**CSCE 5320: Scientific Data Visualization**

**Final Project - Increment 2**

**Project Title: Unemployment Data Analysis**

**Team Members:**

Venkata Naga Prasanna Kumar Bitra (11525657)

Sravanthi Kande (11547212)

Yamuna Bollepalli (11552426)

Prasanth Damarla (11525662)

**Goals and Objective:**

**Motivation:**

A lot of people in this world are unemployed and without jobs. Our aim is to analyze the unemployment across the world and how a lack of job limits people's ability to live their life.

**Significance:**

Unemployment is a big problem in the world which if not addressed will lead to other problems like starvation, robbery, etc. The unemployment rate is the percentage of people who doesn’t have any occupation. Unemployment affects the economic growth. We need to know the detailed analysis of the unemployment to take the significant measure to decrease the unemployment.

**Objective:**

To analyses and find significant details which can help the 5W's and the root of this issue. To find the who? Where? What? When? Why? of unemployment.

**Features:**

We are aiming to visualize the data on charts, graphs and plots for demonstrating the various factors of the unemployment.

**Introduction:**

**Domain:**

According to the OECD (Organization for Economic Co-operation and Development), unemployment is defined as those over a certain age (typically 15) who are neither employed nor self-employed but are currently looking for work during the reference period.

The unemployment rate, which is the number of jobless people as a percentage of the labor force, is used to calculate unemployment (the total number of people employed added to those unemployed).

Unemployment can have many sources, such as the following:

* new technologies and inventions
* the status of the economy, which can be influenced by a recession
* competition caused by globalization and international trade
* policies of the government
* regulation and market

A country's fiscal policies, for example, can affect unemployment and the state of the economy. Furthermore, through monetary policy, a country's monetary authority, such as the central bank, can impact the availability and cost of money.

Aside from unemployment theories, a few classifications of unemployment are used to better precisely model the consequences of unemployment on the economy. Structural unemployment, frictional unemployment, cyclical unemployment, involuntary unemployment, and classical unemployment are some of the most common types of unemployment.

Structural unemployment focuses on the economy's underlying issues and labor market inefficiencies, such as a mismatch between supply and demand for laborers with the requisite skill sets. Causes and solutions related to disruptive technologies and globalization are the focus of structural arguments. Frictional unemployment discussions center on individuals' choice decisions to work based on their own valuation of their own work and how that compares to current wage rates, as well as the time and effort required to find a job. Job entrance thresholds and salary rates are frequently mentioned as causes and solutions for frictional unemployment.

According to the International Labor Organization (ILO), 172 million individuals were unemployed worldwide in 2018 (representing 5% of the global workforce). On account of the trouble in estimating the joblessness rate by, for instance, utilizing studies (as in the United States) or through enlisted jobless residents (as in a few European nations), measurable figures, for example, the business to-populace proportion may be more appropriate for assessing the situation with the labor force and the economy assuming they depended on individuals who are enrolled, for instance, as citizens.

The targeted area of this project is data visualization of the Unemployment in the world and its analysis. We are visualizing the different types of data analysis for the user to understand the data very clearly and easily. It helps to get to know the stats of unemployment data in the world very easily to the user.

**Background:**

Unemployment, according to the OECD (Organization for Economic Co-operation and Development), is people above a specified age (usually 15)[2] not being in paid employment or self-employment but currently available for work during the reference period.[3]

Unemployment is measured by the unemployment rate, which is the number of people who are unemployed as a percentage of the labor force (the total number of people employed added to those unemployed).[4] Unemployment and the status of the economy can be influenced by a country through, for example, fiscal policy. Furthermore, the monetary authority of a country, such as the central bank, can influence the availability and cost for money through its monetary policy.

In addition to theories of unemployment, a few categorizations of unemployment are used for more precisely modelling the effects of unemployment within the economic system. Some of the main types of unemployment include structural unemployment, frictional unemployment, cyclical unemployment, involuntary unemployment and classical unemployment. Structural unemployment focuses on foundational problems in the economy and inefficiencies inherent in labor markets, including a mismatch between the supply and demand of laborers with necessary skill sets. Structural arguments emphasize causes and solutions related to disruptive technologies and globalization. Discussions of frictional unemployment focus on voluntary decisions to work based on individuals' valuation of their own work and how that compares to current wage rates added to the time and effort required to find a job. Causes and solutions for frictional unemployment often address job entry threshold and wage rates.

