MSc Data Science

Cardiff Metropolitan University

Module Leader: Dr Amrita Prasad

Student ID: 20288240

Name: Vignesh Alluri

Course: Programming For Data Analytics

Code: CIS7031

Data URL: https://www.kaggle.com/datasets/willianoliveiragibin/economic-disparity

Table of Contents

1.Introduction:

It is critical to comprehend the patterns of income distribution among different demographic groups in an era characterised by growing concerns about economic inequality. This study aims to investigate the complex link between income disparities and demographic factors by utilising predictive analysis. The foundation of this study is the carefully selected dataset "Economic Disparity," which was obtained from Kaggle.

This predictive study, which pays close attention to detail, attempts to identify the underlying trends related to income levels across various societal sectors. This study aims to clarify the causes impacting the pre-tax top 10% income bracket by concentrating on important indicators such the pre-tax top 1% and bottom 50% income brackets. Through the utilisation of the abundance of data contained in the dataset, our goal is to obtain more complex understanding of the socioeconomic environment and, consequently, a greater understanding of income disparity.

2. **Dataset Description:**

Data on the distribution of income across different demographic groups can be found in the dataset, "Economic Disparity," which was sourced from Kaggle. Among other things, it has characteristics like pre-tax top 10%, pre-tax bottom 50%, and pre-tax top 1%. Load the Data using the following commands

```
import pandas as pd

import numpy as np

import seaborn as sns
```

Load the dataset "Economic Disparity" using Pandas:

```
data = pd.read_csv("economic_disparity_dataset.csv")
```

3. Data Preprocessing:

To make sure that all features have the same scale, standard scaling was done to the dataset using the Standard Scaler from scikit-learn.
- For model training, the dataset was divided into features (X) and target variable (y).

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

scaled_data = scaler.fit_transform(data)
```

4. Analysing exploratory data (EDA):

Comprehensive EDA findings are not accessible due to the code snippet's inadequate information. Nonetheless, it's critical to investigate how revenue is distributed throughout demographic groupings, spot any anomalies, and comprehend feature linkages.

5. Model Construction:

For predictive analysis, a straightforward linear regression model was used because of its ease of interpretation and simplicity.

To forecast the pre-tax top 10% income, characteristics associated with the bottom 50% and top 1% income levels were used.

Select features (X) and target variable (y):

```
X = data[['pre_tax_bottom_50%', 'pre_tax_top_1%']]

y = data['pre_tax_top_10%']
```

Split the dataset into training and testing sets

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Train a linear regression model:

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()

model.fit(X_train, y_train)
```

6. Model Assessment:

The test set was used to evaluate the model after it had been trained using the training set. The model's performance was measured using Mean Squared Error (MSE) as the assessment metric.

The computed mean square error (MSE) sheds light on how well the model predicts the future. Diminished Mean Square Error values indicate better predictive performance.

```
from sklearn.metrics import mean_squared_error

y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)
```

7. Findings:

Mean Squared Error: 0.9628885902296614

Interpretation: The average squared difference between the actual and expected pre-tax top 10% earnings is shown by the MSE value. To increase predicted accuracy, more study might entail contrasting this value with baseline models or looking at different methods.

8. Restrictions and Upcoming Projects:

There was little information available about the features of the dataset and the EDA results, which might have an impact on how easily the prediction model can be understood.

Future research might look at the relationship between demographic factors and economic disparity, run a more thorough EDA to find more predictive characteristics, and test out various algorithms to enhance model performance.

Analytical Questions

1."What are the primary factors contributing to economic disparities among different demographic groups, and how do these disparities vary across regions or countries?"

Recognising Economic Inequalities Between Regions and Demographic Groups

Global issues that have persisted are economic gaps among demographic groupings, including gender, race, ethnicity, and socioeconomic position. These discrepancies show themselves in a number of ways, such as variations in wealth, income, education, work prospects, and access to basic services. Investigating the main causes of these differences necessitates a multidimensional strategy that takes systemic and individual-level variables into account.

