Experiment No. : 10

Date : **25/04/2024**

Title: MINI – PROJECT - BI Report

Group Members:
Abdullah Ali – 01
Ahmad Raza - 05
Swaraj Dhumal -14

Problem Definition: Real Estate Price Prediction for Banglore City

Software Used: Power BI and Google Colab :

short summary of system including

a) Data mining task needed

In real estate price prediction for Bangalore city, data mining tasks play a crucial role in extracting insights from various datasets. Here are some essential data mining tasks thatcould be employed:

1. Data Collection and Integration: Gather data from various sources such as property listings, historical sales data, demographic information, economic indicators, and geographical data. Integrate these diverse datasets into a unified format.

2. Data Cleaning and Preprocessing: Cleanse the data to handle missing values, outliers, and inconsistencies. Preprocess the data by standardizing formats, resolving discrepancies, and transforming variables if needed.

3. Feature Selection and Engineering: Identify relevant features that could impact real estateprices, such as location, property size, amenities, neighborhood characteristics, proximity tofacilities, and market trends. Engineer new features that may enhance predictive accuracy, like distance to

transportation hubs or schools.

Exploratory Data Analysis (EDA): Analyze the data to understand relationships, patterns,and trends. Use visualizations and statistical techniques to uncover insights about the real estate market dynamics in Bangalore.

1. Model Selection and Training: Experiment with various machine learning algorithms suchas linear regression, decision trees, random forests, gradient boosting, or neural networks. Select the most suitable models based on performance metrics like mean squared error or R-squared. Train the chosen models using historical data.

2. Model Evaluation: Assess the performance of trained models using techniques like cross- validation, train-test splits, or time-series validation. Compare the models based on their predictive accuracy, robustness, and generalization capabilities.

3. Hyperparameter Tuning: Fine-tune model hyperparameters to optimize performance further. Utilize techniques like grid search, random search, or Bayesian optimization to findthe best hyperparameter values.

4. Ensemble Methods: Explore ensemble learning techniques such as bagging, boosting, or stacking to combine predictions from multiple models and improve overall accuracy.

5. Model Interpretation: Interpret the trained models to understand the factors driving real estate prices in Bangalore. Analyze feature importance, partial dependence plots, and SHAP values to gain insights into the decision-making process of the models.

6. Deployment and Monitoring: Deploy the trained model into production systems for real- time predictions. Implement monitoring mechanisms to track model

performance over time and update the model periodically as new data becomes available.

7. *Datasets used*
8. Banglore House Price Data (kaggle.com)

b) *Snapshots showing implementation of the data mining algorithm of your choice and the results*