According to the UN's International Labour Organization (ILO), there were 172 million people worldwide (or 5% of the reported global workforce) without work in 2018.[5] Because of the difficulty in measuring the unemployment rate by, for example, using surveys (as in the United States) or through registered unemployed citizens (as in some European countries), statistical figures such as the employment-to-population ratio might be more suitable for evaluating the status of the workforce and the economy if they were based on people who are registered, for example, as taxpayers.[6]

**Data Abstraction:**

* **Dataset (Type and Attributes)**

Dataset is taken from GitHub. Data contains the details of the unemployment in all the countries from 2010 to 2014. Data contains the attributes Country Name, Country Code, 2010, 2011, 2012, 2013 and 2014 with data types string and floats.

* **Detailed description of Dataset**

Data is all about the unemployment rate in the world from 2010 to 2014. From 2010 to 2014, the unemployment rate of each country has been noted and stated in the dataset. Data includes individual nations as well as nations grouped according to income level and supranational groups such as Arab World, EU etc. The economy is not increasing at a fast rate because of a lack of investment and a slow expansion in the manufacturing sector, which produces the most employment.

* **Detail design of Features with diagram**

**Diagram

Description automatically generated**

**Fig1: Workflow diagram of dataset**

1. **Country Name**

This contains the name of the countries like USA, India, China, Pakistan etc.

1. **Country Code**

This contains the country codes of countries like AFG – Afghanistan, AGO - Angola, Argentina – ARG etc.

1. **2010**

It contains the unemployment rate of the all the countries in the world from 2010, which is noted in the float values like AFG - 20.600000, AGO - 10.800000, ARG - 19.500000 etc.

1. **2011**

It contains the unemployment rate of the all the countries in the world from 2011, which is noted in the float values like AFG - 20.600000, AGO - 10.800000, ARG - 19.500000 etc.

1. **2012**

It contains the unemployment rate of the all the countries in the world from 2012, which is noted in the float values like AFG - 20.600000, AGO - 10.800000, ARG - 19.500000 etc.

1. **2013**

It contains the unemployment rate of the all the countries in the world from 2013, which is noted in the float values like AFG - 20.600000, AGO - 10.800000, ARG - 19.500000 etc.

1. **2014**

It contains the unemployment rate of the all the countries in the world from 2014, which is noted in the float values like AFG - 20.600000, AGO - 10.800000, ARG - 19.500000 etc.

* **Data Transformation**

The method involved with transforming crude information into a configuration or design that is more reasonable for model structure and information disclosure overall is known as information change. It's a significant stage in include designing that makes finding experiences more straightforward. The accompanying methodologies for numeric information change are log change, cutting techniques, and information scaling.

The motivation behind information change is the point at which the information dissemination is slanted, the calculation/model is bound to be one-sided. By switching information over completely to a similar scale, the calculation/model is better ready to think about the general connection between data of interest. Information change is a significant part of perception: choosing the factors to show and how much data they require is similarly just about as basic as choosing satisfactory visual encodings.

In our project we used the Unemployment rate data to visualize the various scenarios of unemployment around the world among the years 2010 to 2014. In visualizing various scenarios, we transformed the available data into different forms. From available data we have transformed that into Super National Country for visualizing the Region wise unemployment in the world like 'Arab World', 'Caribbean small states', 'East Asia & Pacific', 'European Union', 'Latin America & Caribbean', 'Middle East & North Africa', 'North America', 'OECD members', 'Other small states', 'South Asia', 'Sub-Saharan Africa', 'World'.

For Visualizing the Unemployment rate among the various income group, we transformed the available data into data Income with attributes 'High income', 'Heavily indebted poor countries (HIPC)', 'Least developed countries: UN classification', 'Low income', 'Lower middle income', 'Low & middle income', 'Middle income', 'Upper middle income', 'World'.

**Task Abstraction:**

1. **Tasks:**

* Data collection
* Tools and frameworks
* Data selection and data frame creation
* Data transformation
* Plotting visualizations
* Web page creation
* Documentation

**2) Workflow diagram with explanation:**

Below diagrams is the workflow diagram of the project. Firstly, collected the data then used the required tools for visualization, Data processing and transformation for visualization od various plots, Visualization, webpage creation and documentation.

