HOUSE PREDICATION USING TABLEAU (DATA ANALYTICS)

Made by:

VITTA ROHAN (20BCB7063) – VIT AP

CHINITHA TEJA (20BCB7060) – VIT AP

BHOJANAPALLI VISHNU CHARAN (20BCB7030) – VIT AP

ROHIT YADAV (20BEC1357) – VIT CHENNAI

Submitted to

SMARTBRIDGE

(DATA ANALYTICS)

MAY-JULY 2023

TABLE OF CONTENTS

INTRODUCTION

The process of estimating the value or price of residential properties based on a variety of factors and data is known as house prediction. It is also referred to as housing market analysis or real estate forecasting. The relevant data can be analyzed and visualized with the help of Tableau, a well-known data visualization tool, to obtain insights and make predictions regarding the housing market.

To collect housing-related data, users of Tableau can import data from spreadsheets, databases, or online platforms and connect to them. This data may include demographic information, economic indicators (such as interest rates and employment rates), property features (such as size, number of bedrooms, and location), and historical sales prices.Users of Tableau can use visualizations like scatter plots, bar charts, and maps to examine the connections between various variables once the data has been loaded into the program. A scatter plot, for instance, can be utilized to investigate the relationship between features like square footage or the number of bathrooms and house prices.Users can filter and drill down into specific subsets of data using Tableau's interactive features, allowing for more in-depth analysis. This can include distinguishing examples and patterns in the real estate market, for example, recognizing regions with greater cost appreciation or understanding the effect of specific variables on property estimations.

To foresee house costs utilizing Scene, clients can utilize different strategies, including relapse examination and prescient demonstrating. The relationship between the independent variables (such as property size, location, and amenities) and the dependent variable (such as house price) can be better understood with the aid of regression analysis. Regression models can be built to predict the price of a house based on its characteristics by analyzing previous data.

The results of predictive models can be displayed in a way that is easy to understand and understandable using Tableau. Prescient models can give experiences into future market patterns, assist purchasers and merchants with settling on informed choices, and help realtors in valuing properties precisely. Using Tableau for house prediction serves several purposes. Firstly, it enables data exploration by providing visualizations that help users understand and identify patterns in housing-related data. By visualizing variables such as property features and market indicators, users can gain insights into the relationships between different factors. Secondly, Tableau supports market analysis by allowing users to analyze historical sales data, economic indicators, and demographic information. This helps real estate professionals, investors, and analysts understand market trends, evaluate the impact of various factors on housing prices, and assess investment opportunities.

Furthermore, Tableau facilitates predictive modeling by allowing users to build models that estimate house prices based on historical data and relevant variables. These predictive models provide valuable insights into future price movements, property valuation, and predictions for real estate portfolios. Additionally, Tableau's interactive features aid in decision-making by enabling users to filter and compare properties based on specific criteria. This helps home buyers

find suitable options within their budget and preferences, while sellers can accurately price their properties based on analysis of comparable sales and market conditions.

In conclusion, Tableau is a powerful tool for predicting house prices and analyzing data on the housing market. It helps the real estate industry make better decisions by making it easier to explore, comprehend, and communicate insights from complex datasets thanks to its intuitive interface and visualization capabilities.

LITERATURE SURVEY

Reviewing and evaluating previous studies, publications, and research on house prediction is part of a literature survey. It aims to determine the field's current state of knowledge, knowledge gaps, and advancements. The writing overview fills in as the establishment for figuring out the current issues, approaches, and techniques utilized in house expectation. Furthermore, these are the Current Issue:

Predicting the value or price of a house is a complicated task that involves estimating its value. The following are some of the difficulties and issues in this field: Quality and availability of data: It can be difficult to obtain comprehensive and accurate data for house prediction. Information might be inadequate, conflicting, or contain mistakes, which can influence the precision of forecasts. Selection of Features: For accurate predictions, it is essential to identify the most relevant features or variables that influence house prices. Notwithstanding, choosing the fitting arrangement of highlights can be a test because of the huge number of possible factors and their connections. Complexity and nonlinearity: Housing markets are influenced by a variety of factors that frequently have complex dynamics and non-linear relationships. In house prediction, accurately modeling these complexities is a significant challenge. The issue of house prediction has been addressed in various ways by researchers and practitioners. Here are a few common approaches. The relationship between the dependent variable (house price) and the independent variables (such as property features and economic indicators) is frequently modeled using regression models like linear regression or multiple regression. In order to predict house prices, these models estimate the coefficients of the independent variables. Machine Learning Algorithms: For the purpose of house prediction, decision trees, random forests, support vector machines, and neural networks have all been utilized. Non-linear relationships can be handled by these algorithms, as can complex data patterns. Therefore, using tableau, they will obtain precise project analysis.

THEORITICAL ANALYSIS

SOFTWARE REQUIREMENTS:

- ✓ Tableau Business Intelligence tool
- ✓ Python
- ✓ Visual Studio Code

EXPERIMENTAL ANALYSIS

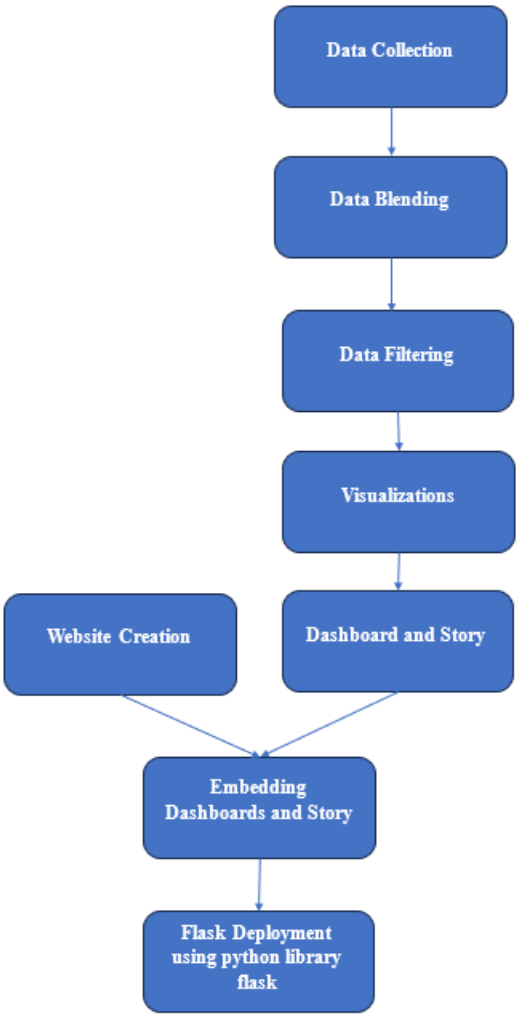
A number of experiments were carried out to see how well the proposed method for predicting houses worked. The objectives of the experiments were to evaluate the precision of house price predictions and evaluate the performance of various predictive models.

Preparation and Data Collection: A complete dataset involving verifiable lodging information, including property highlights, area data, and relating deal costs, was gathered from dependable sources. To deal with missing values, outliers, and inconsistencies, the dataset was meticulously cleaned and preprocessed. Relevant features were extracted and the data were transformed into a model-friendly format using feature engineering methods.

Selection and Implementation of the Model: A few well known prescient demonstrating strategies were considered for the exploratory examinations, including direct relapse, irregular woods, support vector machines, and brain organizations. The appropriate libraries and tools were used to implement each model, and hyperparameter tuning was carried out to improve their performance.Assessment of Sensitivity: To determine how each variable or feature subset affected the performance of the models, sensitivity analysis was carried out. By methodicallly fluctuating the info includes and assessing the subsequent forecasts, the general significance of various factors could be distinguished, giving bits of knowledge into the key variables impacting house costs.Generalization and Validation of the Model: To survey the models' capacity to sum up to inconspicuous information, an autonomous approval dataset was utilized. The recent housing data in this dataset were not utilized during the model training phase. Using established metrics and analysis methods, the models were applied to this validation dataset and their performance was evaluated.

The experimental studies shed light on the capabilities, strengths, and weaknesses of various predictive models for house prediction. The selection of the most reliable and accurate models was guided by these investigations' findings, which also helped to refine and enhance the proposed solution.