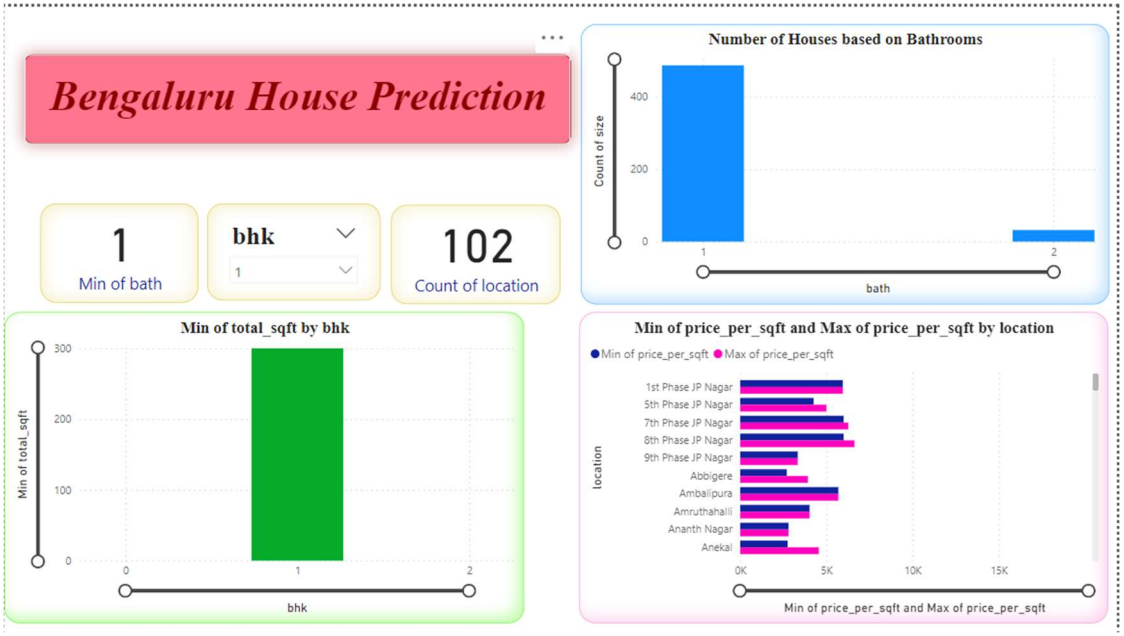

Fig. Overall dashboard 1

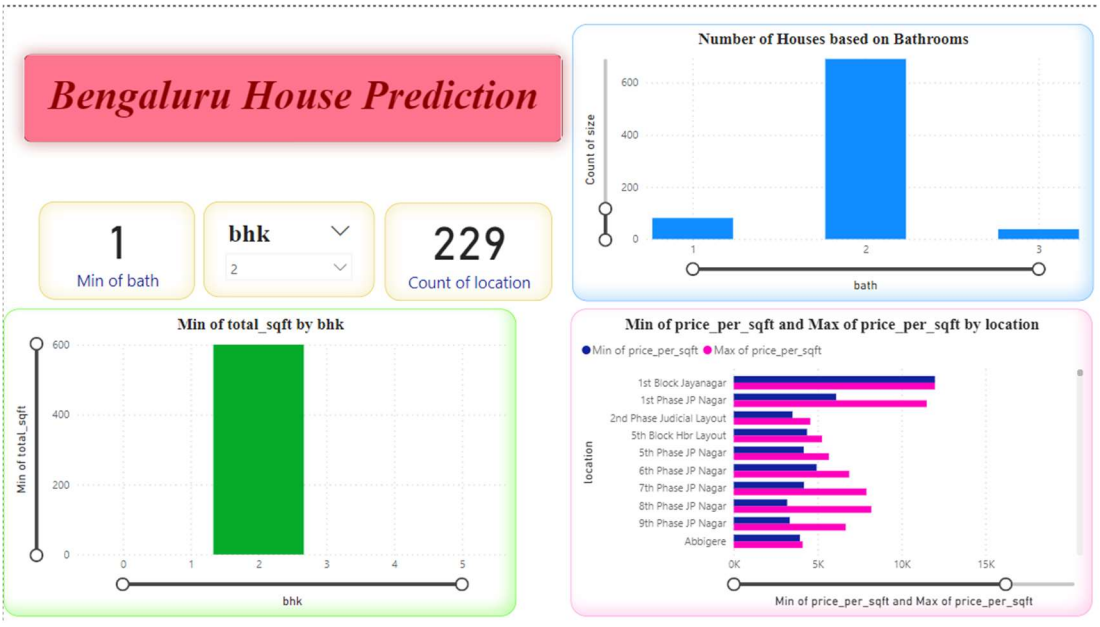

Fig 2. Overall dashboard 2

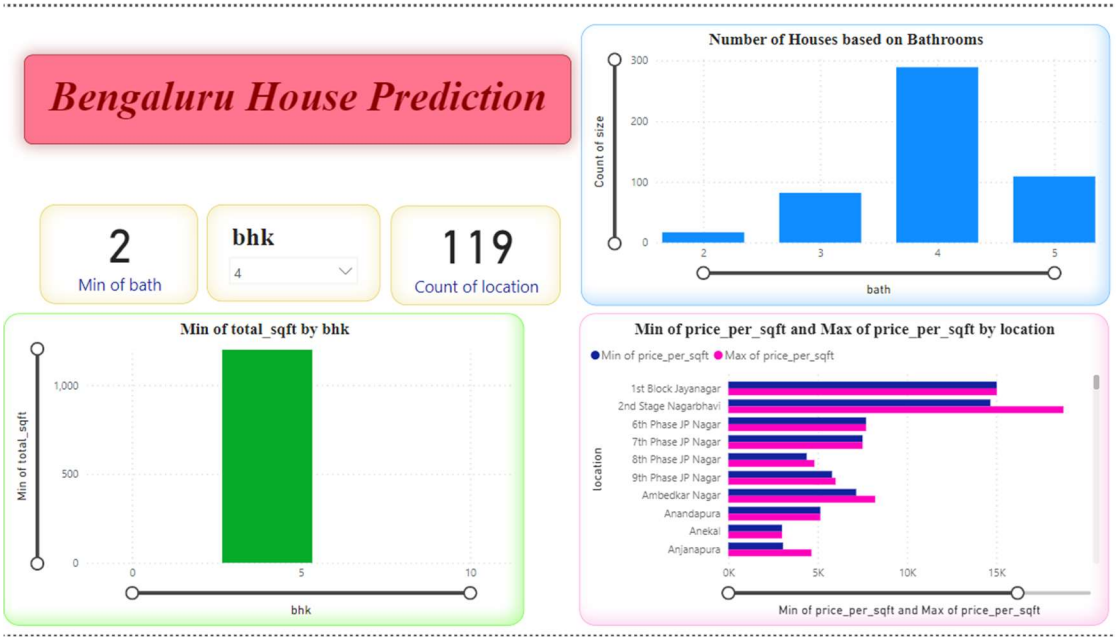

Fig 3. Overall dashboard 3

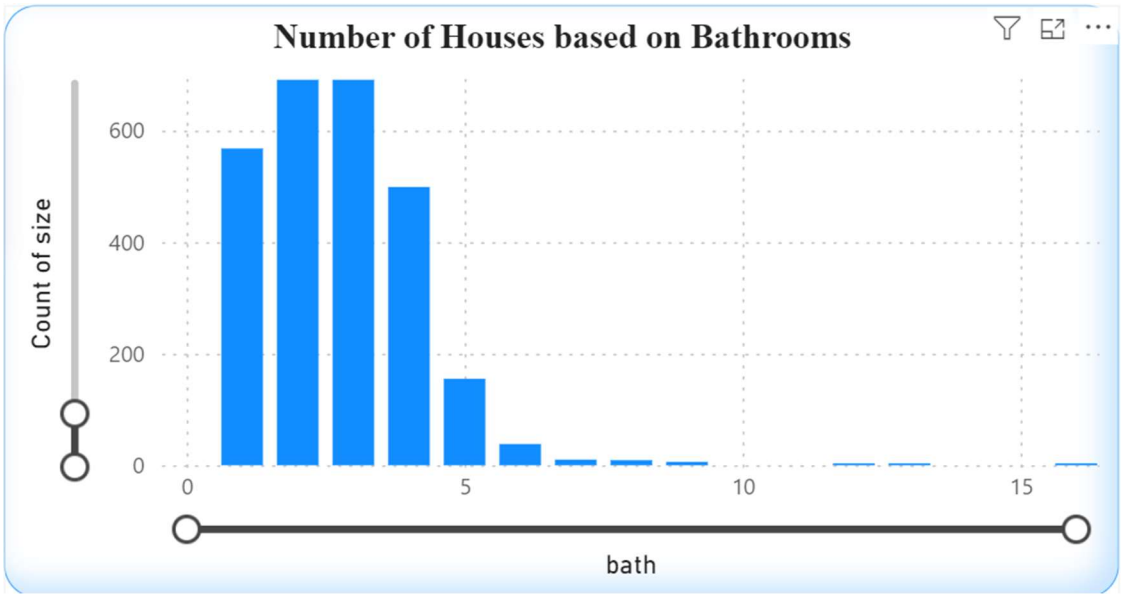

Fig 4. Overall bar Graph

The graph displays the count of houses based on the number of bathrooms they have.
The x-axis is labeled “bath” and ranges from 0 to over 15.
The y-axis is labeled “Count of size” and ranges from 0 to over 600.

There are six bars representing different bathroom counts:
The highest count is for houses with two bathrooms.
As the number of bathrooms increases or decreases, the counts decrease.

