Industrial Training Report (Summer Internship)

Presentation

On

Housing Price Prediction During Covid

Submitted To	Submitted By	Supervised By
Mr. Basudeo Singh Roohani	Shaurya Seth	Ms. Vandna Tomar
Ms. Shanu Priya Chauhan	2101430100162	AP, CSE Deptt.
(Internship Coordinators)	7th Sem, CSE-4	

Department of Computer Science & Engineering
IMS Engineering College, Ghaziabad, India
(2024-25)

Outline & Contents

1. Introduction
   - ➢ Company Profile
2. Requirement Analysis
3. Problem Statement & Objectives
4. System Requirement Specification
   - ➢ H/W requirement specification
   - ➢ S/W requirement specification
5. Technology Used/Methodology
6. Result
7. Conclusion
8. Reference

Company's Introduction

- **Froyo Technologies Private Limited (FT)** is a global leader in skills and talent development, offering diverse learning solutions for individuals, professionals, institutions, and government entities. Founded in 2017 and headquartered in Noida, Uttar Pradesh, FT operates in over three countries, serving clients across various industries.

- FT provides comprehensive training in IT, data analytics, cybersecurity, finance, and soft skills, including industry-recognized certifications, corporate training, and e-learning platforms.

- Our global presence and continuous innovation have earned us industry recognition and numerous accolades. At Froyo Tech , we are committed to customer satisfaction, delivering customized solutions that exceed expectations.

Requirement Analysis

1. **Data Collection and Preprocessing:-**

   ➢ **Data Sources:-** Various sources provide valuable data, including price range, Facilities. This data serves as the foundation for predicting the models

   ➢ **Data Transformation:-** Data transformation often involves scaling and encoding. This converts data into a format suitable for machine learning algorithms, improving model performance.

   ➢ **Data Cleaning:-** Data cleaning is essential for accurate predictions. It involves addressing missing values, inconsistent data formats, and outliers. This step ensures data quality and reliability.

2. Data Analysis:-

   • Collecting and analyzing relevant data

   Features :-

• City, Location, BHK, Security, School, Hospital, Parking, Wifi, ATM, Power Backup Price (Target Variable)

3.Modeling Approach-

•Model Selection:-

1. Start with baseline models (e.g., linear regression) to set a performance benchmark.

2. Explore advanced models such as Decision Trees, Random Forests.

•Model Evaluation:-

1. Split the dataset into training, validation, and test sets.

2. Use evaluation metrics like Mean Absolute Error (MAE), Mean

3. Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared to assess model performance.

•Deployment:-

➢ Choose a deployment strategy (e.g., batch processing, real-time predictions via API).

➢ Use platforms like AWS, Azure, or Google Cloud for deployment.

Problem statement & Objectives

- The objective of this project was to analyze the impact of the COVID-19 pandemic in 2019 on housing prices and predict future prices by applying data science and machine learning techniques.

The Objectives are as follows:-

    a)   Data Cleaning and Pre-processing

    b)   Feature Engineering of Location

    c)   Model Development

    d)   Data Visualization

    e)   Model Evaluation and Optimization

System Requirement Specification

Hardware Requirement Specification:-

1.Computer (Desktop or Laptop):

• Processor (CPU):
  - ➢ **Minimum:** Dual-core processor (e.g., Intel Core i3, AMD Ryzen 3).
  - ➢ **Recommended:** Quad-core or higher (e.g., Intel Core i5/i7, AMD Ryzen 5/7)

•Memory (RAM):

  - ➢ **Minimum:** 8 GB.

  - ➢ **Recommended:** 16 GB or more.

•Display:

➢**Minimum:** Full HD (1920x1080) resolution.

➢**Recommended:** 4K (3840x2160) resolution or dual monitors .

- **Internet Connection:**
  - ➤ **Minimum:** 10 Mbps download/upload speed
  - ➤ **Recommended:** 50 Mbps or higher with low latency
- **Test Devices:-**
  - ➤ Physical mobile devices (smartphones, tablets) can be valuable for testing responsive designs and user experience on different platforms.
- **Storage:-**
  - ➤ **Minimum:** 2 GB SSD

Software Requirement Analysis
1. **Operating System:**
- **Windows, macOS, or Linux**:
  - ➤ Choose an OS that are comfortable with and that supports the development tools you intend to use.
  - ➤ I go through the Windows because I already worked and familiar with it.

2.Data Science Tools:-
- ➢ **Jupyter Notebook:-** Essential for interactive coding, data analysis, and visualization.
- ➢ **Python:-** Programming languages commonly used in data science for analysis and model development.

3.Web Browsers:-
- ➢ **Chrome, Firefox, Safari, Edge:**Popular browsers for testing and debugging websites. I prefer Chrome which offer powerful developer tools for inspecting and debugging HTML, CSS, and JavaScript.

