## **"Dollars and Sense: Exploring the Cost of Living Across the United States"**

**Introduction:**

The primary purpose of the analysis project on the cost of living in the US utilizing the US Family Budget Dataset from the Economic Policy Institute (EPI) is to comprehensively assess and understand the nuanced variations in living expenses across different US counties and metropolitan areas, by leveraging the insights provided by the dataset. Overall, the project is a valuable resource for understanding and navigating the varying cost-of-living landscapes in the United States.

The dataset focuses on revealing the cost of living across various regions within the United States. It derives insights from the Family Budget Calculator the Economic Policy Institute (EPI) created. This dataset provides detailed estimates specific to communities, encompassing ten distinct family types. These families range from one or two adults with varying numbers of children, from none to up to four children. The dataset meticulously covers all 1877 counties and metropolitan areas throughout the United States, offering a comprehensive understanding of the expenses associated with different family compositions in diverse regions of the country.

**NOTE**: Some data need to be included in the dataset, like we do not have cities like New York, Los Angeles, etc.

Link of the data Set - [Cost of Living US](https://www.kaggle.com/datasets/asaniczka/us-cost-of-living-dataset-3171-counties)

The data set has the following column name and description, and there are **31k** rows of data.

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| case\_id | Identification number for the case |
| state | State within the United States |
| metro | Denotes if the area is a metropolitan |
| area name | Name of the area |
| county | Specific county within the area |
| family\_member\_count | Number of family members |
| housing\_cost | The cost associated with housing |
| food cost | Cost related to food expenses |
| transportation\_cost | Expenses linked to transportation |
| healthcare\_cost | The cost associated with healthcare |
| other\_necessities\_cost | Additional necessary expenses |
| childcare\_cost | Cost related to childcare |
| taxes | Tax expenses |
| total\_cost | Total cost of living for the family |
| median\_family\_income | Median income for the family |
|  |  |
|  |  |
|  |  |