**Diagram

Description automatically generated**

**Fig2: Workflow diagram of Task Abstraction**

Task is to plot the distribution of unemployment of all the individual countries for the 5-year duration, comparison of the countries having unemployment rate, comparison of best and worst performers in controlling the unemployment in all the countries, unemployment growth around the globe in those 5 years, unemployment rate in the different regions of the world and unemployment over the years among various income groups. For plotting the above tasks, we are using different types of visualization techniques like Box plot, Scatter plot, Geo chart, Bar chart and line chart. For various types of visualizations, we are using the different plots which visualizes the data clearly.

**Implementation using tools:**

For python coding we have used the Jupyter Notebook and Google Colab,

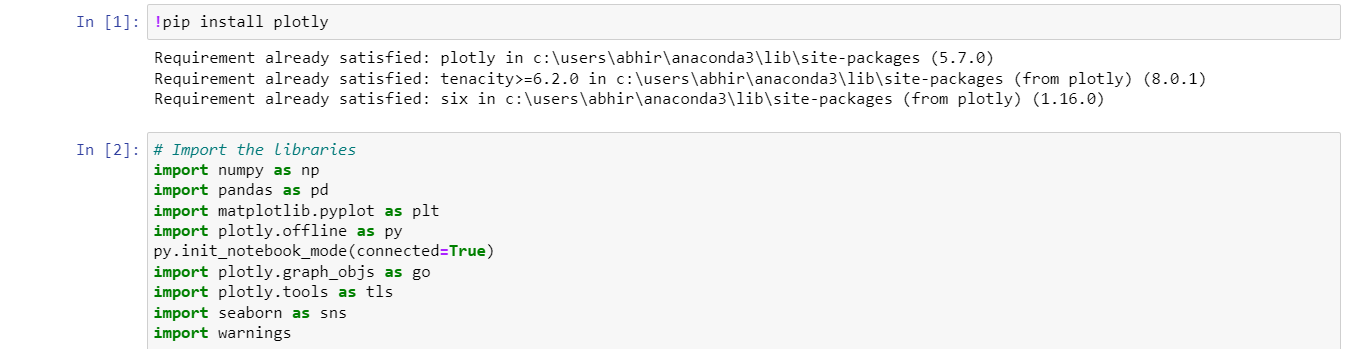
For creating the user interface, we are going to use flask APIs in visual

studio code.

* Jupyter Notebook: Project Jupyter is a project and community whose goal is to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages". It was spun off from IPython in 2014 by Fernando Pérez and Brian Granger.
* Google Colab: Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.
* Visual Studio Code: Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.
* Flask: Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

