHK', y='bathroom', data=Data_frame, palette='Blues', ax=ax)
    plt.title('Bar Plot - Bathroom and BHK')
    plt.xlabel('BHK')
    plt.ylabel('Bathroom')
    st.pyplot(fig)
```

Figure 4.3: Pseudo code for bar plot

### 4.4.4 Pseudo code for box plot

This code corresponds to the "Bathroom Count Analysis" page in a Streamlit web app. It uses Seaborn and Matplotlib to create two plots: a boxplot showing the distribution of bathroom counts (bathroom) and a histogram illustrating the overall distribution of bathroom counts. Both plots are displayed in the Streamlit web app, providing insights into the data's bathroom count characteristics.

```
# Bathroom Count Analysis
elif page == "Bathroom Count Analysis":
    st.title("Bathroom Count Analysis")
    st.subheader("Boxplot for Bathroom Count")
    fig, ax = plt.subplots()
    sns.boxplot(x='bathroom', data=Data_frame, ax=ax)
    plt.title("Bathroom Count Boxplot")
    st.pyplot(fig)

    st.subheader("Histogram for Bathroom Count")
    fig, ax = plt.subplots()
    plt.hist(Data_frame['bathroom'], bins=20, color='skyblue', edgecolor='black')
    plt.xlabel("Number of Bathrooms")
    plt.ylabel("Count")
    plt.title("Bathroom Count Distribution")
    st.pyplot(fig)
```

Figure 4.4: Pseudo code for box plot

### 4.4.5 Pseudo code for histogram

This code corresponds to the "Price Analysis" page in a Streamlit web app. It uses Seaborn and Matplotlib to create two plots: a histogram illustrating the distribution of prices (price) and a boxplot depicting the statistical summary of the price data. Both plots are displayed in the Streamlit web app, providing insights into the distribution and variability of housing prices.

```
# Price Analysis
elif page == "Price Analysis":
    st.title("Price Analysis")
    st.subheader("Histogram for Price")
    fig, ax = plt.subplots()
    sns.histplot(x='price', data=Data_frame, bins=20, kde=True, ax=ax, color='skyblue', edgecolor='black')
    plt.xlim([0, 2000])
    plt.xlabel("Price")
    plt.ylabel("Count")
    plt.title("Price Distribution")
    st.pyplot(fig)

    st.subheader("Boxplot for Price")
    fig, ax = plt.subplots()
    sns.boxplot(x='price', data=Data_frame, ax=ax)
    plt.title("Price Boxplot")
    st.pyplot(fig)
```

Figure 4.5: Pseudo code for histogram

### 4.4.6 Pseudo code for scatter chart

This code generates a Scatter Chart for the "Scatter Chart Analysis" page in a Streamlit web app. It allows users to select a location from a sidebar dropdown and displays a scatter plot with two series: one for 2 BHK properties in blue and another for 3 BHK properties in green with a '+' marker. The x-axis represents the total square feet area, and the y-axis represents the price in Lakh Indian Rupees. The chart provides a visual representation of property prices based on their square footage in the selected location.

```
# Scatter Chart
elif page == "Scatter Chart Analysis":
    st.title("Scatter Chart Analysis")
    selected_location = st.sidebar.selectbox("Select Location", Data_frame['location'].unique())
    bhk2 = Data_frame[(Data_frame.location == selected_location) & (Data_frame.BHK == 2)]
    bhk3 = Data_frame[(Data_frame.location == selected_location) & (Data_frame.BHK == 3)]

