

Cost of Living

Ankit Mistry, Saish Malluri, Shivam Mistry

Objective

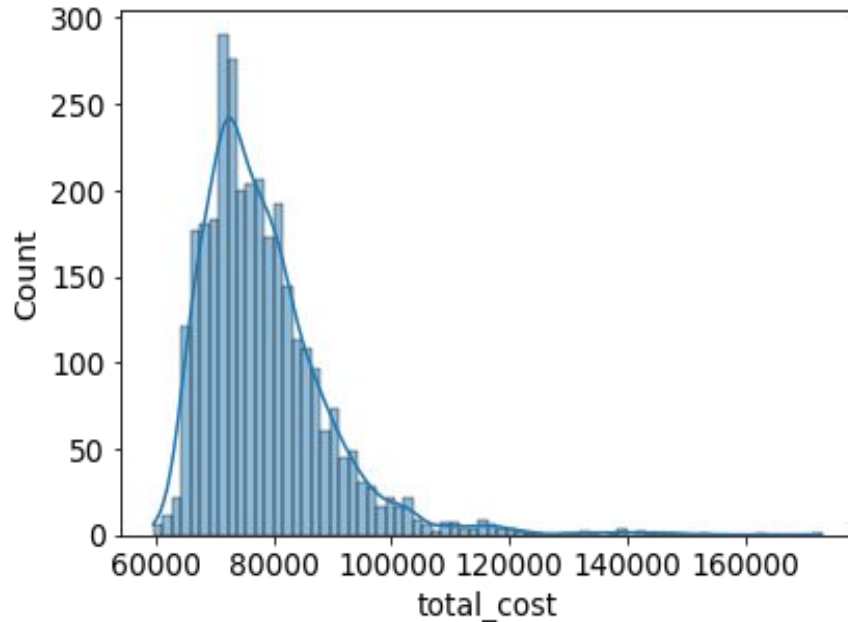
- Goal is to examine the distribution of living expenses across various factors using a dataset from Kaggle.
- Our objective was to understand the connections between these factors and the cost of living, using Python and well-known data analysis libraries for visualization.
- Research Questions:
 - How geographical areas affects various features and Cost of Living?
 - How family size impacts various features and Cost of Living?
 - Which counties are most/least affordable to live in the U.S. and Tennessee?

Data

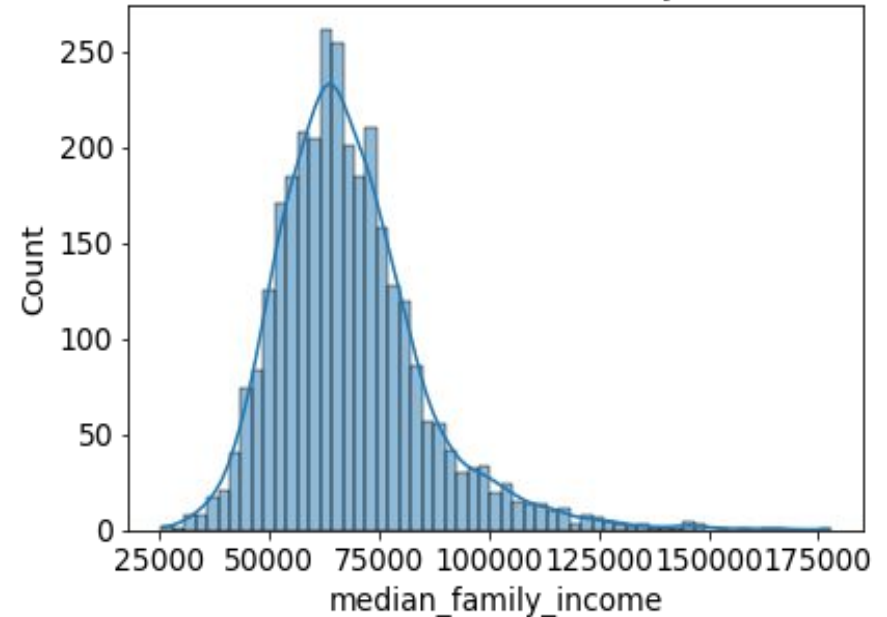
- The dataset is composed of these features:
 - State, County, Metropolitan area
 - Family Size
 - Food Cost
 - Healthcare Cost
 - Transportation Cost
 - Childcare Cost
 - Housing Cost
 - Other Necessities Cost
 - Taxes
 - Median Family Income
 - Total Cost
- The dataset incorporated 31430 entries of community-specific estimates for ten family types, including one or two adults with zero to four children, in all 3143 counties and metro areas across the United States.