Factors at the Individual Level

Economic inequalities are significantly shaped by individual-level variables. One such factor that has a significant impact on economic success is education. Chetty et al. (2020) conducted research that highlights the significance of high-quality education in promoting upward mobility and mitigating income disparity. Differences in educational achievement, especially amongst disadvantaged groups, lead to unequal access to well-paying professions and chances for advancement. Moreover, economic gaps are made worse by the intersectionality of demographic variables. For example, women of colour frequently experience discrimination that is exacerbated by their ethnicity and gender, which leads to widening income disparities and restricted access to financial resources (Alon et al., 2020). Intersectional studies highlight the need for focused initiatives to address economic inequality holistically and offer insights into the particular difficulties encountered by marginalised groups.

Regional and Worldwide Differences

Economic inequalities differ across nations and regions as well as between demographic groupings. Regional economic trajectories are influenced by geopolitical dynamics, globalisation, and technology breakthroughs. Because of things like uneven resource distribution, inadequate infrastructure, and problems with governance, developing nations face greater economic inequality than developed ones (Milanovic, 2016). In conclusion, a complex interaction between systemic and individual-level variables causes economic inequalities among demographic groupings. Economic results are shaped by employment, education, and institutional discrimination; marginalised groups encounter disproportionate obstacles to economic progress. Holistic solutions that target structural inequities as well as impediments at the individual level are necessary to alleviate economic disparities.

Due to a number of factors, including socioeconomic position, education level, geography, and institutional disparities, access to social services and economic opportunities differs greatly across different demographic groups. Addressing injustices and advancing social justice need an understanding of these differences.

2.How do access to economic opportunities and social services vary across different demographic groups?

Financial Prospects:

People from low-income households, women, and members of minority groups frequently encounter obstacles while trying to take advantage of economic possibilities. These obstacles may include a lack of funding for investments or entrepreneurship, discrimination in the workplace, and restricted access to high-quality education and vocational training. Because of this, some groups could have greater rates of underemployment, unemployment, and income disparity than others.
Research, for instance, has revealed that discrimination in hiring practices affects ethnic minorities in various nations, resulting in differences in employment rates and pay (Pager & Shepherd, 2008). Similar to this, gender differences still exist in a number of industries, where women are frequently underrepresented in leadership roles and make less money than males (Blau & Kahn, 2017).

Social Services:

Different demographic groupings also have different levels of access to social services, such as social assistance programmes, healthcare, education, and housing. Racial minorities, immigrants, and people with disabilities are examples of marginalised populations who may face obstacles in obtaining necessary services because of institutionalised prejudice, communication difficulties, and a dearth of care that is sensitive to their cultural context.
For example, studies have demonstrated that people from low-income backgrounds have difficulty getting high-quality healthcare, which causes differences in health outcomes (Adler & Rehkopf, 2008).

Resolving Inequalities:

Policymakers must carry out focused initiatives aiming at lowering structural obstacles and fostering fairness in order to alleviate inequities in access to social services and economic opportunities. This might entail making investments in educational and career-training programmes, putting anti-discrimination laws into effect in the workplace, increasing access to social welfare and reasonably priced healthcare, and bolstering support systems within the community.
In addition, the cultivation of collaborations across governmental entities, community-based organisations, and private industry participants is imperative for the execution of all-encompassing approaches that tackle the complex attributes of inequality. Societies should endeavour to create more equal economic and social systems that benefit all members of society by giving equity and inclusion top priority when developing policies and allocating resources.