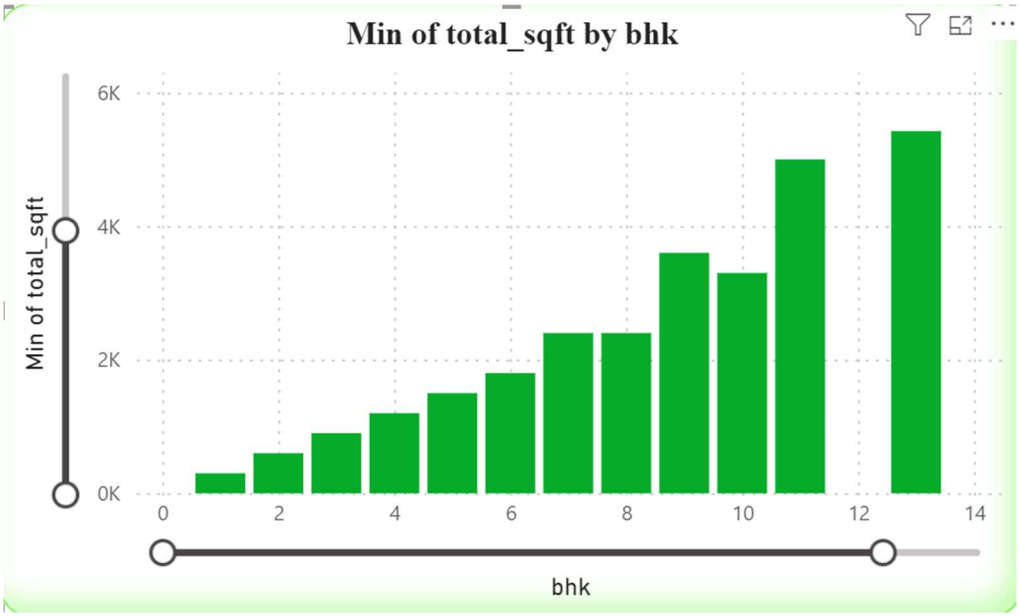

Fig 5. Overall bar graph

1. The graph displays the minimum square footage associated with different numbers of bedrooms (bhk).
2. The x-axis is labeled “bhk” and shows values from 0 to 14.
3. The y-axis is labeled “Min of total_sqft” and ranges from 0K to 6K.
4. There are green bars at each point on the x-axis, starting from ‘2 bhk,’ showing an increasing trend in the minimum square footage as the number of bhks increases.
5. Notably, there’s a significant jump in minimum total sqft at ‘12 bhk’ and ‘14 bhk.’