Exploratory Data Analysis

Distribution of Total Cost

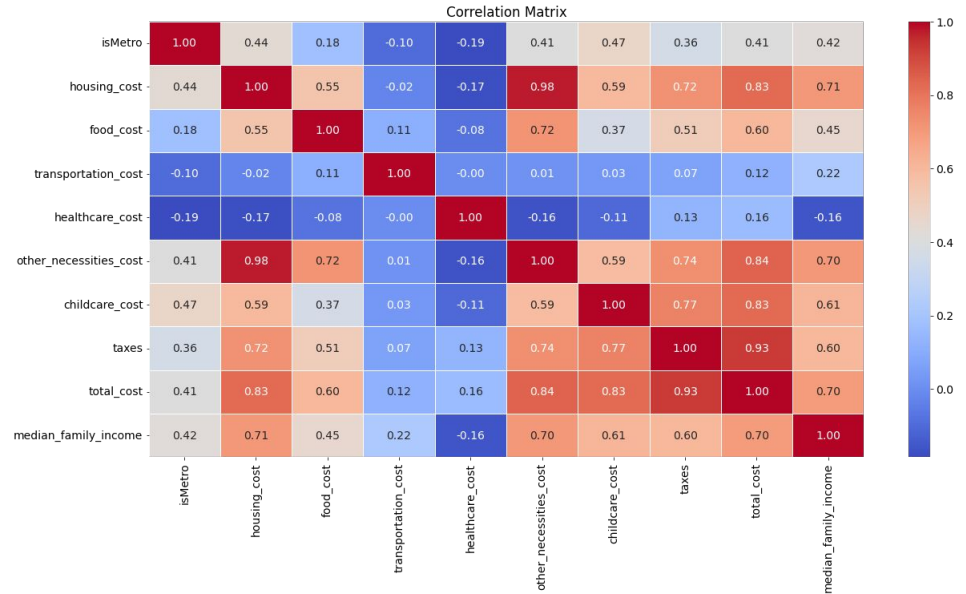


Distribution of Median Family Income



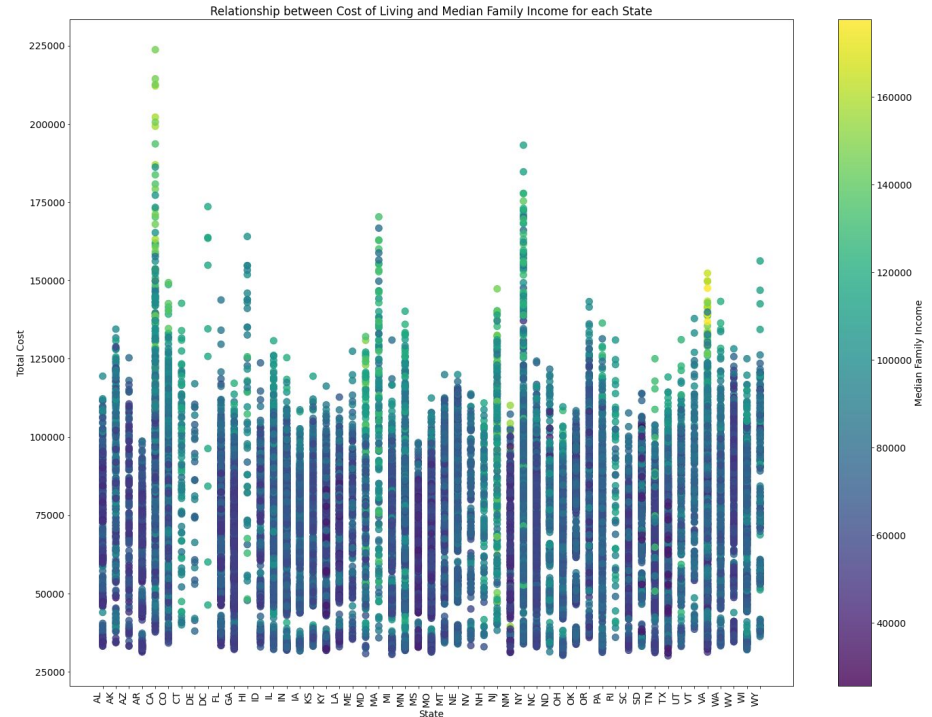
Exploratory Data Analysis

- Correlation matrix is showing how the total cost (cost of living) is correlated with all the other factors and is providing a view of how strongly pairs of variables are related.
- There is a positive correlation between housing cost + taxes and the total cost, which might suggest that areas with higher housing and places with a higher tax burden tend to have a higher cost of living overall.