    fig, ax = plt.subplots()
    ax.scatter(bhk2.total_sqft, bhk2.price, color='blue', label='2 BHK', s=50)
    ax.scatter(bhk3.total_sqft, bhk3.price, marker='+', color='green', label='3 BHK', s=50)
    plt.xlabel("Total Square Feet Area")
    plt.ylabel("Price (Lakh Indian Rupees)")
    plt.title(f"Scatter Chart for {selected_location}")
    plt.legend()
    st.pyplot(fig)
```

Figure 4.6: Pseudo code for scatter chart

# Chapter 5

# Results

### 5.0.1 Home Page

The home page of the House Price Prediction and Visualization web application serves as an entry point, welcoming users and providing an overview of the project. It introduces the app's purpose, utilizing machine learning for Bengaluru house price predictions. Users are encouraged to navigate using the sidebar to explore various analysis and visualization pages.

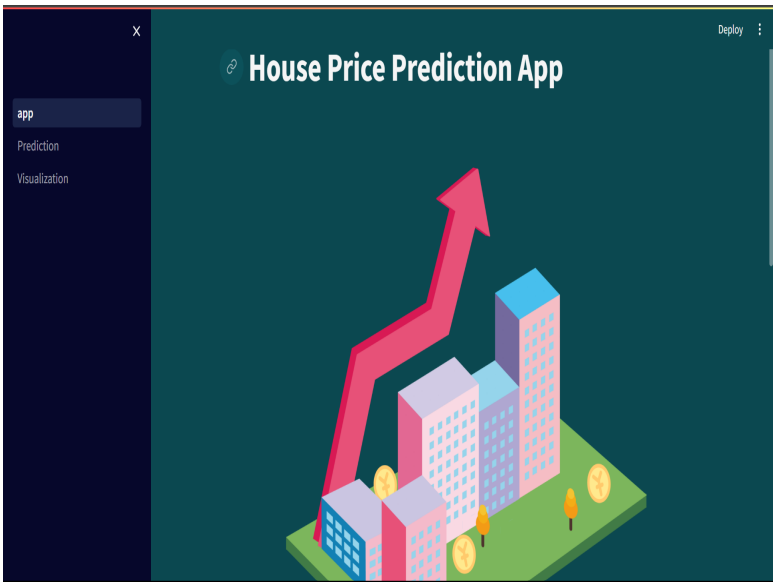


Figure 5.1: Home page

### 5.0.2 Prediction Page

The prediction page of the House Price Prediction and Visualization web application allows users to input data parameters. The Linear Regression model, trained on a cleaned dataset, provides real-time predictions based on user inputs. The page also displays visualizations, including a regression plot and a distribution plot comparing actual and predicted values for model evaluation.

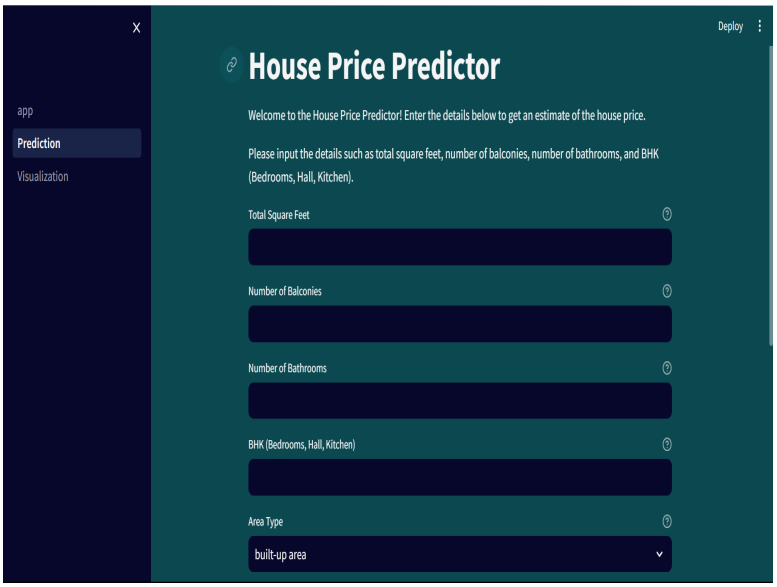


Figure 5.2: Prediction page

### 5.0.3 Visualization Page

The Visualization page in the House Price Prediction and Visualization web app offers diverse analyses. Users can explore scatter charts, histograms, box plots, and more to gain insights into factors like balcony count, total square feet, bathroom distribution, price variations, BHK distribution, and a heatmap revealing correlations.

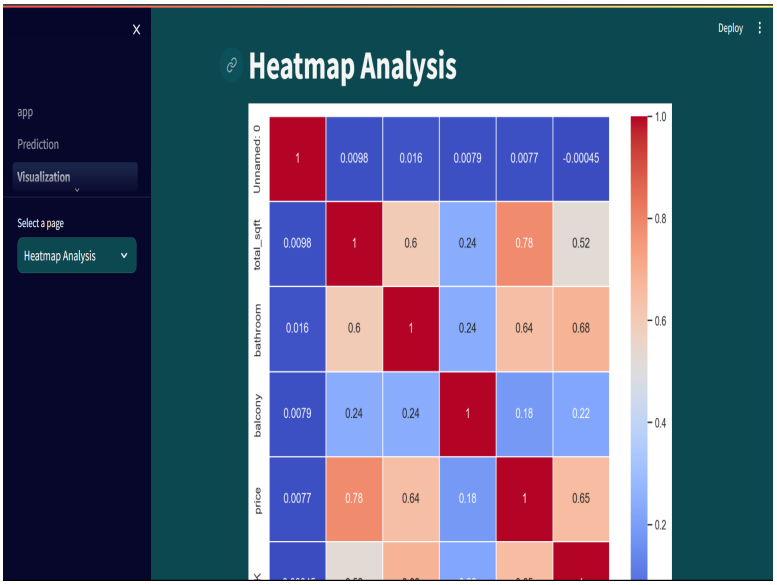


Figure 5.3: Visualization page

# Chapter 6

# Conclusion

The housing price prediction data visualization project emerges as a valuable asset in the realm of real estate analytics. Through the amalgamation of tools like Streamlit and the potential integration of a linear regression model, this project provides a dynamic and accessible platform for users to unravel