**Analysis Methodology Followed**

I used Tableau to perform my analysis for the following reason

1. User-Friendly Interface: It is easy for beginners and experts to create visualizations without complex coding.

2. Interactive Visualizations: Enables dynamic and engaging data representation for exploration.

3. Data Connectivity: Supports various data sources for comprehensive analysis.

4. Powerful Analytics: Provides in-depth data examination and correlation.

5. Scalability: Handles both small and large datasets effectively.

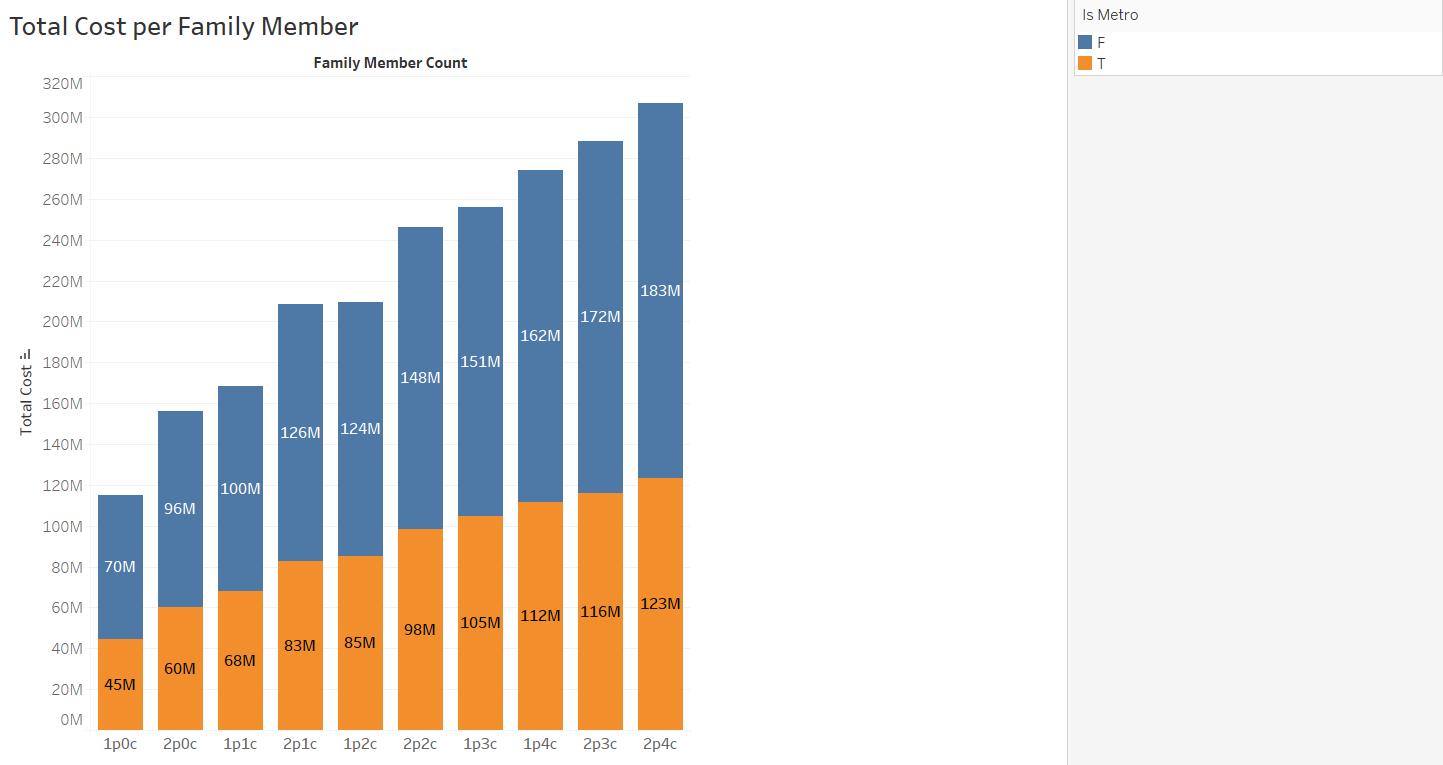
6. Diverse Visualizations: A wide range of visualization options for versatile data representation exists.

7. Real-Time Analysis: Capable of analyzing live data for immediate decision-making.

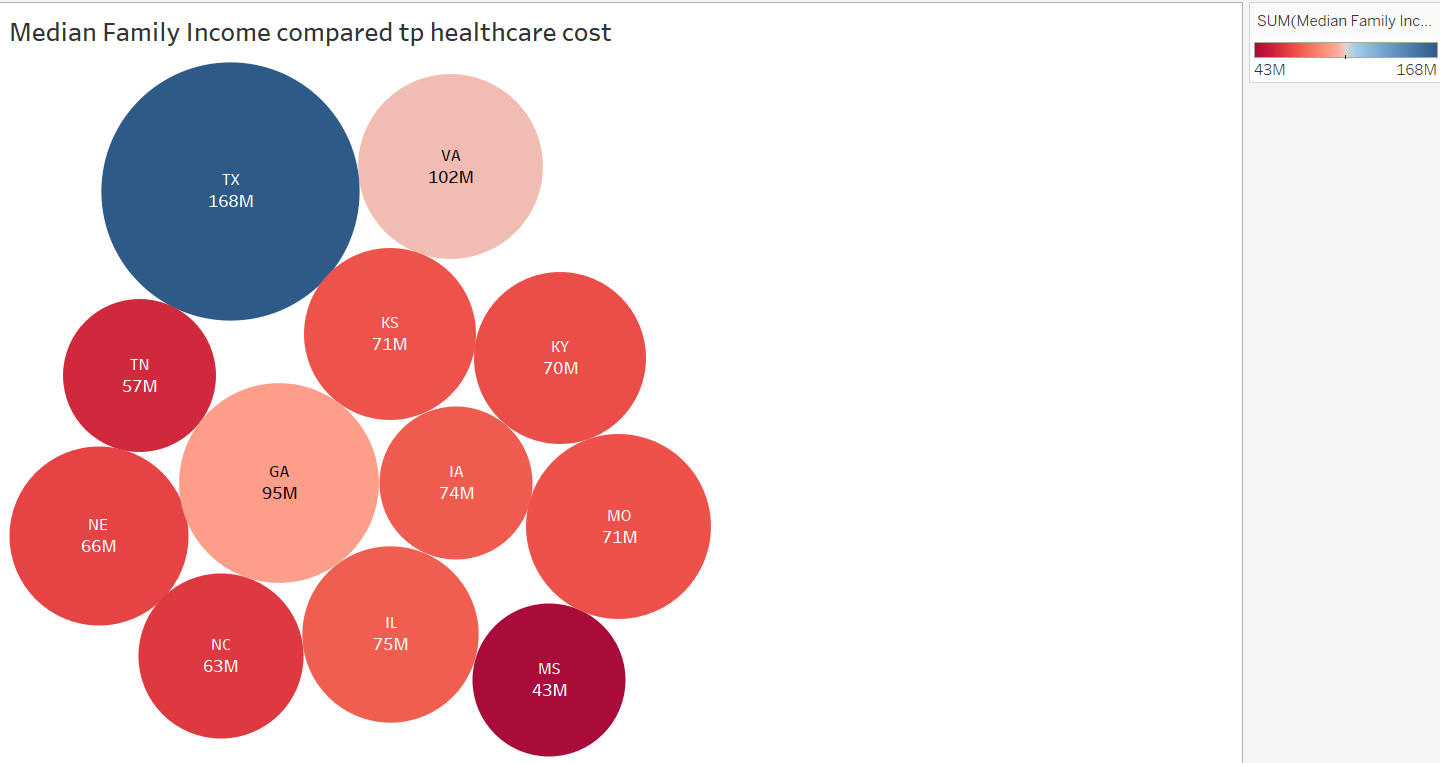
8. Community Support: Benefits from a strong user community and comprehensive support resources.