- There isn't a strong correlation between transportation cost + healthcare cost and the total cost, which might indicate that higher transportation and healthcare costs aren't really associated when determining the cost of living.

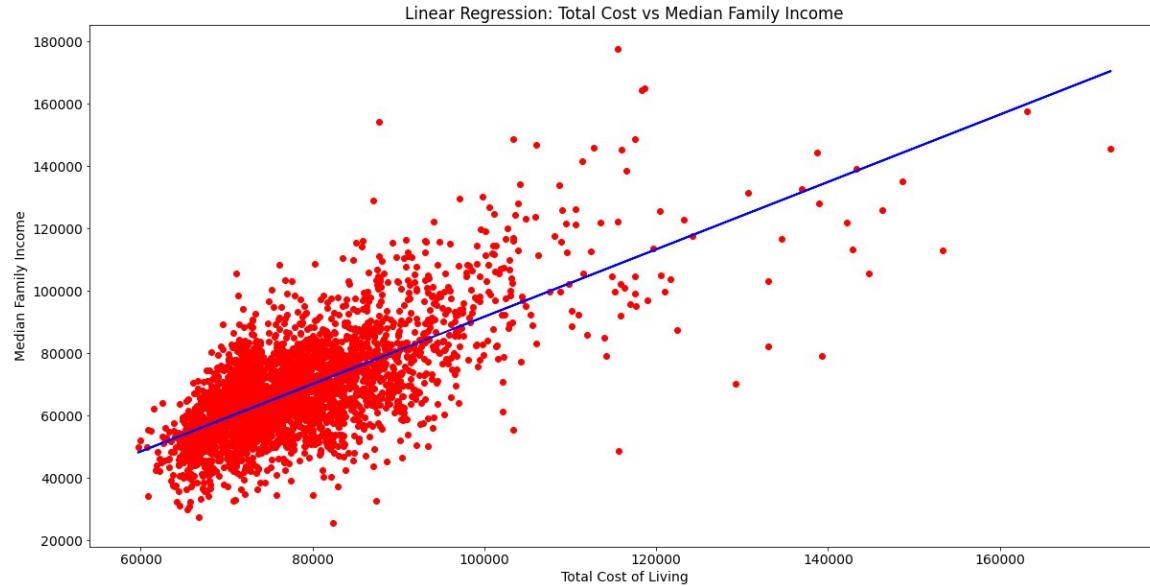


Income vs. Cost of Living

- This plot shows each county's individual family incomes along with their respective cost of living.
- California, Massachusetts, New York, and Virginia have high cost of living.
- Arkansas, Mississippi, Indiana have low cost of living.