4.Code Editor / Integrated Development Environment (IDE):-
- ➢ **Jupyter Notebook**:- Used for writing and executing code, visualizing data, and documenting the analysis process.Frameworks

Technology Used/Methodology

Data Science and Analytics:-

•**Python:-**

  ➢ **Purpose**:- Purpose: A versatile programming language widely used for data analysis, statistical computations, and building machine learning models. Commonly used libraries include Pandas for data manipulation, NumPy for numerical computations, and Matplotlib/Seaborn for data visualization.

Data Handling and Storage:-

•**SQL Databases:-**

  ➢ **Purpose:** Used for managing and querying relational databases. Commonly used databases include MySQL, PostgreSQL, and SQLite.
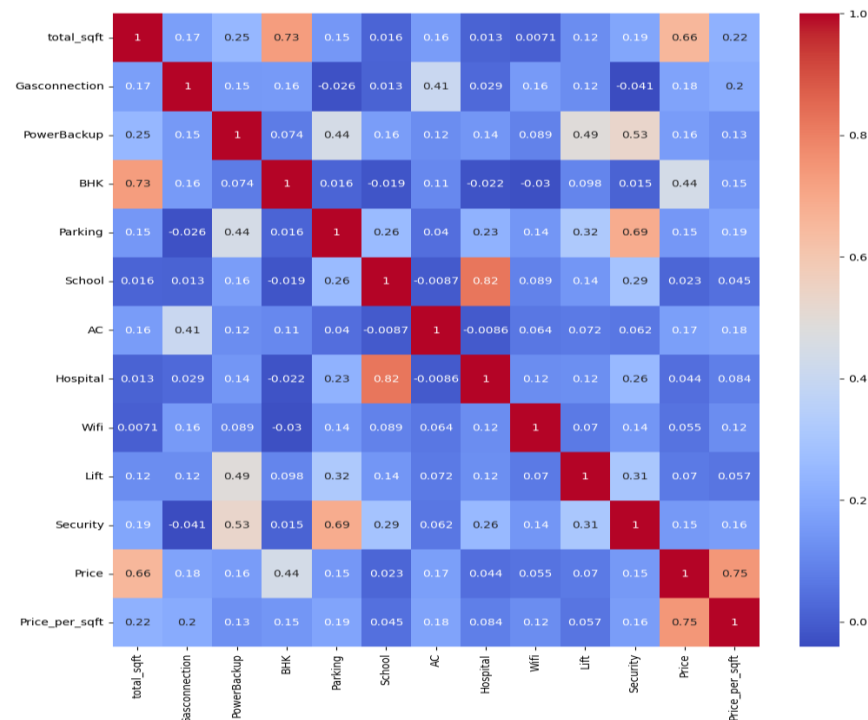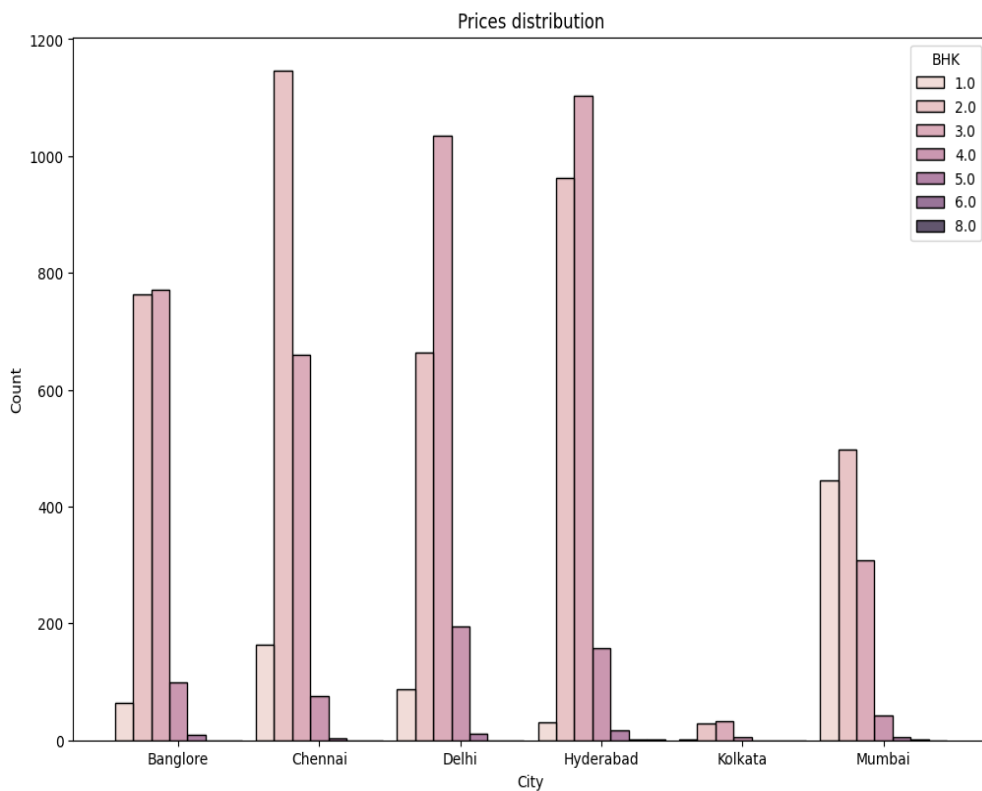
•**NoSQL Databases:-**

