HOUSE PRICE PREDICTION AND DASHBOARD

SUMMER INTERNSHIP MAJOR PROJECT

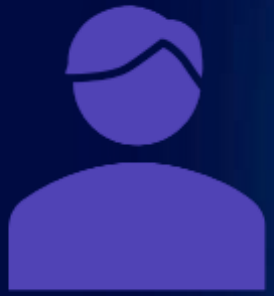
ANUSHKA PALO, ASHOK SWAIN, CH.AYUSHMAN PATRO
BBSP012406072665, BBSP012406072668, BBSP012406072669
Under the Guidance of Dr. Ritu Maity & Gitanjali Digha

DATA ANALYTICS
CENTRAL TOOL ROOM AND TRAINING CENTRE,
BHUBANESWAR, ODISHA
GOVERNMENT OF INDIA SOCIETY,
MINISTRY OF MSME
June-July 2024

TEAM DETAILS

ASHOK SWAIN
BBSP0124060726668

ANUSHKA PALO
BBSP0124060726665



CH. AYUSHMAN PATRO
BBSP0124060726669

CONTENTS

I. INTRODUCTION
II. PROJECT OVERVIEW
III. PROBLEM STATEMENT
IV. DESCRIPTION
V. SOFTWARE REQUIREMENTS
VI. DATA EXPLORATION
VII. APPROACH
VIII. ADVANTAGES & DISADVANTAGES
IX. DASHBOARD & PREDICTION
X. CONCLUSION

INTRODUCTION

The housing market is a critical aspect of any economy, influencing financial stability and personal wealth. Understanding the factors that drive house prices is essential for buyers, sellers, investors, and policymakers. This project leverages a comprehensive dataset of residential property prices in India to analyze and predict house prices using statistical and machine learning techniques.

PROJECT OVERVIEW

The goal of this project is to analyze the factors influencing residential property prices in India and develop a predictive model to estimate house prices based on various features. The project involves data exploration, preprocessing, model building, and visualization using tools like Tableau to gain insights and present findings.

PROBLEM STATEMENT

The primary objective of this project is to predict the sale prices of residential properties in India based on various features of the properties. The project aims to answer the following questions:

1. What are the key factors influencing house prices in India?
2. How accurately can we predict house prices using the available data?

DESCRIPTION

This dataset contains detailed information on residential property prices in India, including:

- Property characteristics (e.g., number of bedrooms, bathrooms, living area, lot area, number of floors)
- Location features (e.g., waterfront presence, number of views, geographical coordinates)
- Quality indicators (e.g., condition and grade of the house)
- Temporal aspects (e.g., year built, year renovated)
- Neighborhood factors (e.g., number of schools nearby, distance from the airport)
- Target variable: Price

SOFTWARE REQUIREMENTS

1. PYTHON (V3.11-V3.12)
2. JUPYTER NOTEBOOK
3. TABLEAU 2019.4
4. MS EXCEL
5. PYTHON LIBRARIES:-
    I. NUMPY
    II. PANDAS
    III. MATPLOTLIB
    IV. SEABORN
    V. SCI-KIT LEARN

DATA EXPLORATION

Data exploration involves:

1. **Descriptive Statistics:** Summary statistics for numerical and categorical features.
2. **Correlation Analysis:** Identifying relationships between features and the target variable (Price).
3. **Visualization:** Using plots (e.g., histograms, scatter plots, box plots) to understand the distribution and relationships of features.

APPROACH

1. **Data Preprocessing:** Handling missing values, encoding categorical variables, and normalizing numerical features.
2. **Feature Engineering:** Creating new features and selecting important ones for the model.
3. **Model Building:** Developing and training a Linear Regression model for price prediction.
4. **Model Evaluation:** Assessing model performance using metrics like R-squared and Mean Absolute Error.
5. **Dashboard Creation:** Using Tableau to create an interactive dashboard to present insights and model results.

ADVANTAGES

01 COMPREHENSIVE DATASET

The dataset includes a wide range of features that can provide deep insights into the factors affecting house prices.

02 PREDICTIVE MODELLING

Ability to build robust predictive models to estimate house prices, aiding in decision-making for buyers, sellers, and investors.

03 VISUALIZATION

Use of Tableau for creating interactive dashboards to visualize trends and patterns.