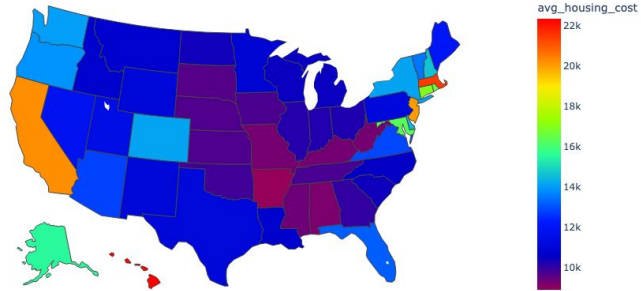


Income vs. Cost of Living: Regression Model

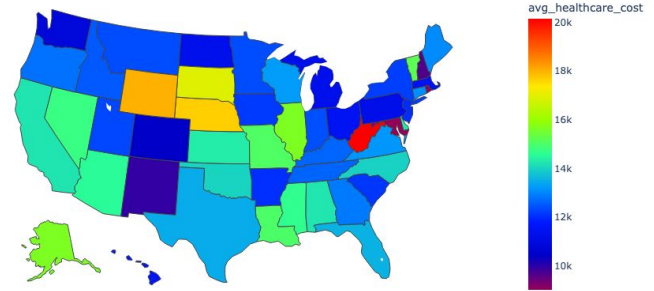


Feature Visualization on Map

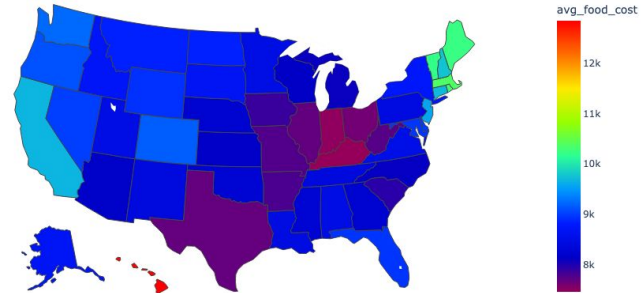
avg_housing_cost



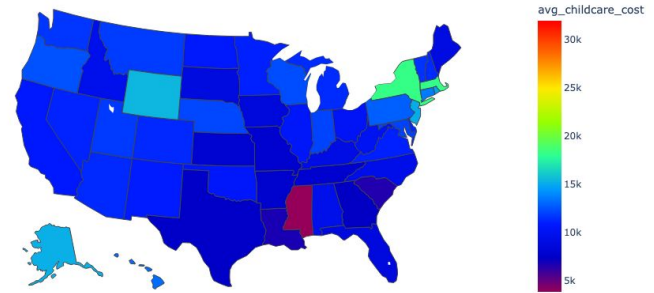
avg_healthcare_cost



avg_food_cost

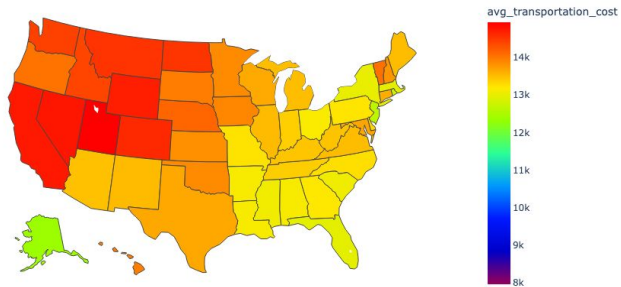


avg_childcare_cost

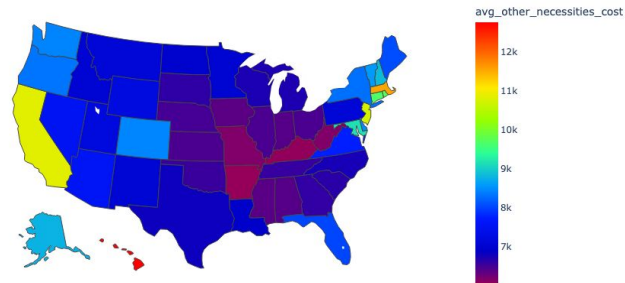


