Impact of Population and Unemployment on Per Capita Income

ARUNKUMAR SUDHAKARAN
NAIR
School Of Computing
National College Of Ireland
Dublin, Ireland
x19153317@student.ncirl.ie

EVLIN BRIDGIT EMMANUEL
School of Computing
National College Of Ireland
Dublin, Ireland
x19153813@student.ncirl.ie

ARATHY VIJAYAN
School of Computing
National College Of Ireland
Dublin, Ireland
x19148615@student.ncirl.ie

Abstract— **Identifying the relationship between population growth and per capita income has been a debate topic among the researchers for years. The instrumental variables projections show that a 1% rise in per capita income over ten years raises population density in countries by about 0.1%[1]. Likewise, unemployment is another factor that is equally important while analyzing a country's economy. The question of employment was to the degree that demand for labor or job opportunities in the nation did not balance the availability of labor. The effects of shocks on national income due to population and unemployment are discussed in this paper.**

I. INTRODUCTION

The world population has risen unprecedently in the past half-century. The environmental, socio-political, and economic problems linked to the massive and rapid population growth of our time is an issue of concern for many development planning professionals. Leading theories suggest that a negative correlation between income and population growth may be motivated by technological developments that not only raise incomes but also raise childbearing opportunities [1]. If you break GDP by the resident amount, you wil get the GDP per capita or per capita income. It is one of the main definitions for national income to equate the wealth of nations. Unemployment is one main factor affecting the per-capita income of a country. The influential US economist Arthur Okun has found that growth of GDP has an impact on unemployment such that a 2% rise in GDP causes unemployment to fall by 1%[2]. Unemployment and population growth are strongly interrelated to each other and per capita income is of no less importance. Also, per capita income is the main factor used to measure a country's standard of living. This was the main factor which made us choose this topic.

II. RELATED WORK

The important contributions to studies related to income and population size date back to the 18th century. As described by Malthus in the year1798, income gains translated directly into population growth, with per capita income stagnant and population density only increasing. However, the demographic dynamics shifted from the Malthusian paradigm into the modern growth system during the industrial revolution, marked by economic growth and decreasing fertility[1]. Likewise, the Malthusian model, Okun's law explores the statistical link between a country's unemployment rate and its economic growth rate. Okun's rule of thumb defines the measurable relationship between unemployment rate adjustments and actual gross national product which is similar to or equal to the per capita income of a country. Okun noted that, since labor force sizes and productivity rates continue to increase, real GDP growth is usually required close to its potential growth rate only to sustain the unemployment rate. Therefore, the economy needs to expand at a pace higher than its ability to reduce the unemployment rate [2].

This paper also adds to the literature on the health and fertility impacts of wages. In 1997, Lee explores research on the relationship between salary-fertility in pre-industries and suggests that pay adjustments are more likely to be mistaken for structural and technical development in these economies than in developed countries. For most countries, he shows positive fertility income elasticity. In 2009 Cotet and Tsui analyzed how petroleum exploration affects the population and health outcomes of countries. In countries with and without significant oil finds during the 1960s, they are contrasting improvements in these tests. It is interesting to note that oil discoveries in 8 countries have seen faster population growth and lower mortality among children [1].

The European Union (EU) unemployment rate is one of the main economic and social challenges. It is becoming very difficult to compare individual EU countries. This issue is common not only for economically poorer countries such as Greece but also for the Czech Republic, where in recent years the unemployment rate has been stagnating. For a long time, the Czech Republic has been below its capacity for job growth. This condition can be better explained by the history of Austria and Germany, the two neighboring nations. Austria is more like the Czech Republic in terms of history, geography and demography and demography [3].

III. RESEARCH QUESTION

Our overall analysis plans in explaining the following research questions in our report:

1. What is the effect of population growth and unemployment rate on per capita income.
2. Region wise effect of unemployment on percapita income.
3. Effect of region wise population growth on per capita income.

IV. METHODOLOGY

A. Dataset Description:

For extracting data for our analysis, we have used the publically available data repository. All the 3 datasets were downloaded from public.opendatasoft.com in JSON format.

1)Population by Urban Area

The Population dataset shows the annual estimate of the resident population in the Metropolitan Areas of the United States. The data was published by the U.S. Census Bureau.

2) Local Area Unemployment Statistics

The dataset gives an overview of the unemployment rate in the Metropolitan areas from 2014 to 2018. The data was published by the Bureau of Labor Statistics.

