Real Estate Project

Course-end Project 1

Description

- **A banking institution requires actionable insights into mortgage-backed securities, geographic business investment, and real estate analysis.**
- **The mortgage bank would like to identify potential monthly mortgage expenses for each region based on monthly family income and rental of the real estate.**
- **A statistical model needs to be created to predict the potential demand in dollars amount of loan for each of the region in the USA. Also, there is a need to create a dashboard which would refresh periodically post data retrieval from the agencies.**
- **The dashboard must demonstrate relationships and trends for the key metrics as follows: number of loans, average rental income, monthly mortgage and owner's cost, family income vs mortgage cost comparison across different regions. The metrics described here do not limit the dashboard to these few.**

Project Task: Week 1 Summary:

During Week 1 of the project, the focus was on data import and preparation, as well as exploratory data analysis (EDA). Here's a summary of the tasks and observations made:

Data Import and Preparation:

Imported the data and identified the primary key.
Evaluated the fill rate of variables and planned for missing value treatment. Chose specific treatments for each variable based on the nature of the data and the impact of missing values on the analysis.

Exploratory Data Analysis (EDA):

Conducted debt analysis by exploring the top 2,500 locations with the highest percentage of households having a second mortgage and ownership above 10 percent. Visualized the results using a geo-map and set an upper limit of 50 percent for the second mortgage households.

Calculated bad debt using the provided equation and created pie charts to display overall debt and bad debt.

Created box and whisker plots to analyze the distribution of variables such as second mortgage, home equity, good debt, and bad debt across different cities.

Collated income distribution data for family income, household income, and remaining income and visualized the findings.

Derived new fields for population density and median age, using appropriate variables and weighted averages. Visualized the results.

Created population bins with appropriate class intervals and analyzed the married, separated, and divorced population within each bin.

Examined rent as a percentage of income at an overall level and for different states, providing observations on the findings.

Conducted correlation analysis for relevant variables and created a heatmap to visualize the correlations.

Project Task: Week 2 Summary:

During Week 2, the focus shifted to data pre-processing, factor analysis, data modeling, and data reporting. Here's a summary of the tasks and objectives:

Data Pre-processing:

Applied factor analysis to identify latent variables in the dataset and explore linear relationships among the measured variables.

Identified and plotted the loadings for the latent variables, including high school graduation rates, median population age, second mortgage statistics, percent own, and bad debt expense.

Data Modeling:

Built a linear regression model to predict the total monthly expenditure for home mortgages loans, specifically focusing on the mean monthly mortgage and owner costs.

Ran the model at both the national and state levels, excluding loans with NaN values for the predicted variable.

Considered variables with significant impact, avoided multi-collinearity, and aimed for an R-squared of 60 percent or higher.

Data Reporting:

Created a Tableau dashboard to present the analysis and insights to stakeholders.

Included appropriate chart types and metrics to address specific business needs, such as a box plot for average rent by type of place, pie charts for debt analysis, a geo-map for locations with high second mortgage percentages, a correlation matrix heatmap, and a pie chart for population distribution across different types of places.

Overall, this project involved comprehensive data analysis, modelling, and reporting to gain insights into the real estate market. It encompassed tasks ranging from data preparation and EDA to factor analysis, regression modelling, and visual reporting using Tableau. The project aimed to support decision-making processes for stakeholders in the real estate industry and provided valuable insights into various market dynamics.

Conclusion:

The project's modelling and Tableau dashboard reporting provided valuable insights into the real estate market, enabling stakeholders to identify profitable investment opportunities, understand rental market dynamics, develop targeted marketing strategies, and make informed policy and urban development decisions.