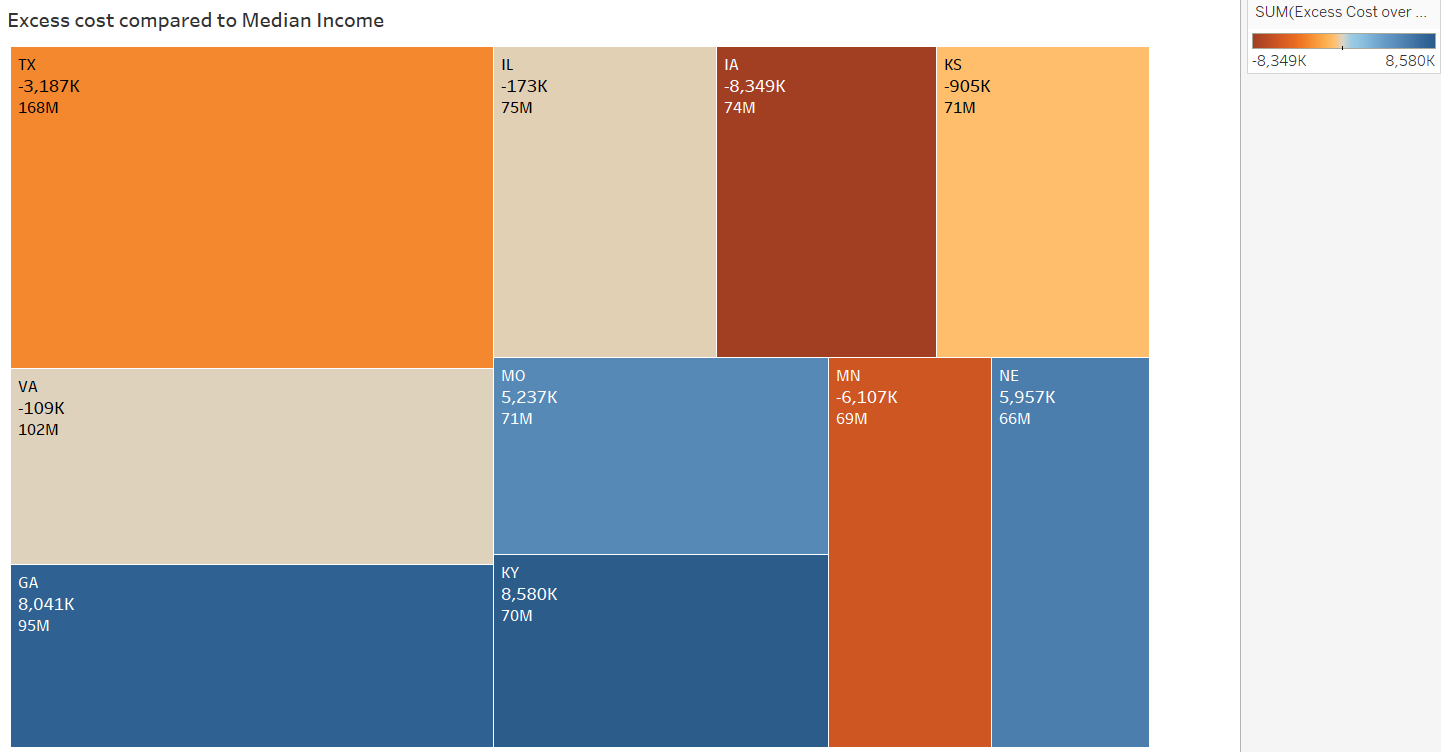
**Visualization creation process**

For the visualization, I have created three dashboards based on Total Cost, Median Family Income, and Impact of Metro. I have used graphs like Bar Graphs, Geographical representations, tree maps, bubble charts, and scatter plots. Further, I have created some calculated parameters and used them in the graphs. Added filter wherever required to get the top 5 states regarding different costs and then added filter in the category.

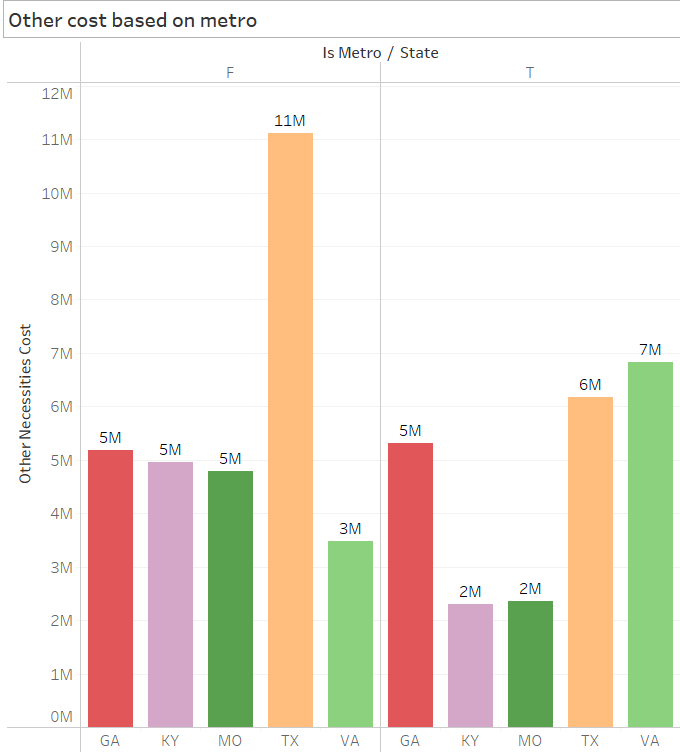


For this visualization, I have created a bar graph in terms of Total cost per family member and included how much the price is based on whether there is metro facility.

This graph represents the median family income and how much they spend on health, which is indicated by the bubble size; the more significant the size, the more prominent the families spend on their health.