3) Real Per Capita Income by Metropolitan Statistical Area

The dataset gives an audit regarding the per capita income for a person. Time and the Metropolitan area are the other factors which are considered in the dataset. The data was published by GeoFRED.

A. Data Analysis Flow:
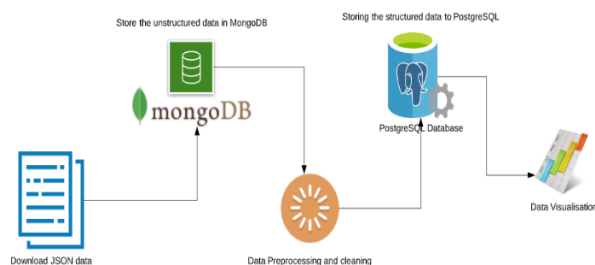
Data Flow Diagram 1

i. Data Gathering:

The data used for the project was directly downloaded from the website public.opendatasoft.com. All the three datasets were downloaded from the above-mentioned website in JSON format.

ii. Storage of Unstructured JSON Data:

Since the downloaded data was large and was unstructured, we chose the schema-less database MongoDB to store our data. Each of the team members chose one data set respectively and stored them in MongoDB installed in their system respectively.

iii. Data Transformation:

Each team members were supposed to work on the datasets which they have chosen. In this step, the data was cleaned and transformed into a structured format. For this, the data was extracted from MongoDB and converted into a data frame first. The data cleaning operations such as dropping of the 'NA' values, unnecessary fields and splitting of the data were done on the data. The data was then converted to a CSV file. The data cleaning and transformation part was necessary since all the fields in the original dataset are not required for visualisation

iv. Structured Data Storage:

The entire cleaned data was then stored in PostgreSQL. Each team member installed PostgreSQL locally in their system respectively and stored the cleaned data. Since PostgreSQL is an object-relational database, it supports user-defined objects and their behaviours. This makes PostgreSQL stand out among the other databases like MySQL and SQLite.

v. Data Analysis:

The data was extracted from PostgreSQL and was combined as a CSV file which contains the data of all the team members. The file was converted to data frame using Python Pandas library and data cleaning was performed on the final data. K-fold cross-validation was performed on the final dataset. The k value was chosen as 3 for a total number of rows 1038 in the final data. Linear regression was performed using Scikit-Learn.

The output of performing KNN Regression on population and unemployment rate is as follows.

```
[-0.004789608921169641, -0.00666460684213388, -0.008084087668874362]
```

The output of performing KNN Regression on population and per capita income is as follows.

```
[-0.0049533221733613075, -0.0008679205366188826, -0.0007153931034693528]
```

On performing the same on unemployment rate and per capita income gave the following results .

```
[-0.013018964603139782, -0.002302332602676316, -0.0005948269847542154]
```

Performing KNN Regrssion gave the best results for unemployment rate and per capita income.

vi. Data Visualisation:

The next step was to visualize the final combined data. For

data visualisation part, the python plotly express library was used. We have plotted the Bar graph and the Scatter Dot Diagram of the Combined data to Analyse the overall effect of Population and Unemployment Rate in Per Capita Income of a Region.

V.   RESULT

Through the data visualisation process, we were able to infer the following conclusions:

- **Effects of Unemployment, Population Growth and Per Capita Income on Different Regions on the US.**