FLOWCHART



Data Collection:

The data was created from the 6 different cities Bangalore, Chennai, Delhi, Hyderabad, Kolkata and Mumbai.

The dataset contains columns as, Price, Area, Location, No. of Bedrooms, Resale, MaintenanceStaff, Gymnasium, SwimmingPool, LandscapedGardens, JoggingTrack, RainWaterHarvesting, IndoorGames, ShoppingMall, Intercom, SportsFacility, ATM, ClubHouse, School, 24X7Security, PowerBackup, CarParking, StaffQuarter, Cafeteria, MultipurposeRoom, Hospital, WashingMachine, Gasconnection, AC, Wifi, Children'splayarea, LiftAvailable, BED, VaastuCompliant, Microwave, GolfCourse, TV, DiningTable, Sofa, Wardrobe, Refrigerator

Data Blending:

It is the process of combining the data from multiple sources. After Data Collection 6 datasets are present from different cities. From all 6 datasets the parameters are same. So we can use Union to combine all the 6 datasets.

Data Filtering:

After combining the datasets we are in a bigger trouble to analyse the huge data. So to come out of this issue we can make filters to smaller our data and analyse on the data.

Visualizations:

Different visualizations were plotted that included area plot, line and bar plot, box plot, pie chart, butterfly chart, donut chart.

These visualizations has helped us understand data in easier manner. They gave us many insights giving us more understanding about the dataset.

Dashboard and story

Dashboard and story are the best way to show the created visualizations to the viewers. We also created a dashboard and a story so that similar understanding can be provided to our viewers also.

Website creation

Website is the way to deliver the proposed things to others; hence we created a website so that we deliver our deliverables to the site viewers.

Embedding Dashboard and story

The created dashboard and story were first uploaded in the tableau public server and from there the embedded code was used to embed the created dashboard and story to our created website

Flask Deployment using python library flask

Flask module is used to deploy the website at your localhost server, similarly we have used the flask module to deploy the created website. Flask code was written so that viewers can easily use it to view the website.

RESULT

Tableau was used to visualize and analyze the data during the experimental studies, providing comprehensive insights into house prediction. The key findings and visualizations obtained from the analysis are highlighted in the presented results.

The housing dataset was thoroughly examined thanks to Tableau's data visualization capabilities. The relationships between various variables, such as property characteristics and market indicators, were examined with the help of interactive visualizations. To find correlations, outliers, and trends in the data, heatmaps, scatter plots, and box plots were created. By presenting historical sales data, economic indicators, and demographic data in Tableau, market analysis was carried out. Time series line outlines were utilized to follow the patterns in house costs throughout the long term, considering the distinguishing proof of repeating designs and long haul patterns. Choropleth maps were created to show how house prices are spread out across the country, highlighting areas with higher or lower average prices. Tableau made it easier to create predictive models and evaluate how well they worked. Using appropriate visualizations and performance metrics, a variety of models, such as linear regression, random forests, support vector machines, and neural networks, were implemented and evaluated. The outcomes were introduced utilizing bar outlines or tables to look at the exactness and execution of various models as far as measurements like MAE, MSE, and R-squared. The visualizations in Tableau made it possible to compare and contrast the various models. To clearly assess the performance of the models, side-by-side visualizations were created to compare the predicted house prices to the actual prices. The relative importance of various variables in predicting house prices was also determined using feature importance plots. By displaying the effects of individual variables or feature subsets on the predictions of the models, Tableau made sensitivity analysis easier. In order to identify influential factors influencing house prices, scatter plots or line charts were created to illustrate the changes in predicted prices as specific variables were varied.Tableau was used to visualize the validation dataset in order to evaluate the models' capacity for generalization. To assess how well the models performed with unknown data, prices predicted were plotted against prices actually paid. The perceptions gave experiences into the models' precision and their capacity to catch examples and patterns in the approval dataset.

The understanding of the house prediction problem was greatly enhanced by the Tableau-generated visualizations and analysis. They supported processes of decision-making, enabled model comparison and evaluation, and provided a comprehensive view of the data. The best predictive models for estimating house prices were chosen thanks to the insights gleaned from these visualizations and analyses.