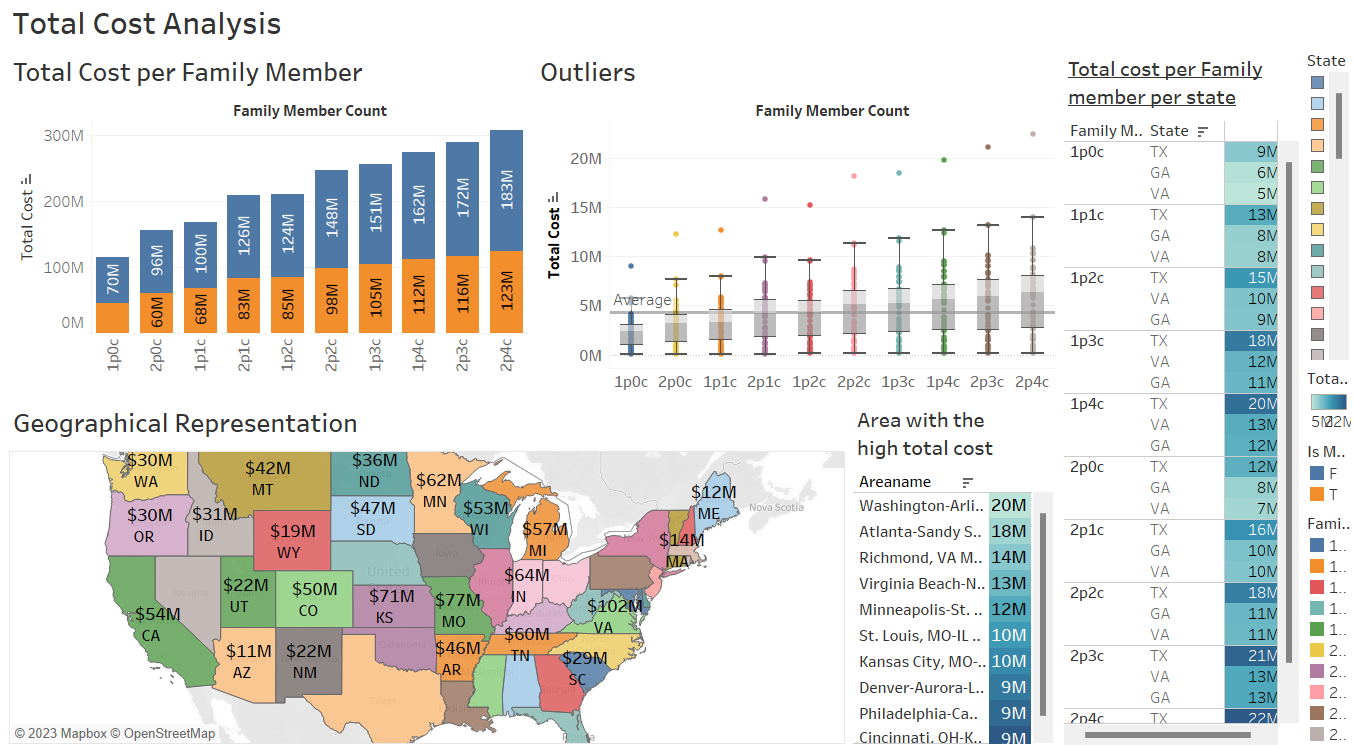


The above graph is a Tree map, which represents the excess cost of the median family income.