Feature Visualization on Map

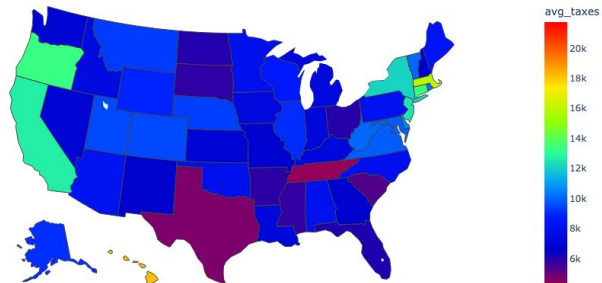
avg_transportation_cost



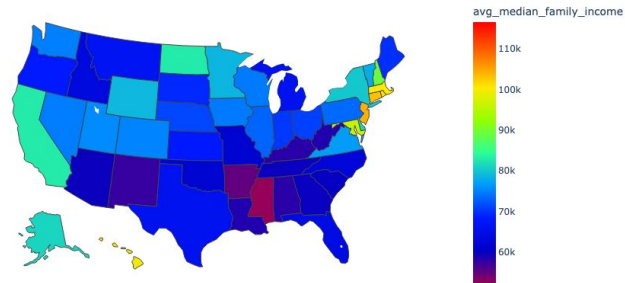
avg_other_necessities_cost



avg_taxes

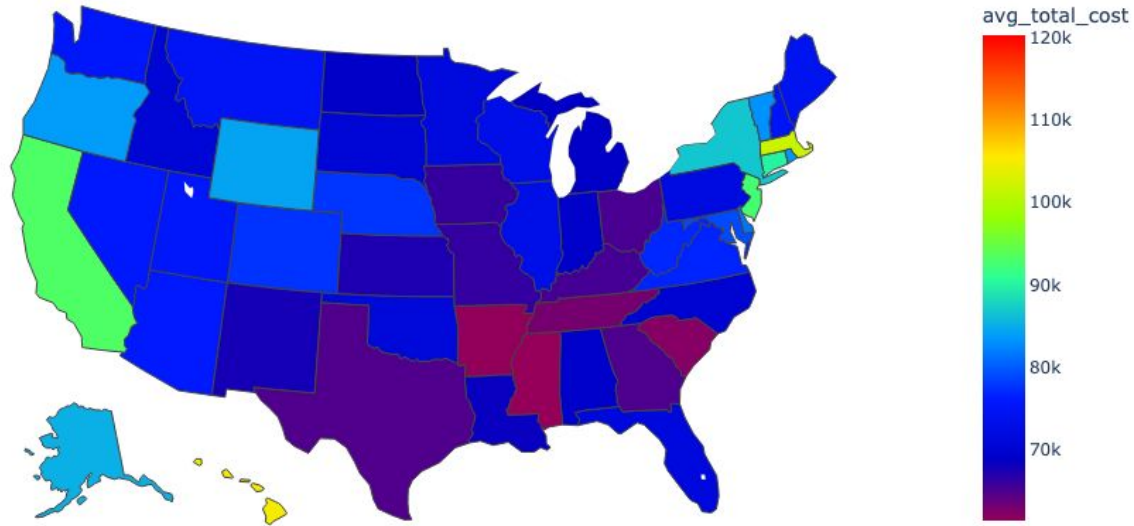


avg_median_family_income



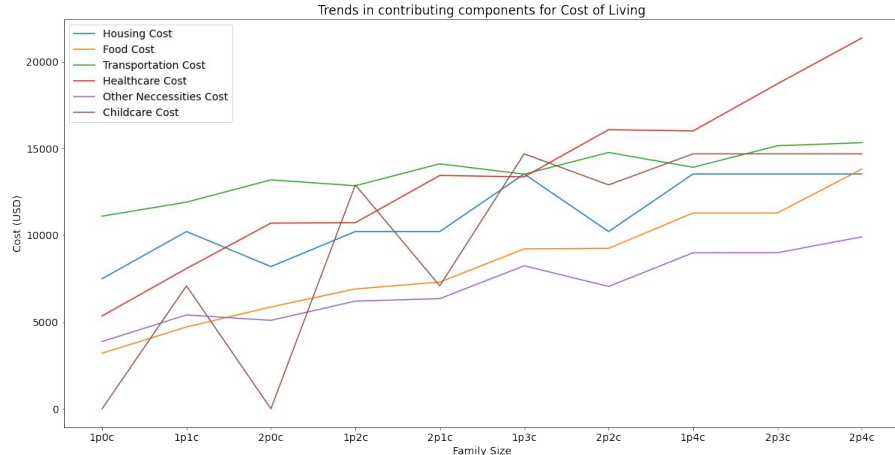
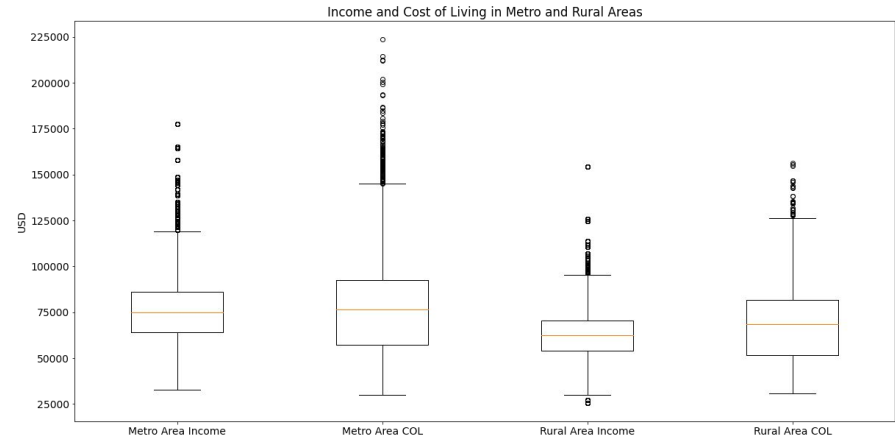
Total Cost of Living in the U.S.