  ➢ **Purpose:-** Used for handling unstructured or semi-structured data. Examples include MongoDB for document-oriented storage and Redis for in-memory data structure store.

Methodology

Data Collection

Data Cleaning

Data Analysis

Model Building

Model Evaluation

Deployment
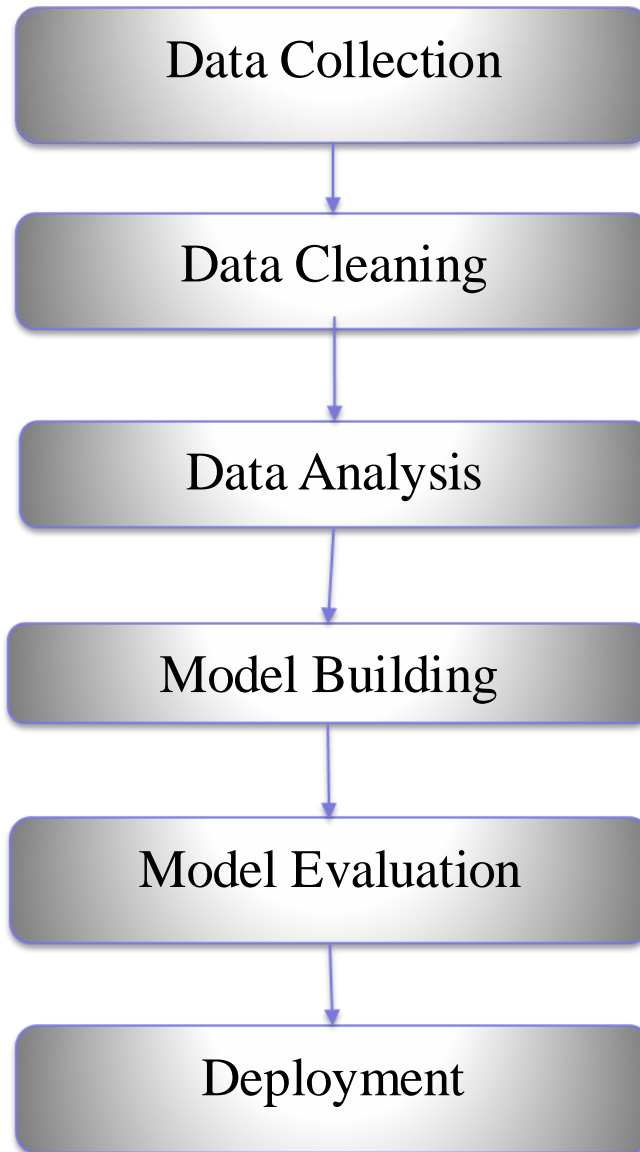
Implementation

```python
App.py > ...
 1  import streamlit as st
 2  import pandas as pd
 3  import numpy as np
 4  import matplotlib.pyplot as plt
 5  import seaborn as sns
 6
 7  # Load your datasets
 8  file_path = 'Indian House Prices.csv'
 9  DF = pd.read_csv(file_path)
10
11  # Rename 'No. of Bedrooms' to 'BHK' for consistency
12  DF.rename(columns={'No. of Bedrooms': 'BHK'}, inplace=True)
13
14  # Set a custom color palette for consistent visuals
15  sns.set_palette("Set2")
16
17  # Streamlit layout and title
18  st.markdown("<h1 style='text-align: center; color: #4CAF50;'>Housing Price Analysis</h1>", unsafe_allow_html=True)
19
20  # Display column names to verify
21  st.markdown("<h3 style='color: #FF6347;'>Column Names in the Dataset</h3>", unsafe_allow_html=True)
22  st.write(DF.columns)
23
24  # Check the first few rows of the data to verify structure
25  st.markdown("<h3 style='color: #1E90FF;'>Data Overview</h3>", unsafe_allow_html=True)
26  st.write(DF.head())
27
28  # --- Function to remove outliers based on City and Price ---
29  def remove_city_outliers(DF):
30      data_new = pd.DataFrame()
31      for key, subdata in DF.groupby('City'):
32          mean = np.mean(subdata['Price'])  # Use 'Price' instead of 'Price_per_sqft'
```

