homeLLC Report

Ahmad Bader

February 2024

Introduction

This report aims to analyze the impact of various features on house prices in the United States. The data used for this analysis includes a comprehensive DataFrame containing information on key economic and demographic factors.

Data Overview

The dataset consists of various features, including Urban Population, Consumer Price Index (CPI), Houses Supplied, Median Income, Interest Rates, Working Population, Subsidies (billion $), CSUSHPISA, Unemployment Rate, Construction Materials, and GDP per Capita. The target/proxy variable is CSUSHPISA, represented by the 'target' column.

Statistical Analysis

To gain insights into the dataset, descriptive statistics and correlation coefficients were calculated. The correlation heatmap revealed potential relationships between features. Additionally, Variance Inflation Factor (VIF) values were computed to identify potential multicollinearity issues. Notable features with high VIF values included Urban Population, Median Income, Working Population, Subsidies, CSUSHPISA, Unemployment Rate, Construction Materials, and GDP per Capita.

Model Training

Linear Regression model was employed to analyze the impact of these features on house prices. Cross-validation was performed to assess model performance, and Root Mean Squared Error (RMSE) was used as an evaluation metric. The model indicated all features significantly affecting house prices.

Prominent Features

Recursive Feature Elimination (RFE) was utilized to identify the most prominent features influencing house prices. Features such as Urban Population, Median Income, Working Population, Subsidies, CSUSHPISA, and GDP per Capita consistently appeared influential in the model.

Conclusion

In conclusion, the analysis suggests that all factors such as Urban Population, Median Income, Working Population, Subsidies, CSUSHPISA, and GDP per Capita play crucial roles in impacting house prices in the US. Further research and detailed modeling could provide deeper insights into the complex relationships among these features.