The below graph shows the region-wise population rate for different regions of the United States. The graph was created by using plotly.express.library.
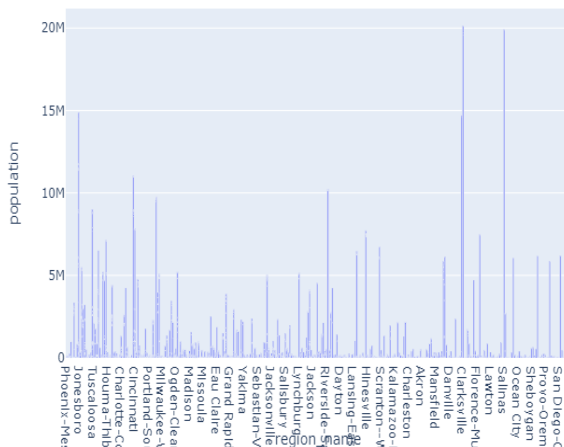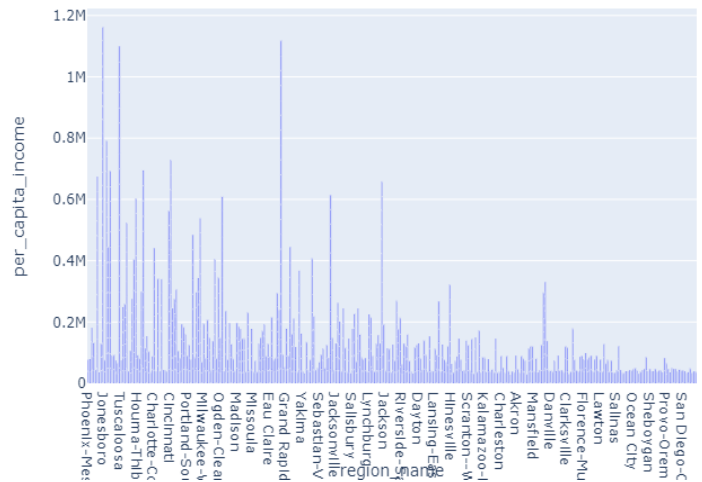

From the above diagram, it can be observed that the regions, Salinas and Clarksville have a population of about 20M and that of the region Jonesboro is about 15M. From the graph, it can be inferred that the population of the other regions of the US are less than 15M and the majority of the regions fall under the category of 10M.

The second graph which can be analysed is that of unemployment_rate to the regions in the United States. On analysing the below-given diagram, it can be inferred that the region Jonesboro has an Unemployment rate of around 140-160, while the region Grand Rapid's unemployment rate is 140 and that of Tuscaloosa is around 100-120. However, it can be also observed that the rest of the regions have an unemployment rate of less than 80.
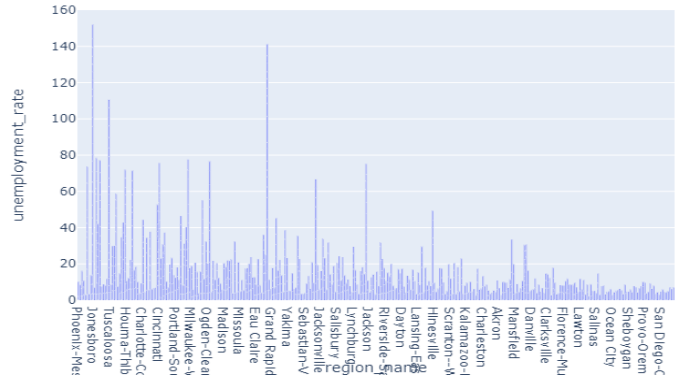

The third graph which can be analysed is that of the Per Capita Income to regions in the United States.

From the above diagram, it can be observed that the regions Grand Rapid, Jonesbroand and Tuscaloosa has the highest Per Capita Income among the US cities. The Per Capita Income for the above-mentioned regions comes around 1M to 1.2M.

