A Brief Analysis of Employment and Poverty in the United States

A Report by: The Data Vikings

Overview:

- Introduction
- Questions → The questions we had regarding our data.
- Datasets → The list of datasets that were used and where we found them.
- (E)xtract - (T)ransform - (L)oad → An overview of our basic ETL process.
- Technologies & Languages → A list of all the different technologies we used throughout the project with the coding language used for that technology.
- Machine Learning → An overview of the ML model used.
- Conclusion

Introduction:

For our Capstone Project, we decided to look into data surrounding employment and poverty. Our goal was to not only examine the two concepts individually, but to use data to show any relationships between the two. To gain a deeper understanding into this data, we divided our analysis into three different levels. We examined employment and poverty at the national, state and county level. This report outlines some of the core details surrounding our project.

Questions:

- National Level:
    - How has the poverty rate changed over time?
    - How does median income compare to the poverty threshold over time?
    - In the year 2022, how many occupations have an annual median income of 100k plus?
- State Level:
    - Do certain geographical areas tend to have higher poverty?
    - Do areas with generally higher or lower poverty tend to have different proportions of employees working in different job sectors?
    - How have poverty rates changed relative to income in recent years?
- County Level:
    - Is there a correlation between income and poverty rate?
    - Is there a correlation between unemployment and poverty rate?

Datasets:

Dataset	Website Obtained From
Employment	https://www.bls.gov/oes/tables.htm
Unemployment	https://apps.deed.state.mn.us/lmi/laus/Default.aspx
Poverty (API Call)	https://www.census.gov/data/developers/data-sets/Poverty-Statistics.html
Poverty Threshold	https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines
Location and Population	https://simplemaps.com/data/us-counties

(E)xtract - (T)ransform - (L)oad:

- Extract:
  - All data was either downloaded from websites or collected through an API call.
- Transform:
  - A major transformation that took place is the appending of the employment tables as each table was for a specific year. Other transformations that took place were renaming columns, changing data types, filtering, etc.
- Load:
  - Data was either loaded to a SQL database using Azure Databricks, or loaded to a CSV file.

Technologies & Languages:

- Visual Studio Code - Python/Pandas
- Azure Databricks - Python
- Azure Data Studio - SQL
- Power BI

Machine Learning:

- Models tested → Linear Regression, GXBoost, Random Forest
- Model used → Random Forest
  - This model predicts the poverty rate of all ages.

Conclusion:

While our findings may not have been conclusive, they do offer some insight into the state of poverty and employment in the United States, and these insights can be used to offer some suggestions for future research on this topic. Location was found to be the number one predictor of poverty in the United States, but more information is needed to determine why poverty rates can vary so greatly in different locations, most notably when poverty is higher in one location than in a different location with lower income. We suggest that additional research be conducted looking at geographical factors such as health indicators, socioeconomic indicators, and governmental factors. Given more time, we find it would be worthwhile to do additional research to find other potential topics that may affect employment and/or poverty.