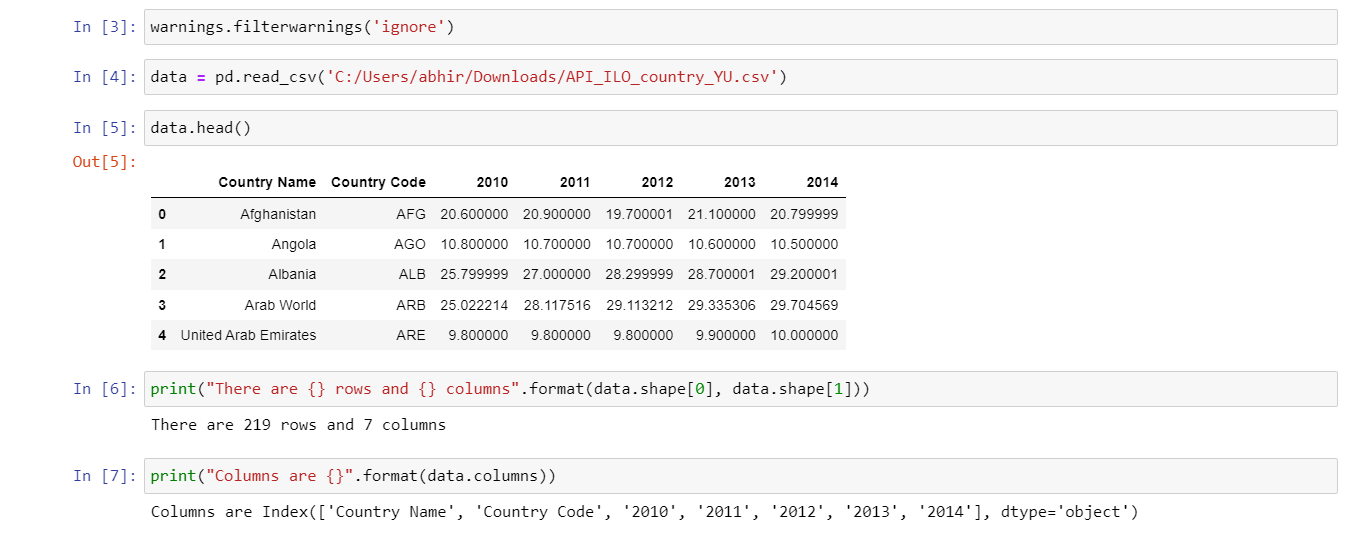
**Preliminary Results for Analysis:**

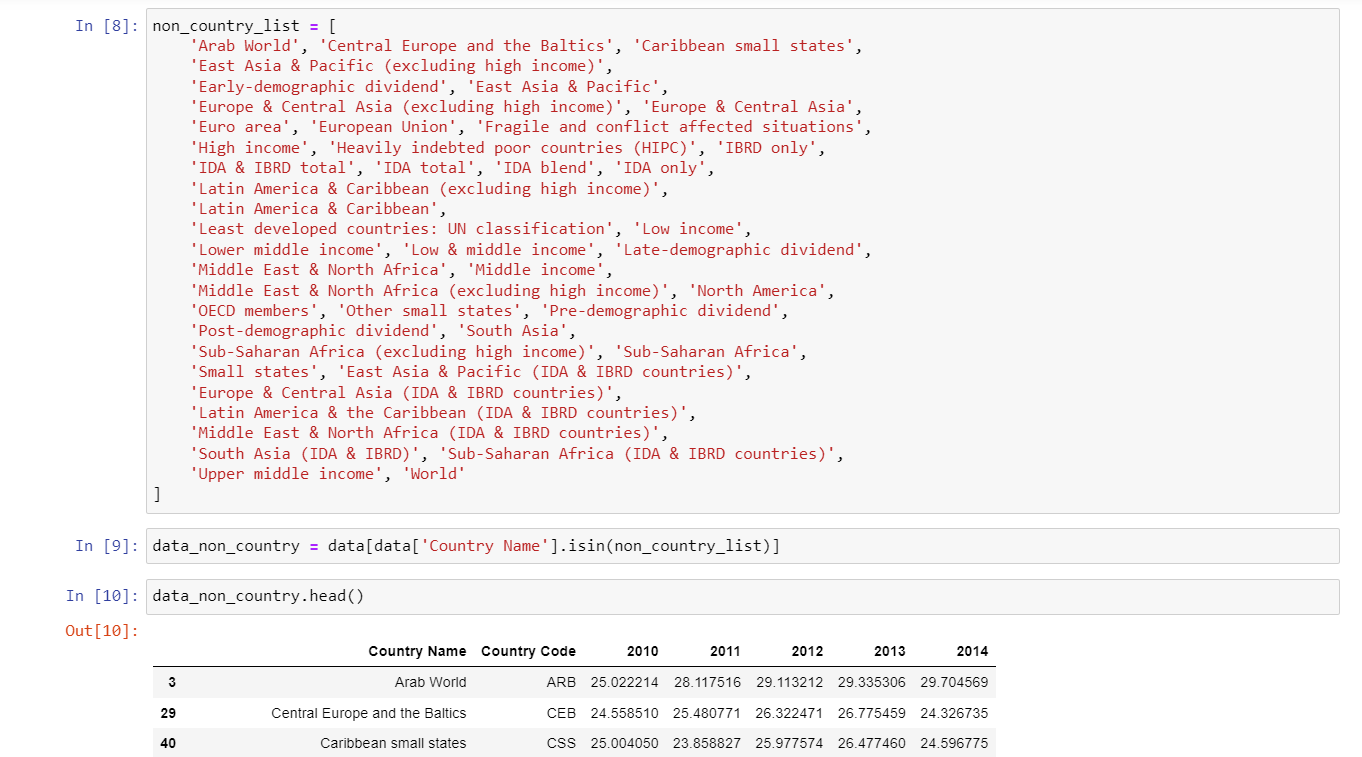
1) importing the required libraries



Imported the required libraries for visualization like NumPy, pandas, plotly, seaborn, warnings.