04 MARKET INSIGHTS

The analysis provides valuable insights into market trends and regional variations in house prices, which can help investors and policymakers make informed decisions.

DISADVANTAGES

01 DATA QUALITY

Potential issues with data quality, such as missing or inconsistent values, may affect model accuracy.

02 REGIONAL SPECIFICITY

The dataset is specific to India, which may limit the generalizability of findings to other regions.

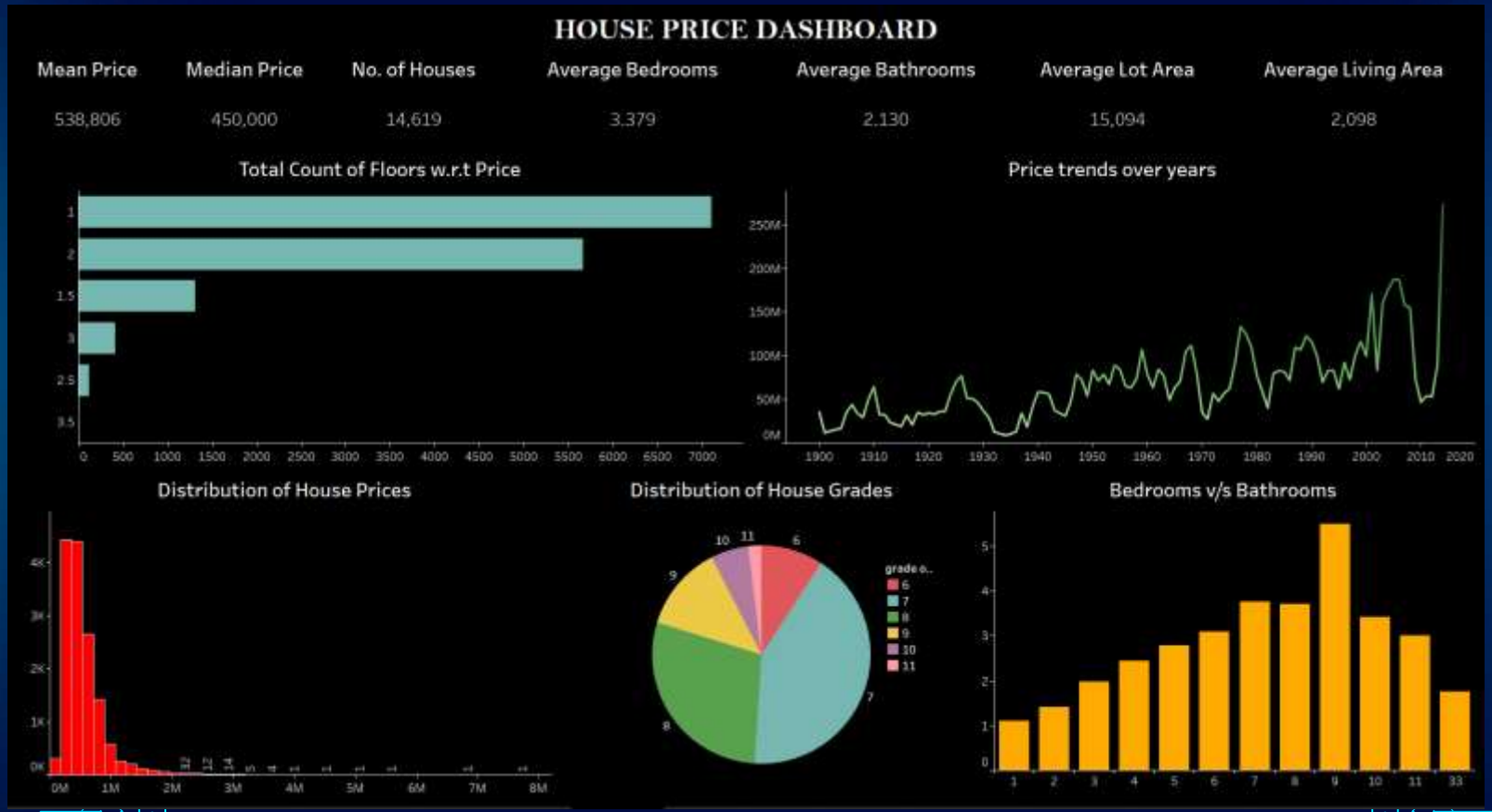
03 MODEL LIMITATIONS

Achieving high accuracy in predictive models can be challenging due to the complexity of the housing market.