Result

This is the final result of the development of Chart and Streamlit UI
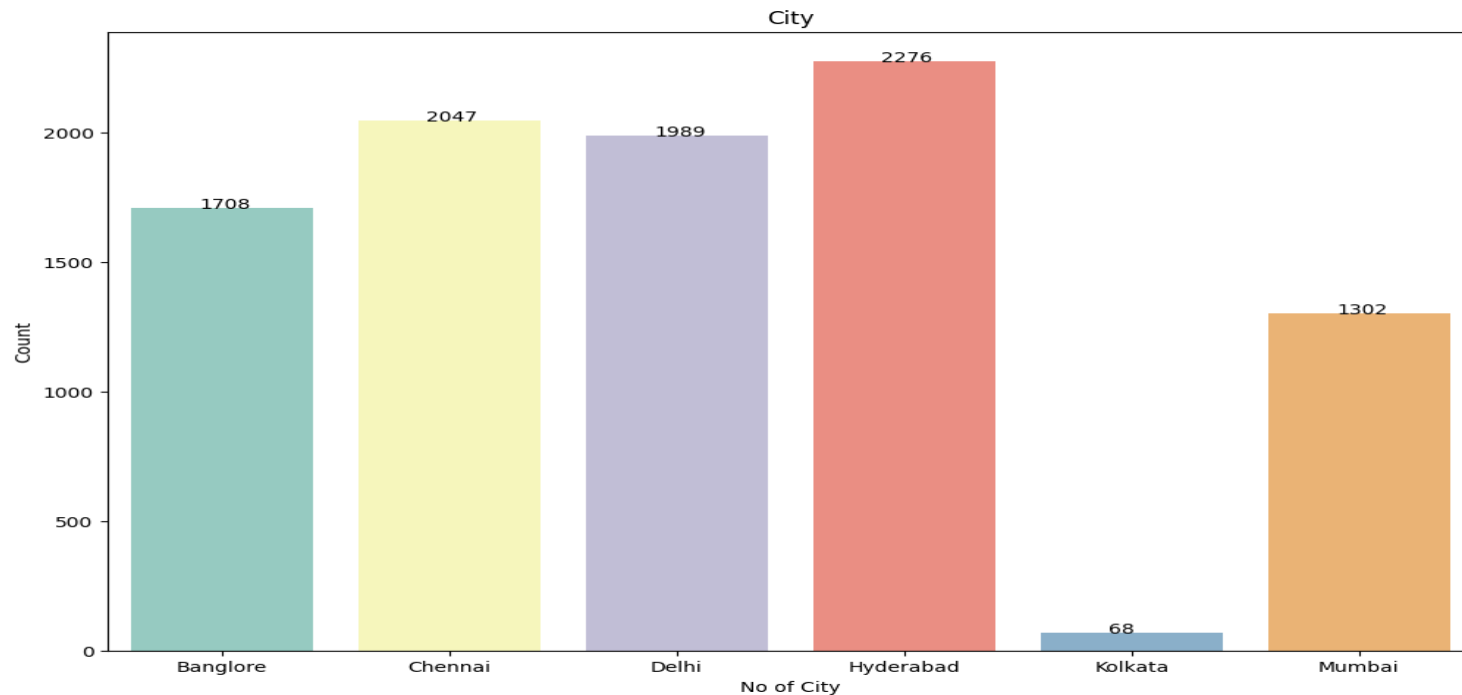
Conclusion

Through a structured and data-driven methodology, we successfully analyzed the effects of COVID-19 on housing prices and constructed predictive models to forecast future trends. Our notable achievements included the development of detailed location-based features and the optimization of machine learning algorithms, resulting in reliable and actionable predictions. Among the evaluated models, **Random Forest Regression** stood out as the most effective, achieving the lowest mean squared error (MSE) among our tests. This project fulfills its initial objectives but also establishes a solid groundwork for future enhancements as new data emerges and market dynamics continue to shift.

References

- [1] A. Gupta et al., "House Price Prediction Using Machine Learning Techniques," Int. Res. J. of Modernization in Engineering Technology and Science, vol. 3, no. 5, 2024.

- [2] R. Sharma et al., "Predicting Real Estate Prices Using Random Forest Regression," Int. J. of Adv. Research in Computer and Comm. Engg., vol. 10, no. 2, 2024.

- [3] M. Verma et al., "Automated House Price Prediction Model Using Data Science," E3S Web of Conferences, vol. ICMPC 2024, 01053, 2024. DOI: 10.1051/e3sconf/202443001053.

- [4] S. Patel et al., "Real-Time House Price Prediction Using Streamlit and Machine Learning," AIP Conf. Proc., vol. 2555, 020018, 2024. DOI: 10.1063/5.0108928.

Thank you!!