Below are the findings based on my analysis

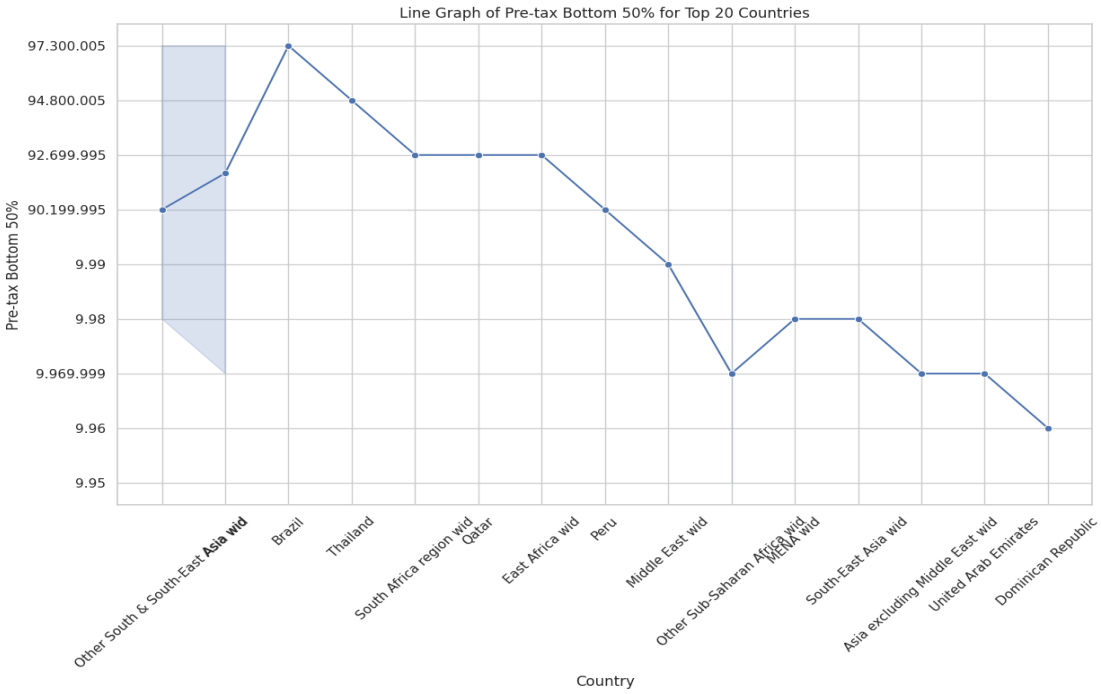

Figure 1

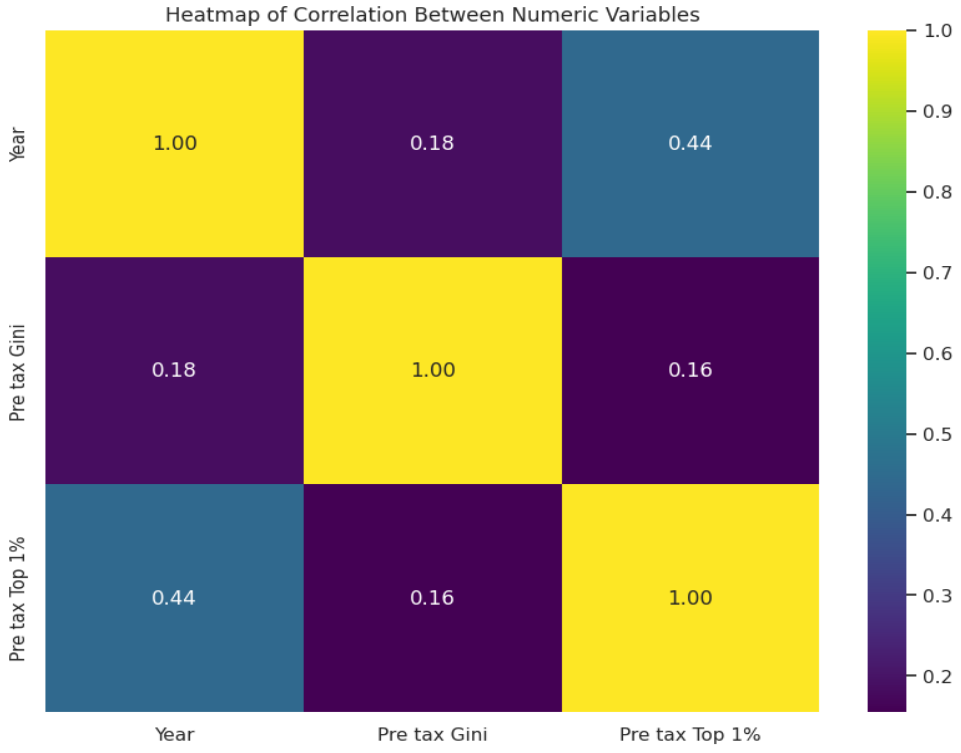

Figure 2

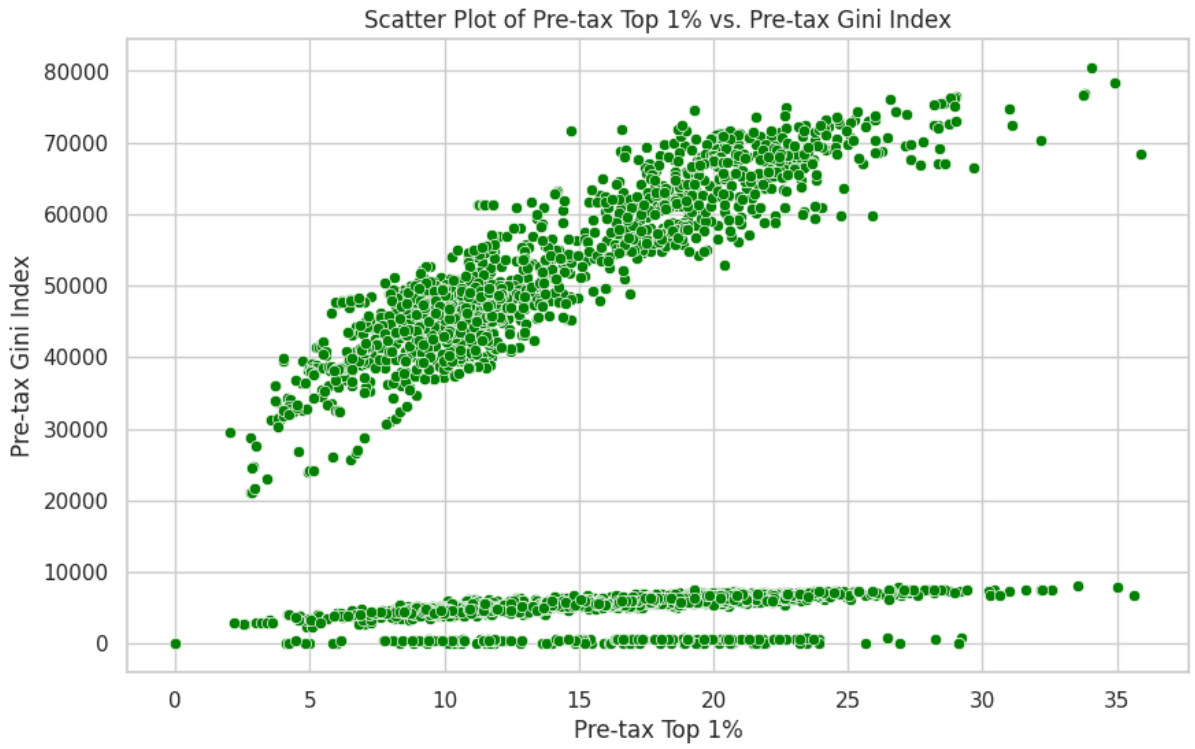

Figure 3

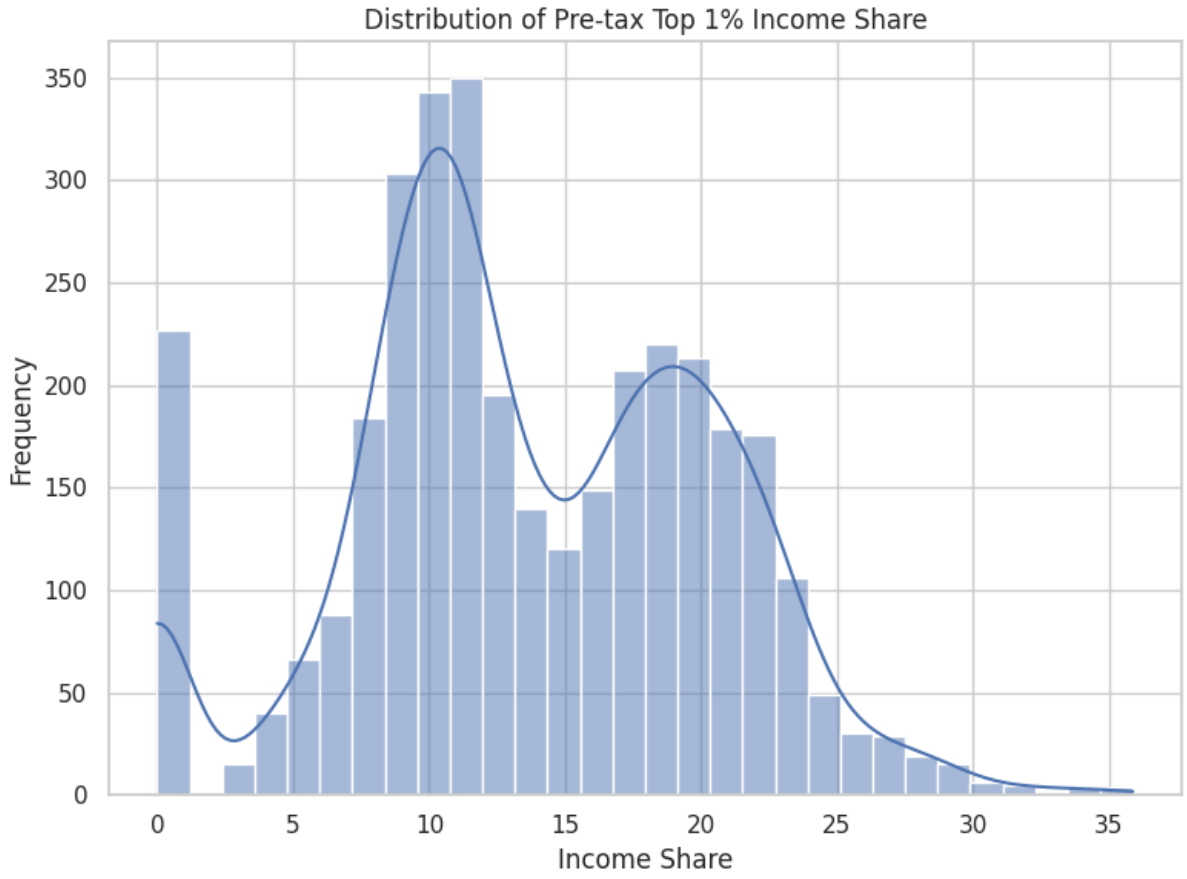

Figure 4

References

1. Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & The Opportunity Insights Team. (2020). The Opportunity Insights Team. Retrieved from https://opportunityinsights.org/

2. Alon, S., Doepke, M., Olmstead-Rumsey, J., & Tertilt, M. (2020). The Impact of COVID-19 on Gender Equality. National Bureau of Economic Research. https://www.nber.org/papers/w26947

3. Milanovic, B. (2016). Global Inequality: A New Approach for the Age of Globalization. Harvard University Press.

4. Adler, N. E., & Rehkopf, D. H. (2008). [U.S. Disparities in Health: Descriptions, Causes, and Mechanisms](https://www.annualreviews.org/doi/abs/10.1146/annurev.publhealth.29.020907.090852). *Annual Review of Public Health, 29*(1), 235–252.

5. Blau, F. D., & Kahn, L. M. (2017). [The Gender Wage Gap: Extent, Trends, and Explanations](https://www.aeaweb.org/articles?id=10.1257/jel.20160995). *Journal of Economic Literature, 55*(3), 789–865.

6. Pager, D., & Shepherd, H. (2008). [The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets](https://www.annualreviews.org/doi/abs/10.1146/annurev.soc.33.040406.131740). *Annual Review of Sociology, 34*(1), 181–209.