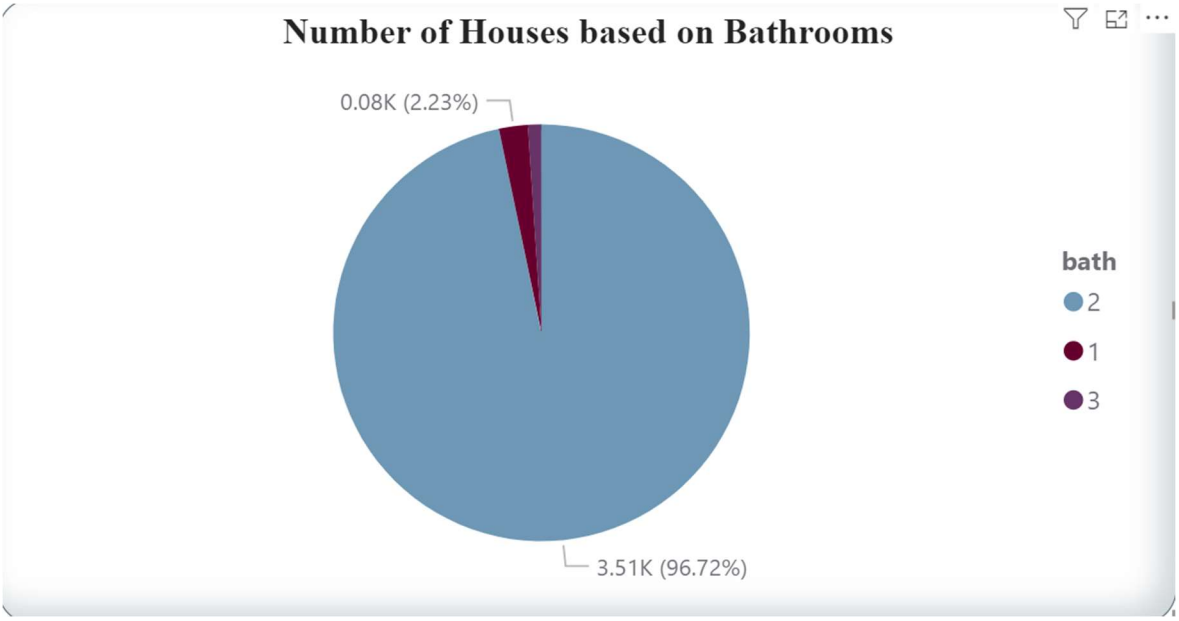

Fig 5. Overall Pie graph

1. The pie chart visualizes data on the number of houses categorized by the number of bathrooms they have.
2. The title “Number of Houses based on Bathrooms” is displayed at the top.
3. There are three categories represented:
    - **1 bathroom (blue)**: The largest section, accounting for 96.72% of houses (3.51K).
    - **2 bathrooms (red)**: A very small slice, representing 2.23% of houses (0.08K).
    - **3 bathrooms (purple)**: There is no visible section or data for houses with three bathrooms in this pie chart.

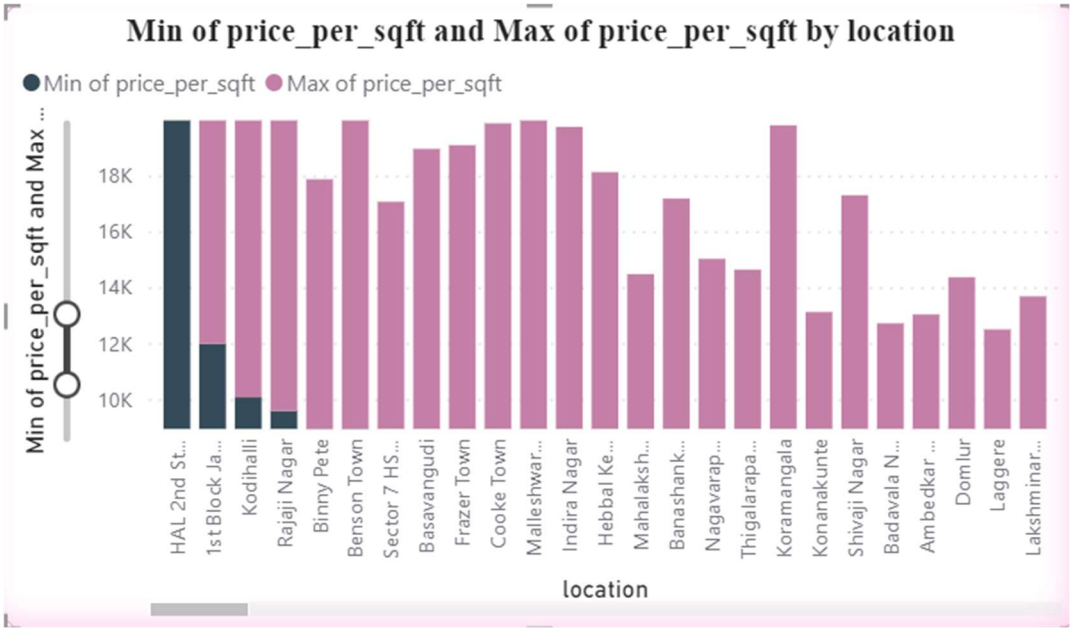

Fig 7. Overall bar graph

1. The graph represents the minimum and maximum price per square foot of properties in various locations.
2. The y-axis shows the price per square foot, ranging from 10K to 16K.
3. The x-axis lists different locations, including H.A.L 2nd St., Kothanur, Rainy Hospital, Sector 7 HSR Layout, Cooke Town, Mallasandra, Hebbal Kempapura, Indira Nagar, Horamavu Agara, Malleshwaram, Rajaji Nagar, Yelahanka New Town, Thanisandra Main Road, Thigalarapalya, Konanakunte Cross, Sarjapur Road Water Tank, Nagarbhavi Stage, Shivaji Nagar, Domlur, Ambalipura, and Lakshmipura.
4. Each location has two bars:
    - The darker bars represent the minimum price per square foot.
    - The lighter bars represent the maximum price per square foot.