the complexities of property valuation. The interactive interface, adorned with charts, graphs, and maps, offers a visual journey into the intricate relationships between diverse factors such as square footage, bedrooms, bathrooms, and geographical location. This not only empowers potential buyers and sellers with informed insights but also aids real estate professionals in navigating the nuanced market trends. Beyond its immediate applications, the project showcases the transformative potential of data visualization in democratizing information. It bridges the gap between raw data and actionable insights, fostering a more transparent and informed decision-making process. As the real estate market continues to evolve, the adaptability and scalability of this project make it a cornerstone for future developments in predictive modeling and data-driven analytics. In essence, the housing price prediction data visualization project not only elevates user experience but also contributes significantly to the broader landscape of real estate intelligence. By promoting transparency, accessibility, and understanding, it stands as a testament to the potential of data-driven tools in shaping the future of the real estate industry.

# References


[1] Predicting House Price in India Using Linear Regression Machine Learning Algorithms - Ankush Verma, Chetan Nagar, Nikita Singhi, Nupur Dongariya, Nidhi Sethi.

[2] Prediction of House Price Using Machine Learning Algorithms - G Kiran Kumar, D Malathi Rani, Neeraja Koppula, Syed Ashraf.

[3] House Price Prediction Using Linier Regression - Noviyanti T M Sagala, Laura Hestia Cendriawan.

[4] House Price Prediction Using Regression Techniques: A Comparative Study - CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha.

[5] House Price Prediction using regression techniques - Rupam Dwivedi, Rishabh Gupta, Prashant Kumar Pal.

[6] Real Estate Price Prediction using Machine Learning - S P Sreeja, V Asha, Binju Saju, Nivash CP, P Om Prakash, Adarsh Kumar Singh.

[7] House Price Prediction using Machine Learning Algorithm - Shailendra Sharma, Deepti Arora, Gori Shankar, Priyanka Sharma, Vihaan Motwani.