avg_total_cost



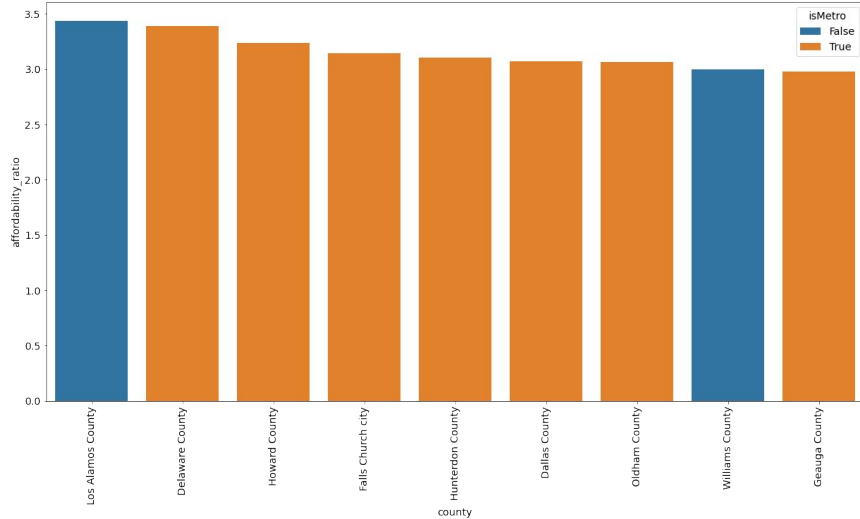
Impact of Family Size

- The box and whisker plots on the right are illustrating the distribution of income and cost of living in both rural and metro areas
- In general, the median income and cost of living in metro areas is typically higher, and are about the same.
- However, in rural areas, the cost of living is higher than income.
- All the costs tend to increase with increase in family size
- Housing cost and other necessities cost tends to decrease when number of children decreases

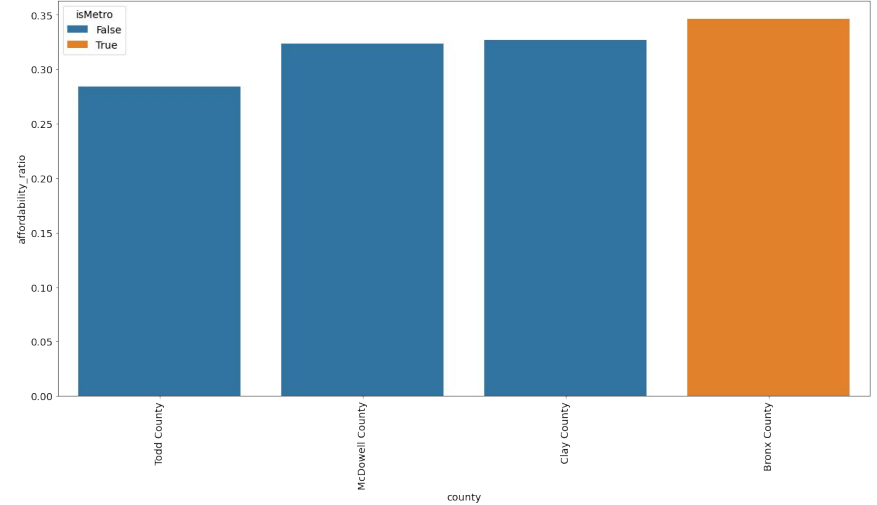


Affordability Analysis (U.S.A.)