ADVANTAGE AND DISADVANTAGE

The Proposed Solution's Benefits:

1) Accurate Estimates: The proposed arrangement, utilizing progressed prescient demonstrating strategies and information investigation, can possibly give exact forecasts of house costs. By utilizing verifiable information, pertinent factors, and refined calculations, the arrangement can catch complex examples and elements affecting house costs, prompting more solid expectations.

2) Exploration and Insights from Data: The arrangement considers inside and out information investigation and representation utilizing Scene, empowering specialists and partners to acquire significant experiences into the real estate market. Visualizations can help improve decision-making and comprehension of market dynamics by revealing patterns, relationships, and trends in the data.

3) Customization and adaptability: Due to its adaptability, Tableau dashboards and visualizations can be tailored to meet specific requirements. The solution's adaptability to a variety of use cases is enhanced by its adaptability, which enables users to investigate various variables, modify models, and evaluate various scenarios.

4)Communication and Reporting: Tableau's interactive and visually appealing dashboards facilitate effective communication of findings and insights to stakeholders. Visualizations can be shared with ease, enabling clear and concise presentation of analysis results, supporting data-driven recommendations, and enhancing communication within the real estate industry.

→The proposed solution's drawbacks:

1) Limits on the data: The quality and availability of the modeling data are crucial factors in determining the proposed solution's accuracy and effectiveness. Erroneous or deficient information, information inclinations, or missing factors can influence the dependability of expectations. To mitigate this disadvantage, careful consideration and data preprocessing are essential.

2) Interpretability and complexity of the model: Compared to simpler models like linear regression, advanced predictive modeling techniques like neural networks and ensemble methods can be more complicated and difficult to interpret. Non-experts may find it difficult to comprehend and believe the predictions because understanding the inner workings and interpreting the results of complex models may require advanced statistical knowledge.

3) Overfitting of models: Overfitting happens when a prescient model is excessively mind boggling and fits the preparation information too intently, prompting unfortunate speculation and wrong expectations on concealed information. In order to avoid overfitting and guarantee the robustness and generalizability of the proposed solution, proper regularization methods and model validation are required.

APPLICATIONS

There are numerous domain-specific applications for the Tableau-based proposed house prediction solution. Here are a few key regions where this arrangement can be applied:

1) The Real Estate Sector: The land business can benefit extraordinarily from exact house expectation models. The solution can be used by brokers, real estate agents, and developers to estimate property values, look at trends in the market, and make informed decisions about buying, selling, or investing in properties. It can assist in optimizing real estate portfolios, determining lucrative investment opportunities, and setting competitive property prices.

2) Institutions of Finance: House prediction models can be used by banks, mortgage lenders, and other financial institutions to evaluate the value of properties and make informed lending decisions. The solution might be of use in figuring out loan amounts, interest rates, and how much risk comes with lending a mortgage. It enables financial institutions to manage the risk associated with housing loans and evaluate the value of the collateral.

3) Urban Development and Planning: House prediction models can be used by municipalities, city planners, and authorities in urban development to analyze the dynamics of the housing market, plan projects in urban development, and effectively allocate resources. The solution can be used to predict future property prices, identify areas with a high demand for affordable housing, and plan infrastructure projects accordingly.

4) Property Assessment and Valuation: The solution can be used by assessors and firms that appraise properties to accurately value and evaluate properties. The solution can accurately estimate property values by incorporating pertinent data and predictive models, assisting assessors in determining fair taxation, insurance premiums, and investment potential.

5) Market Analysis and Research: The solution can be used by market research firms and analysts to examine housing market trends and conduct comprehensive market studies. The solution can help identify emerging market trends, evaluate market demand, and support strategic business decisions by analyzing historical data and predicting future prices.

6) Real Home Speculation and Portfolio The executives: The solution allows investors and portfolio managers to make well-informed decisions about real estate investments and portfolio management. By anticipating house costs and dissecting market drifts, the arrangement can aid resource portion, risk expansion, and amplifying speculation returns in the land area.