For a better comparison, we have combined the graphs which represent the population, unemployment_rate and per capita income into a single plot below. From the comparison, we can infer that the areas such as Jonesboro which is one of the densely populated areas with a population of around 15M also has the highest unemployment_rate of 140-160. Further, it can be also observed that the Per Capita Income of the region lies in the highest margin which is 1M to 1.2M. While considering the region, Tuscaloosa we can see that the Unemployment rate is around 100-120 and Per Capita Income is 1M to 1.2M. However, if we consider the region, Clarksville, the population of the region falls in the highest bracket but the Per Capita Income is below 0.2M.
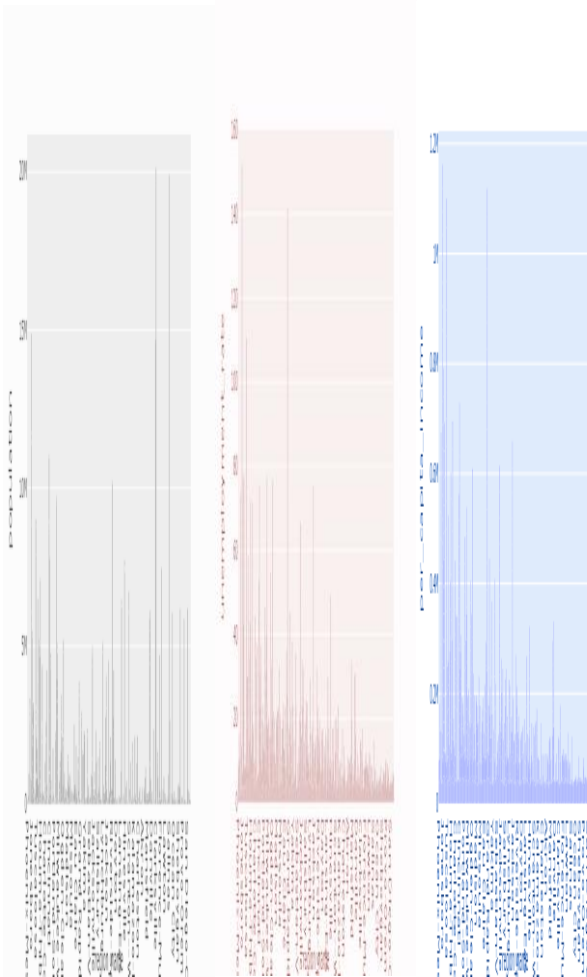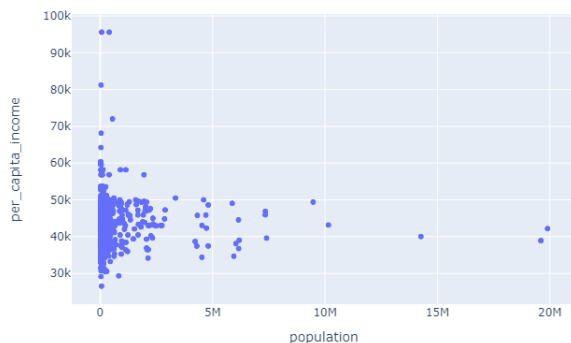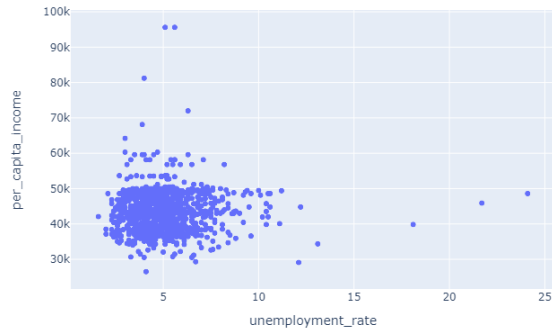
An overview regarding the relationship between the Unemployment rate and the Per Capita Income can be analysed from the below scatter plot:

The above scatter plot also shows a strong correlation between the variables. The unemployment rate and per capita income are strongly correlated to each other.

VI. CONCLUSION AND FUTURE WORK

After analysing the data and visulaising it, we were able to conclude that the regions Salinas & Clarksville is densly populated and Joneboro comes behind them in the list. Whereas, Jonesboro and Tuscaloosa has the highest per-capita income in the country .Despite of being highly populated and having the highest per-capita income, Jonesboro is the region with the highest unemployment rate.

The research was conducted for analysing the impact of population and unemployment have on per-capita income. The future scope of this project will be to extend the scope of this analysis to different parts of the world and getting insights of how to improve the per-capita income on different parts of the world. This can inturn lead to an improved economy and standard of living.

VII. REEFERENCE

[1]*Ftp.iza.org*, 2020. [Online]. Available: http://ftp.iza.org/dp7422.pdf. [Accessed: 23- Apr- 2020]

[2]"Unemployment and Economic Growth: Okun's Law", *Investopedia*, 2020. [Online]. Available: https://www.investopedia.com/articles/economics/12/okuns -law.asp. [Accessed: 23- Apr- 2020]

[3]2020. [Online]. Available: https://www.researchgate.net/publication/282466990_The_ Effect_of_GDP_per_Capita_on_Employment_Growth_in_ Germany_Austria_and_the_Czech_Republic_Macroecono mic_Analysis. [Accessed: 23- Apr- 2020]

[4]"Linear Regression in Python with Pandas & Scikit-Learn", *Medium*, 2020. [Online]. Available: https://becominghuman.ai/linear-regression-in-python-with-

• **Summarising the Result:**

The below scatter plot summarises the relationship between Population and Per Capita Income :

From the above diagram, it can be observed that the Per Capita Income and the Population are highly correlated if the outliers are ignored.

pandas-scikit-learn-72574a2ec1a5. [Accessed: 23- Apr-2020]