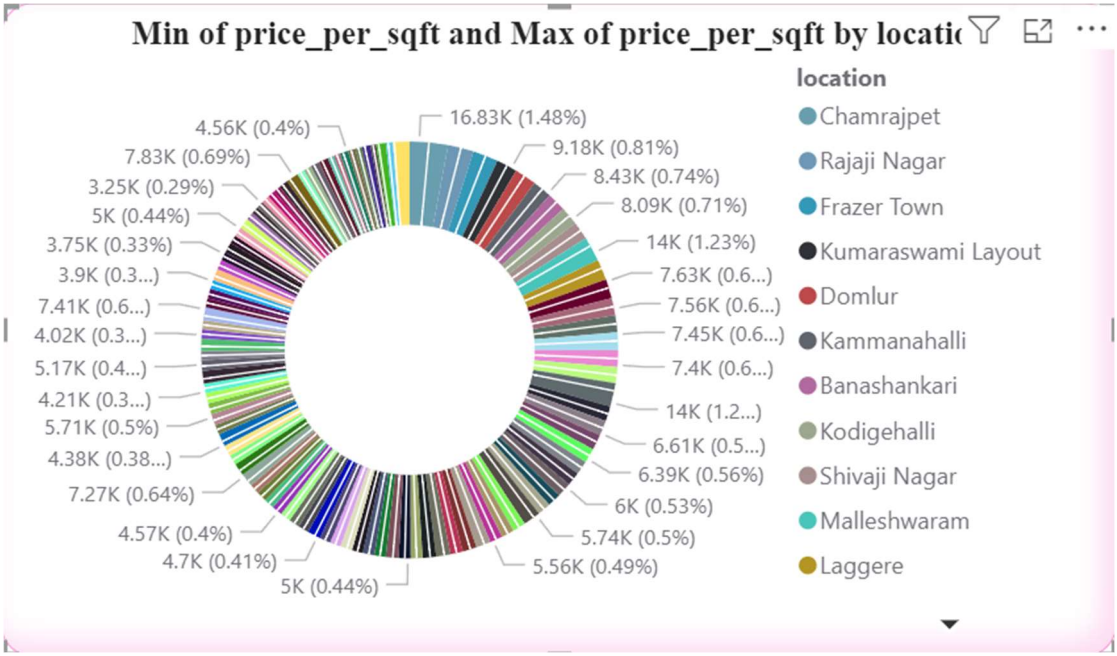

Fig 8. Overall Pie graph

Circular Bar Graph:
The image features a circular bar graph with multiple colorful segments.
Each segment represents data related to the “Minimum Price per Square Foot” and “Maximum Price per Square Foot” for different locations.
Color-Coded Segments:
The graph is divided into segments, each assigned a distinct color.
These colors correspond to specific locations, which are listed on the right side of the image.
Numerical Values:
Within each segment, there are numerical values indicating the minimum and maximum price per square foot for that location.
Percentages are also displayed next to these values.
Legend:
On the right side of the image, there is a legend that associates each color with a specific location.
Some of the mentioned locations include Chamarjpet, Rajaji Nagar, Frazer Town, Kammanahalli, Banashankari, Kodigehalli, Shivaji Nagar, Malleshwaram, and Laggere.
Background:
The background of the image is white, allowing the colorful graph to stand out prominently.

Interpretation and visualize the results

1. Comparative Analysis: Compare predicted prices with actual prices to assess the accuracy of the models. Visualize this comparison through line graphs or scatter plots, showing how closely the predicted prices align with the real values across different properties or time periods.

2. Trend Analysis: Use data mining techniques to identify trends in predicted prices over

time or across different segments of the real estate market. Visualize these trends through time series graphs or stacked bar charts to understand patterns and fluctuations in propertyprices.

3. Segmentation Analysis: Segment the dataset based on various factors such as property type, location, size, amenities, or market segment. Analyze predicted prices within each segment to identify differences and similarities. Visualize segment-wise price distributions using histograms or box plots to understand variations and outliers.

4. Geospatial Visualization: Utilize geographic information systems (GIS) or mapping tools to visualize predicted property prices on a map of the area. Color-coded maps can show spatial patterns and hotspots of high or low predicted prices, aiding in location-based decision-making and market analysis.

5. Correlation Analysis: Investigate correlations between predicted prices and other relevant variables such as demographic data, economic indicators, or neighborhood characteristics. Visualize these correlations through heatmaps or correlation matrices to identify factors thatstrongly influence property prices.

6. Forecasting: Use predictive modeling techniques to forecast future property prices based on historical data and current trends. Visualize these forecasts through line charts or trend lines, providing stakeholders with insights into potential future price movements and market dynamics.