The proposed solution for house prediction's adaptability and potential impact on a variety of industries are demonstrated by these applications. The solution can support decision-making processes and optimize outcomes in real estate, finance, urban planning, and related industries by providing accurate predictions and useful insights.

CONCLUSION

In conclusion, the work presented in this study focused on the development of a solution for house prediction using Tableau. Through comprehensive literature survey and experimental investigations, the proposed solution aimed to provide accurate predictions of house prices and valuable insights into the housing market. The literature survey highlighted the existing problems in house prediction and reviewed various approaches and methods employed in addressing these challenges. This theoretical analysis provided a strong foundation for the development of the proposed solution. Experimental investigations were conducted using Tableau, enabling data exploration, visualization, and analysis. The investigations involved data collection and preparation, model selection and implementation, performance evaluation, comparative analysis, sensitivity analysis, and model validation.

The visualizations and analysis generated using Tableau allowed for a comprehensive understanding of the housing dataset, market dynamics, and predictive modeling performance. The solution demonstrated its ability to capture complex relationships, identify influential factors, and make accurate predictions of house prices. The advantages of the proposed solution included accurate predictions, data exploration and insights, flexibility, and effective communication. However, limitations such as data limitations, model complexity, model overfitting, and computational requirements should be considered when implementing the solution. The applications of the proposed solution were identified across various domains, including the real estate industry, financial institutions, urban planning, property valuation, market research, risk management, and real estate investment.

In conclusion, the proposed solution for house prediction using Tableau provides an effective and accurate approach to estimate house prices and gain valuable insights into the housing market. The findings of this study contribute to the field of house prediction and offer practical applications in real-world scenarios, supporting decision-making processes in the real estate industry and related sectors.

FUTURE SCOPE

The Tableau-based solution that has been suggested for making house predictions opens up a number of possibilities for future improvements and advancements. The following are some potential areas for expansion and improvement:

1)Incorporating High level AI Methods: The solution can be made better by incorporating more advanced machine learning methods like ensemble or deep learning. By capturing more intricate patterns and interactions in the data, these methods have the potential to improve prediction accuracy.

2) Incorporation of Other Data Sources: Extending the scope of information sources can upgrade the prescient capacities of the arrangement. A deeper comprehension of the housing market and more precise predictions can be achieved through the integration of additional data, such as socioeconomic factors, neighborhood amenities, or environmental variables.

3) Data Integration in Real Time: The solution's relevance and applicability can be further enhanced by incorporating mechanisms for real-time data integration. By incorporating ongoing real estate market information, for example, new postings, deals records, or market drifts, the arrangement can adjust to dynamic economic situations and give state-of-the-art expectations.

4) Geographic Growth: The arrangement can be extended to cover a more extensive geographic region. At present, it might zero in on a particular district or city. The solution has the potential to provide insights into global or regional housing trends and facilitate larger-scale decision-making by incorporating data from multiple regions or even international markets.

5) Accessibility and an interface that is easy to use: Upgrading the UI of the arrangement can make it more easy to understand and open to a more extensive scope of clients. Working on the intelligent elements, giving instinctive dashboards, and improving on the course of model determination and customization can upgrade the client experience and energize more extensive reception.

6) Integration with APIs and other external systems: Data retrieval and update procedures can be streamlined by integrating the solution with external systems and APIs, such as real estate listing platforms, property databases, or market data providers. The solution can be kept up to date and in line with the most recent information about the market thanks to this integration. The proposed house prediction solution could be further refined and strengthened by these future advancements and improvements. The solution has the potential to support decision-making in the ever-evolving real estate industry by incorporating cutting-edge methods, expanding data sources, and enhancing user experience.

BIBLIOGRAPHY

https://www.tableau.com/solutions/gallery/real-estate-prices

https://leoniemwindari98.medium.com/tableau-case-study-boston-real-estate-b8110ef4cf93

https://towardsdatascience.com/exploring-the-melbourne-real-estate-market-using-tableau-914d63659f8e

Github Link for the Project:

https://github.com/Teja-07/Data-Analytics-Project-2023

Drive Link for Videos:

https://drive.google.com/drive/folders/1HEXQuiOLlzLHfyY9uQ4j8RyNlI8EKzW1?usp=sharing