Living Comfy

Elena Ingraham
Cooper Timmerman

Description

We are going to analyze income tax paid per zip code and housing price across the United States to try and evaluate how comfortable the residents in every area are living. Then we will compare those results with several other datasets including election data, health data, and others to search for unseen patterns.

Prior Work

There has been a lot of work, a lot of which is done by government agencies, to evaluate the cost of living and financial security of U.S. Citizens. There are large datasets that measure poverty and health concerns but we want to develop a better visual representation of these while overlaying additional interesting datasets.

Datasets

Individual Income Tax - Zip Code Statistics

https://catalog.data.gov/dataset?tags=elections

https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi

Along with datasets on…

- Election Data
- Health Statistics
- and more…

Proposed Work

- Data cleaning:
  - Getting rid of unnecessary attributes from the income tax statistics that don't have relevancy to gauging level of comfortability across the nation, and making sure separate files only contain relevant data as well.
- Data preprocessing:
  - We may not have to do a lot of pre-processing because most of our data is in integers and the majority of our project's result will be based on integers.
- Data integration:
  - To integrate our large dataset of income tax statistics with our proposed separate datasets, we will need to confine smaller datasets to specific US regions and take averages of certain attributes where we don't have the number of regional data points to match the number of income statistics.

List of Tools

-Jupyter Notebook

-WEKA data visualization

-Plotly

Evaluation

There can be interesting patterns to be found by overlaying several datasets on top of our US individual income tax statistics; with election data we can see historical trends of regional migration and ideological changes over time, real estate sales and cost-of-living variables can show financial stability or lack thereof, and more.