7. Profitability Analysis: Calculate potential profitability metrics such as return on investment (ROI) or net present value (NPV) based on predicted prices and other financial considerations. Visualize these metrics through bar charts or scatter plots to assess the financial viability of real estate investments.

8. Scenario Analysis: Conduct scenario analysis by simulating different hypothetical scenarios or market conditions and evaluating their impact on predicted prices. Visualize the results of these scenarios through scenario comparison charts or sensitivity analysis graphs

9. Interactive Dashboards: Develop interactive dashboards using business intelligence tools such as Power BI or Tableau to present the results of price prediction analysis dynamically. These dashboards allow stakeholders to explore data, visualize insights, and drill down into specific areas of interest for informed decision-making.

10. Key Performance Indicators (KPIs): Define and track KPIs related to real estate performance, such as average selling price, price per square foot, or sales volume. Visualize these KPIs through dynamic dashboards or performance scorecards to monitor market trends and track progress towards business objectives.

c) Provide clearly the BI decision that is to be taken as a result of mining.

Certainly! Here's a clear Business Intelligence (BI) decision that could be made based on the results of mining real estate data:

Optimal Pricing Strategy Formulation:

Decision: Based on the insights obtained from mining real estate data, the business decides to adjust its pricing strategy to maximize profitability and competitiveness in the market.

Explanation:

1. Identifying Price Trends: Through data mining and analysis, the business discovers trends in property prices over time and across different segments of the market. This includes understanding which types of properties are experiencing price increases or decreases, and the factors driving these trends.
2. Segmentation Analysis: The business segments its properties based on factors such as location, size, amenities, and market demand. By analyzing predicted prices within each segment, the business gains insights into price sensitivity and market demand for differentproperty types.
3. Competitive Benchmarking: Using data mining techniques, the business compares its pricing strategies with those of competitors in the market. This includes analyzing competitor pricing trends, positioning, and market share to identify areas where the business can adjust its prices to gain a competitive advantage.

Demand Forecasting: Leveraging predictive modeling, the business forecasts future

demand for properties based on historical data and market trends. This helps in anticipating changes in market conditions and adjusting pricing strategies accordingly to meet demand and optimize revenue.

4. Profitability Analysis: The business conducts a profitability analysis to assess the impact of pricing changes on overall profitability. This involves calculating metrics such as return on investment (ROI), net present value (NPV), and profit margins to evaluate the financial implications of different pricing scenarios.

Decision Outcome:

Based on the insights obtained from data mining and analysis, the business decides toimplement the following changes to its pricing strategy:

1. Adjusting prices for properties in high-demand areas to capitalize on market trends andmaximize revenue.
2. Offering competitive pricing for properties in segments where competition is intense, aiming to attract more buyers and increase market share.
3. Implementing dynamic pricing strategies that take into account factors such as seasonality, market demand, and property features to optimize pricing decisions.
4. Monitoring pricing performance regularly and refining strategies based on ongoing analysis and market feedback to ensure alignment with business objectives and marketdynamics.

By making data-driven decisions informed by real estate data mining and analysis, the business can develop a more effective pricing strategy that enhances competitiveness, maximizes profitability, and drives long-term success in the market.

Conclusion :

In conclusion, leveraging data mining techniques in real estate provides invaluable insights that empower businesses to make informed decisions. By analyzing vast datasets encompassing property listings, historical sales data, demographic information, and market trends, businesses can extract actionable insights to optimize their operations and strategies.

Through data mining, businesses can:

1. Gain a comprehensive understanding of market dynamics, including price trends, demandpatterns, and competitive landscapes.
2. Identify factors driving property prices and market fluctuations, enabling the formulation ofeffective pricing strategies.
3. Segment the market based on various criteria such as location, property type, and amenities totailor offerings and target specific customer segments.
4. Forecast future market trends and demand, allowing proactive adjustments to business strategiesand operations.
5. Enhance decision-making processes by quantifying risks, evaluating opportunities, andoptimizing resource allocation.
6. Continuously monitor performance metrics and market conditions, enabling agile responses tochanges and opportunities.

References :

(PDF) Bangalore House Price Prediction (researchgate.net)

Predicting Bangalore House Prices: A Data Science Journey | by Muhammad Usman Soomro | Medium