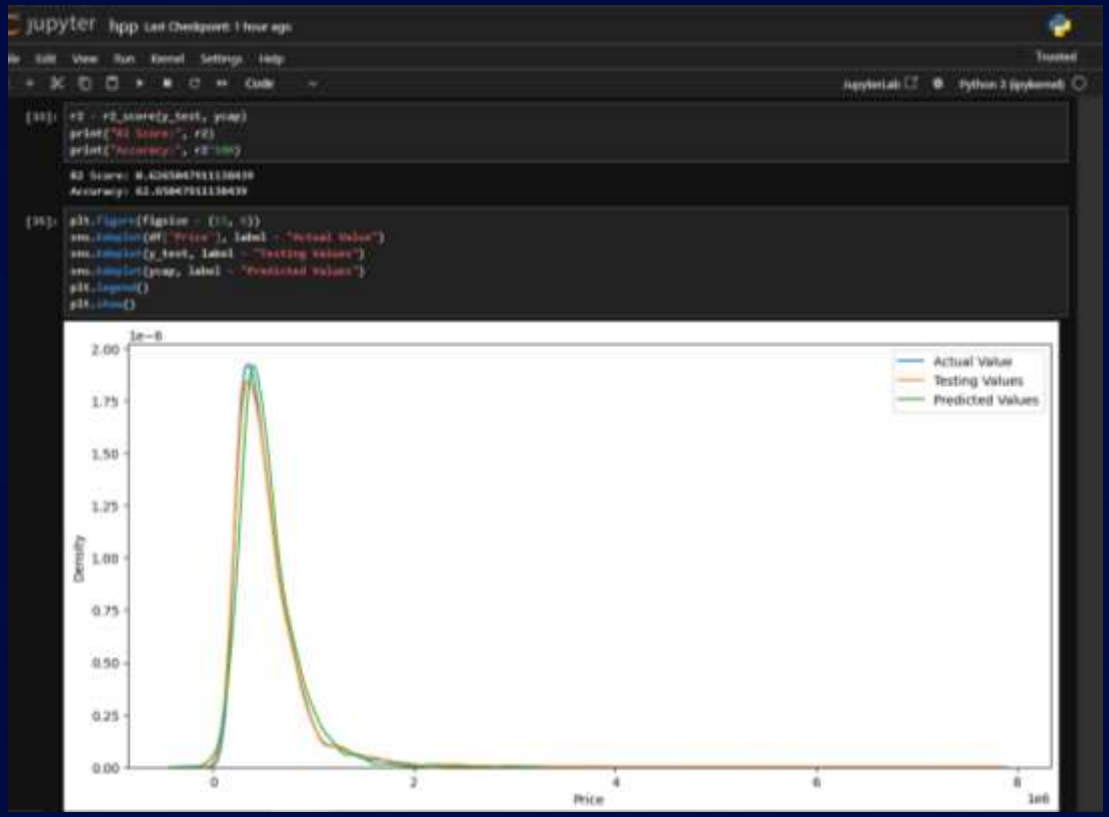
04 DATA COMPLEXITY

The complexity and high dimensionality of the dataset may lead to over-fitting in predictive models, requiring careful feature selection and regularization techniques.

DASHBOARD


HOUSE PRICE DASHBOARD
Mean Price
Median Price
No. of Houses
Average Bedrooms
Average Bathrooms
Average Lot Area
Average Living Area
538,806
450,000
14,619
3.379
2.130
15,094
2,098
Total Count of Floors w.r.t Price
Price trends over years
Distribution of House Prices
Distribution of House Grades
Bedrooms v/s Bathrooms

PREDICTION



R2 SCORE : 0.62

CONCLUSION

- The project successfully identified key factors influencing house prices and developed a predictive model with an accuracy of 60-65%.
- The interactive dashboard created in Tableau provides a comprehensive view of the analysis, allowing users to explore different aspects of the housing market in India. While the model's accuracy suggests room for improvement, the insights gained from this project can guide further research and model refinement.

Thanks!!!