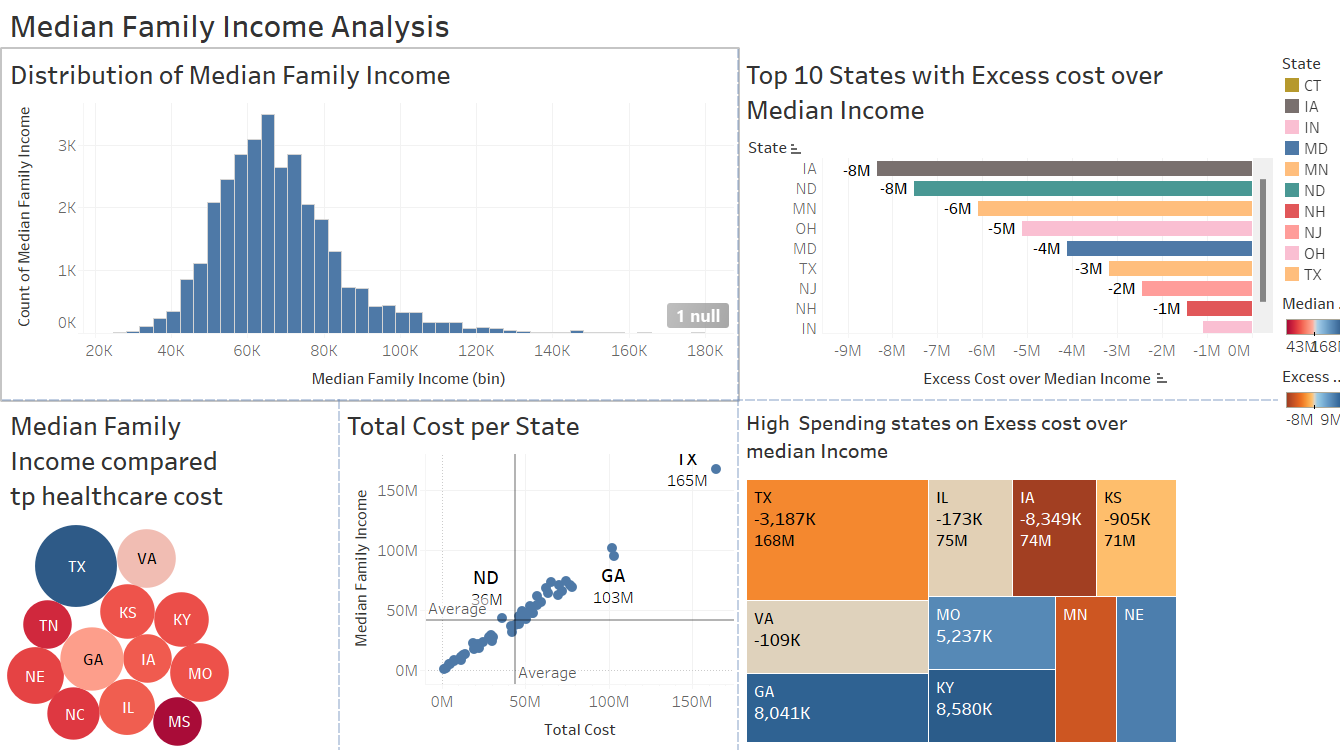


For this graph, I have used the bar chart to represent the top 5 states in terms of other costs and how they are affected by the presence of Metro.

