Housing Cost Burden

Dataset Link: Housing Cost Burden. | Video | Github

Description: This dataset shows the percentage of households in California—across different regions, counties, cities, and neighborhoods—that spend more than 30% or 50% of their income on housing. It highlights the extent of housing cost burdens using data from HUD and the U.S. Census. This information is important to our overall project because it helps us understand how housing affordability affects different communities, especially low-income and minority households. It supports our goal of examining the links between housing, food affordability, unemployment, birth counts, gini coefficient and how its health outcomes.

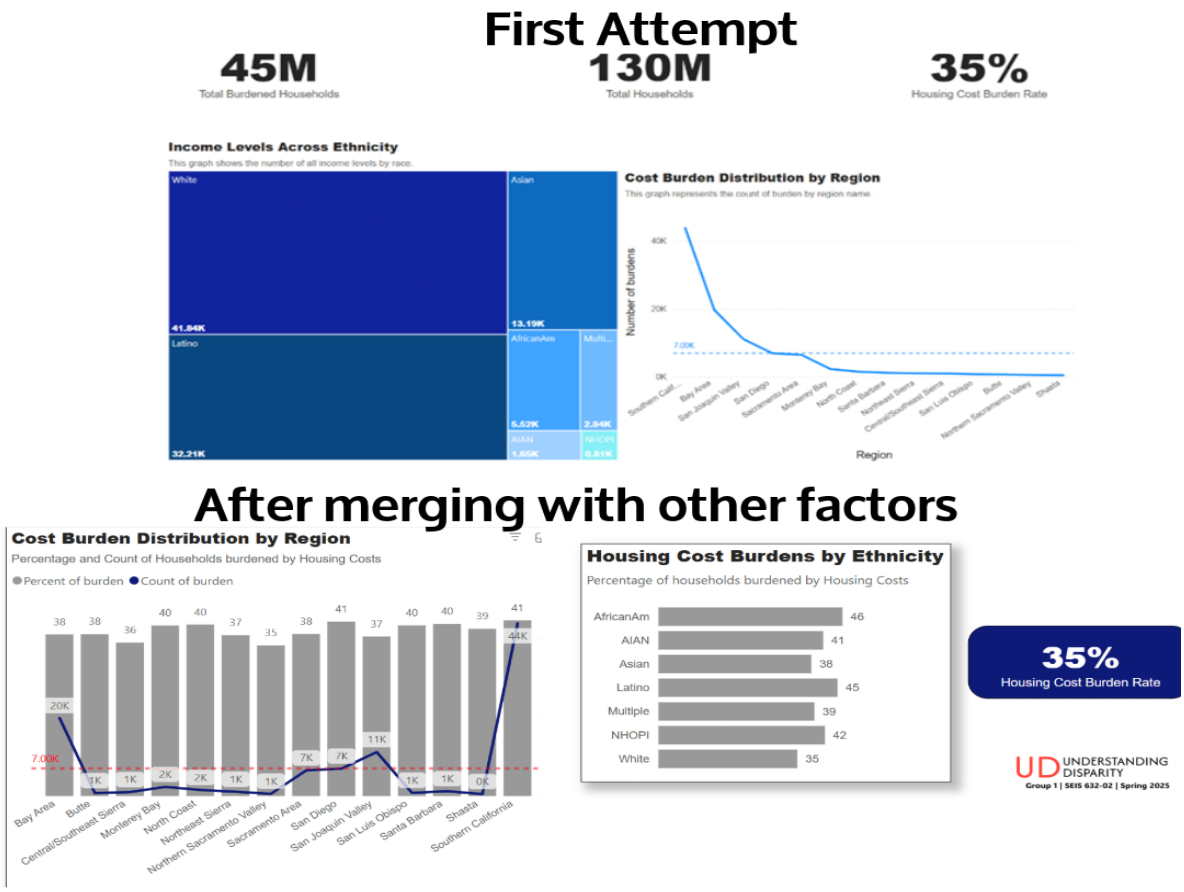
Dataset cleaning:

- Renamed the data table from APPEND_TABLE to Housing_Cost_Burden.
- Excluded total from race_eth_name (counts only races).
- Converted county_fips datatype to integer for consistency with other datasets.
- Created a new column county_region_name by concatenating county_name and region_name using DAX formula:
    - county_region_name = [county_name] & ", " & [region_name].
- Converted county_region_name datatype to string.
- Removed the version column (not needed) and the CA_decile column (all values were null).
- Filtered out blank rows in total_households, county_fips, rse, CA_RR, and se columns.
- Removed empty values from burdened_households, percent, and LL95CI.

Key Takeaways:

- Significant Housing Cost Burden:
    - As shown in the cards up there that a high portion of households, specifically 35%, are experiencing a housing cost burden. This represents 45 million out of 130 million total households.
- Income Levels by Ethnicity:
    - As shown in the treemap graph up there, it shows the number of people from different ethnic groups according to their income data. This shows that there are a lot of groups suffering from low income and high housing cost burdens.
- Cost Burden Hotspots:
    - As shown in the line graph up there Southern California has the highest number of burdened households by a large margin (over 40K) which means that the housing cost burden is not evenly distributed geographically.

Visuals:



For the Housing Cost Burden Rate card, I used the following DAX formula:

BurdenedToNormalPercentage = DIVIDE(SUM([burdened_households]), SUM([total_households]))

Reflection: I liked that our group communicated very well and worked hard to choose a topic that engaged everyone. In our first meeting, we set up a shared calendar with milestones, created a MS Teams channel for communication, and organized a OneDrive folder to keep all our materials in one place. We also divided the project roles effectively. The main reason everything went smoothly was the strong communication within our group. The main challenge we faced was how to synchronize all our work together on Power BI. However, we managed to solve this problem by making each person make their changes one at a time.