Most Affordable Counties in US

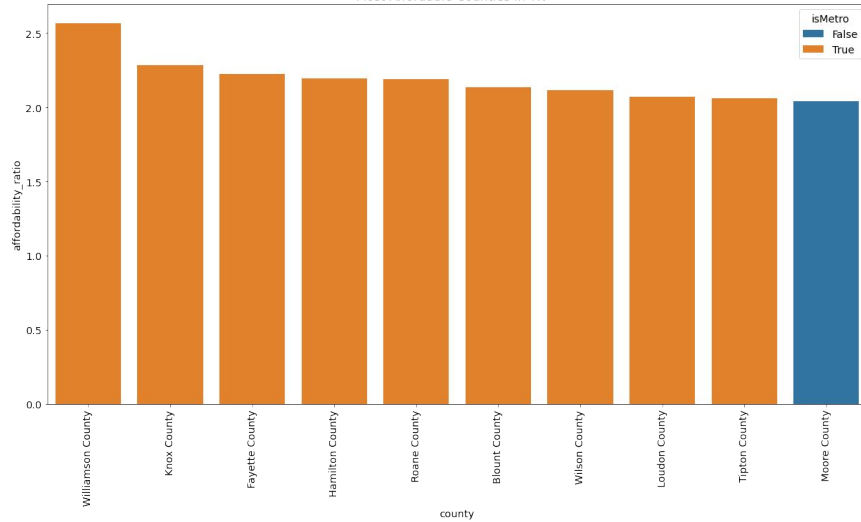


Least Affordable Counties in US

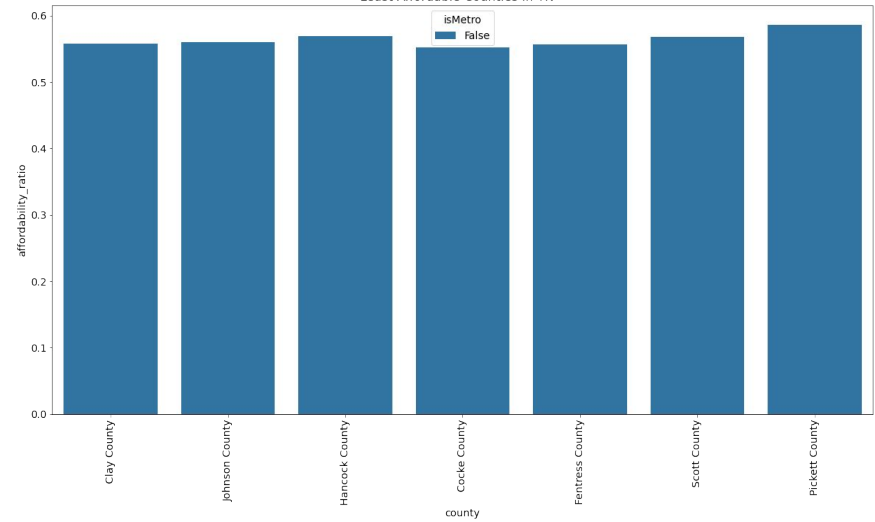


Affordability Analysis (TN)

Most Affordable Counties in TN



Least Affordable Counties in TN



Challenges and Future Work

- **Challenges**

- The dataset didn't have a Cost of Living Index, which is a comprehensive representation of various features in Cost of Living. We had to work with raw costs in USD, which can be hard to compare and visualize.
- Installing plotly and having the map visualizations come up

- **Future Work**

- Try to calculate Cost of Living Index, which can be better utilized for comparison purposes.
- Implement machine learning models to predict cost of living based on various factors.
- Extend the analysis to different cultural contexts to understand how cost of living factors may vary globally