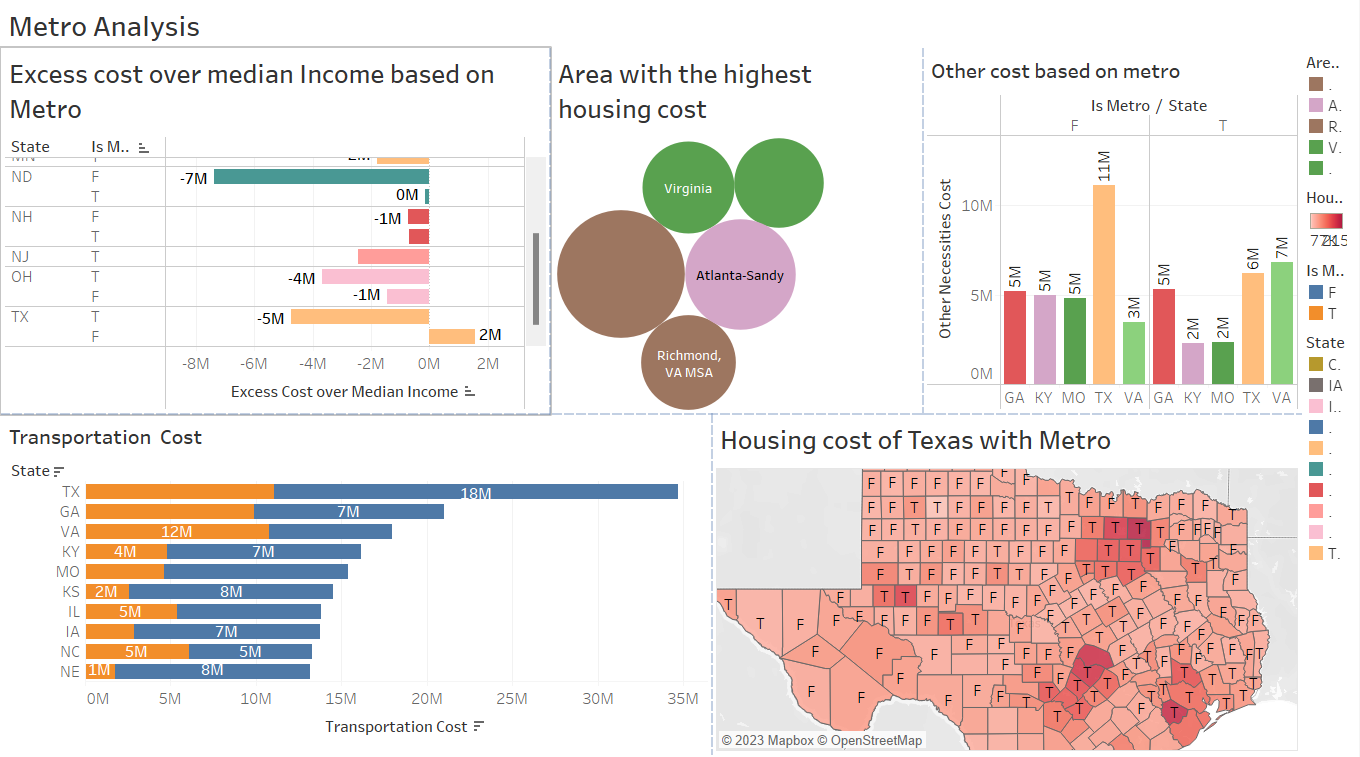
**Dashboard 1- Total cost analysis**



**Dashboard 2 – Median Family Income Analysis**



**Dashboard -3 Impact of Metro Analysis**



**Analysis Results**

|  |  |  |
| --- | --- | --- |
| **Total Cost Analysis** | | |
| Positive Insights | Negative Insights | Other Comments |
| 1. The breakdown of total costs from the bar graph into categories reveals that housing, healthcare, and transportation are the primary contributors to most regions' overall cost of living. This insight is crucial for budgeting and financial planning.  2. Washington-Arlington, Atlanta- Sandy Springs, Virginia -Richmond are the wealthiest areas to live as per the total cost spent, and two places from Virginia suggest their facilities are excellent so that people are willing to pay that much from the table area with the highest total cost.  3. The geographical representation highlights significant regional disparities in the cost of living across the United States. Users can quickly identify areas with higher or lower living costs, providing valuable insights for individuals or businesses considering relocation.  4. From the bar graph of total cost per family member, I found that the families spend an average of 4.3 million dollars on total costs towards their families, including every facility. People with fewer family member has the advantage of reducing their total cost. | There are outliers in each family member group in terms of Total cost, and that difference is very high compared to median income. Those outliers may be the result of specific individuals who have high incomes. |  |

|  |  |  |
| --- | --- | --- |
| **Median Family Income Analysis** | | |
| Positive Insights | Negative Insights | Other Comments |
| 1. From the distribution graph, I find that the Average median family income is 63K dollars from the distribution graph. Very few people are above 100k dollars, and there is a positive correlation between the total cost and median family income from the tree map.  2. From the bubble chart, I find that Texas spends more on health care than any other state, which may be why it has high excess costs over its median income.  3. From the tree map, I find that few states have favorable excess costs, which means they are saving all that money, and all these states have low healthcare costs, which results in reasonable extra prices over their median family income.  4. From the scatter chart, I find many states above the average median family per total cost line, which is a good sign that the conditions are improving, which will, in turn, help the country. | States with the highest median income have a negative excess cost, which means that the place with a high total price spends way more than their average income. |  |

|  |  |  |
| --- | --- | --- |
| **Impact of Metro** | | |
| Positive Insights | Negative Insights | Other Comments |
| 1. From the bar vertical bar, I find that the metro impacts the median family income and helps to reduce the excess cost.  2. From the bubble chart, I find that all the area with the highest housing cost has metro facilities, which increases the land value but decreases the transportation cost.  3. From the horizontal bar chart, I find that for Virginia state, the transportation cost is less even if they don’t have metro facilities, which indicates that most people live in an area where all the basic needs are close by.  4. From the geographic chart of Texas, I find that areas where there is a Metro, the housing increase, and it directly affect the surrounding areas | Many states, except for Virginia, have high transportation costs; even if they have a metro facility deeper analysis is required on what leads to that. |  |

**Conclusion:**

The analysis from Tableau's Total Cost, Median Family Income, and Metro Analysis dashboards reveals significant disparities between the cost of living and median family incomes across various regions. High-cost metros are evident, indicating potential financial strain for residents due to income levels needing to keep pace with living expenses. This disparity underscores challenges in maintaining a comfortable standard of living in these areas. Such findings might lead to migration trends away from high-cost regions, impacting local economies and demographics. Policymakers, businesses, and individuals can use this data to address affordability issues, highlighting the urgency for strategies to bridge the gap between the cost of living and income levels to ensure better economic stability and quality of life for residents.