Project Title: Housing Affordability Project Report

1. Introduction

- Background and Motivation: The aim of this project is to analyze the factors influencing housing affordability and develop a predictive model to estimate housing costs based on various variables.

- Project Objective: To provide insights into the key predictors of housing affordability and propose strategies for improving affordability.

2. Data Collection and Preprocessing

- Dataset Description: The dataset used for this project is the National Survey of Affordable Housing (NSAH) dataset, containing information on housing-related variables such as income, housing costs, bedrooms, age, and utility costs.

- Data Cleaning and Preprocessing Steps: The dataset was cleaned by handling missing values, removing outliers, and ensuring data consistency.

3. Exploratory Data Analysis

- Dataset Overview: Descriptive statistics and visualizations were used to understand the characteristics and distributions of housing-related variables.

- Visualizations and Statistical Summaries: Plots and summary statistics were generated to identify patterns, trends, and correlations among the variables.

4. Feature Selection and Model Building

- Selected Features for Housing Affordability Prediction: Based on correlation analysis and feature importance scores, the most relevant variables for predicting housing affordability were identified, including income, bedrooms, age, utility costs, and housing costs.

- Multilinear Regression Algorithm: A multilinear regression model was chosen for its interpretability and ability to capture relationships between multiple predictors and the target variable.

- Model Training and Evaluation: The model was trained using historical data from the NSAH dataset, and its performance was evaluated using metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared (R2).

5. Model Interpretation and Findings

- Coefficient Interpretation: The coefficients of the selected features were analyzed to determine their impact on housing affordability. The household size, number of bedrooms, income, and utility costs were found to have significant influences on housing costs.

- Key Predictors and their Impact: The number of bedrooms and household income were identified as strong predictors of housing affordability. As the number of bedrooms increases, housing costs tend to rise, while higher income levels are associated with lower housing costs.

- Insights and Findings: The model suggests that the size of the household and the availability of affordable housing options play crucial roles in determining housing affordability.

6. Model Assumptions and Validation

We had difficulties in this phase, we need more advice from experts.

7. Results and Visualization

- Presentation of Visualizations: Various visualizations, including scatter plots, histograms, and residual plots, were created to present the results and insights derived from the model.

- Interpretation of Visualizations: The visualizations illustrated the relationships between the selected features and housing affordability, highlighting the key factors that impact housing costs.

8.

Conclusion and Recommendations

- Summary of the Project and Findings: This project aimed to analyze housing affordability and develop a predictive model. The findings highlighted the importance of household size, income, and number of bedrooms in determining housing costs.

- Key Insights into Housing Affordability: The analysis emphasized the need for policies and interventions to increase affordable housing options, support income growth, and address the specific needs of households with different sizes and income levels.

- Recommendations for Improving Affordability: Based on the findings, recommendations include implementing affordable housing initiatives, providing financial assistance programs, and promoting economic opportunities to enhance housing affordability.

- Limitations and Future Research: It is important to acknowledge that the model's accuracy may be affected by data limitations and other external factors. Further research could focus on incorporating additional variables or exploring advanced predictive modeling techniques.

References:

Dataset: <https://www.huduser.gov/portal/datasets/hads/hads.html>

Github: <https://github.com/bdamags/Project>