2) Reading and processing the data.

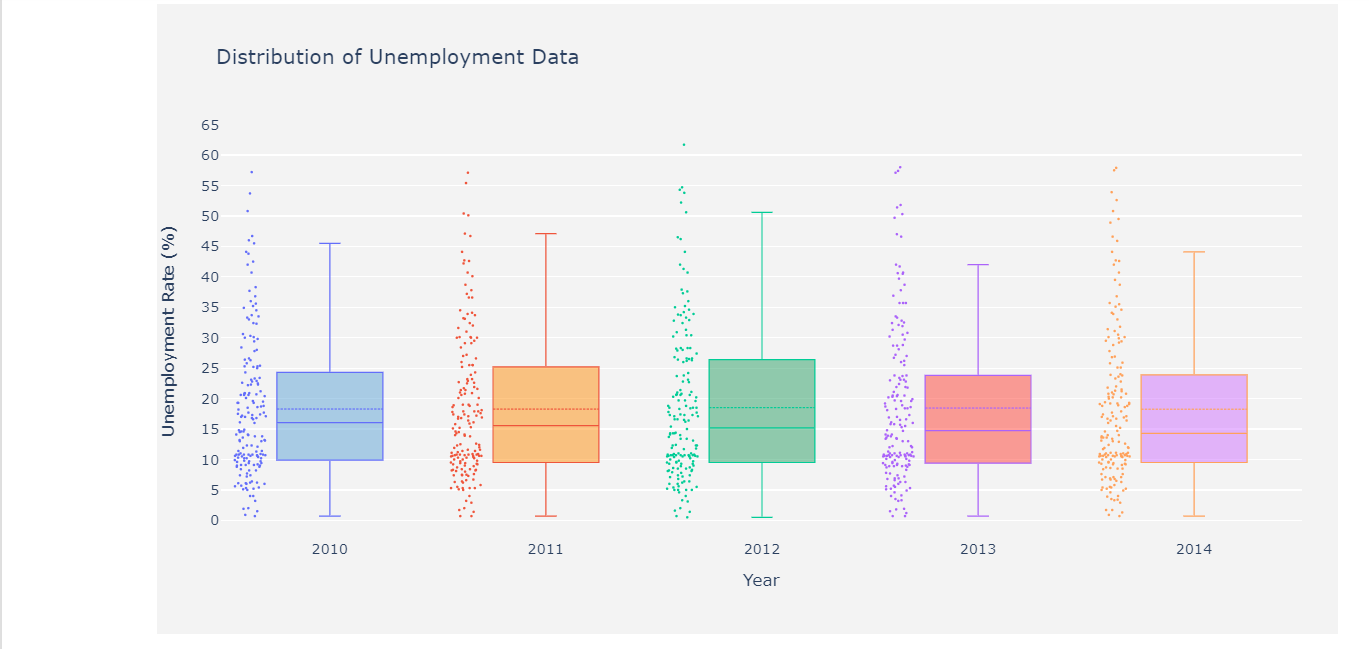




Firstly, downloaded the dataset from GitHub and read the file using the file path and displayed the data using head. Then modified the data as required like non country list and displayed the data.

3) Creating a plot for unemployment rate of all countries:

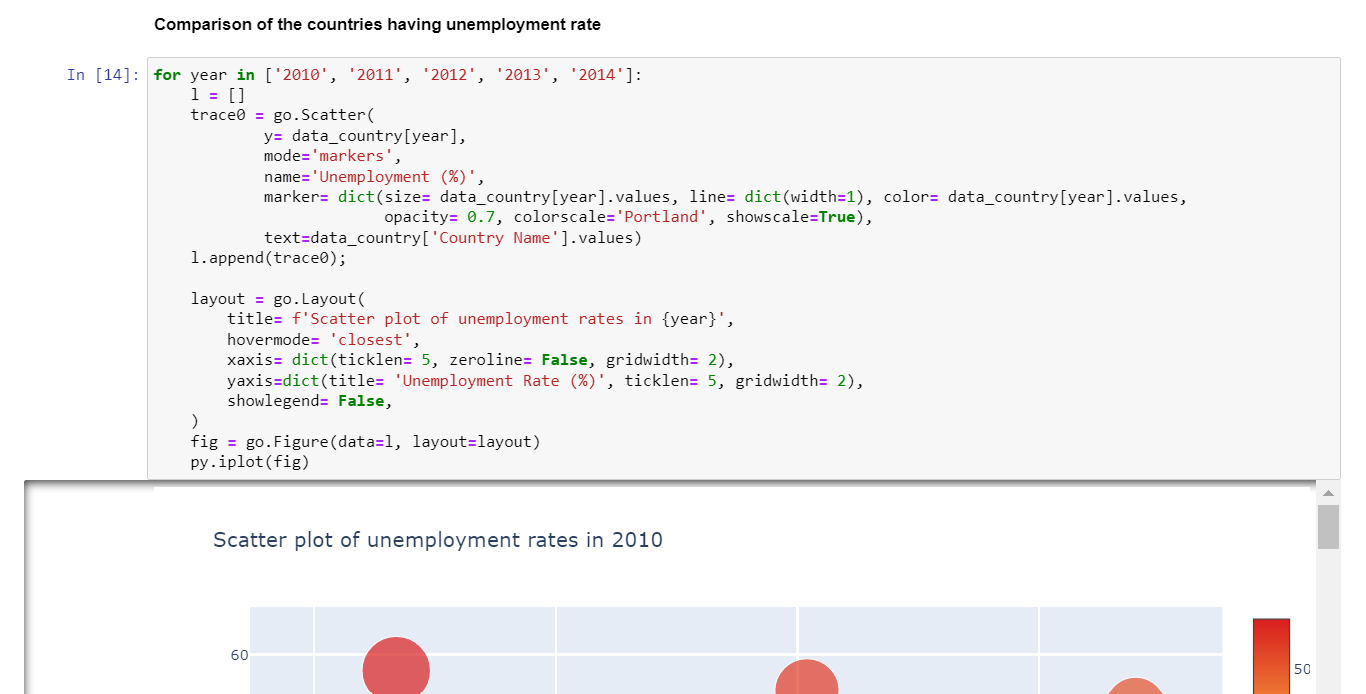


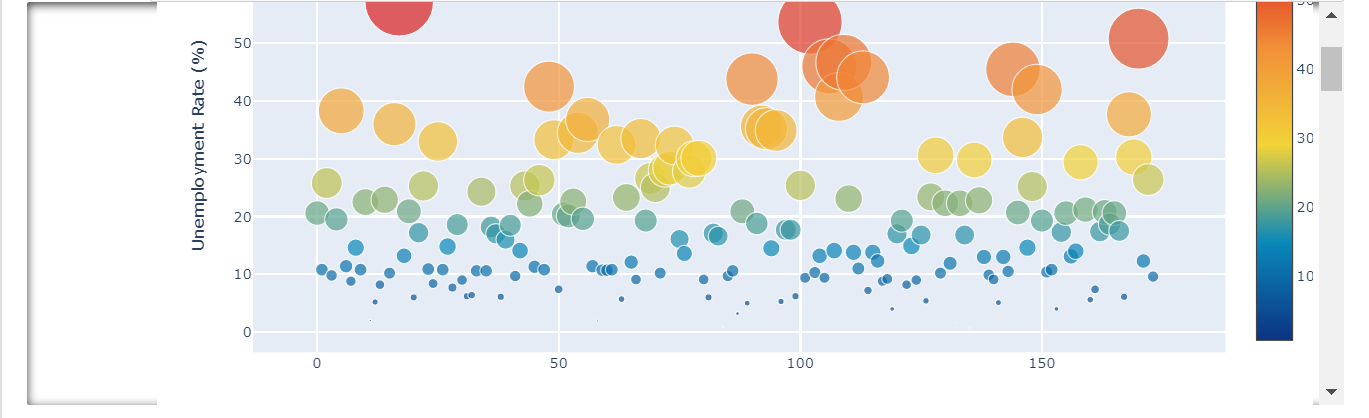


For displaying the unemployment rate in all the countries, we have used the box plot which displays the data distribution of numerical values and their percentiles and averages. In X axis, we took the years and in Y axis we took the unemployment rate. We used different colors to the years for visualization of unemployment.

We can better understand the data if we use a box plot and scatter plot together. There is a clear trend in the data over time, with a mean of 18.285 percent from 2010 to 2014 and a range of 18.2546 percent. statistical plots such as box plots, scatter plots, etc. of data from certain countries and getting a fast look at how different countries compare against one another in terms of statistical analysis.

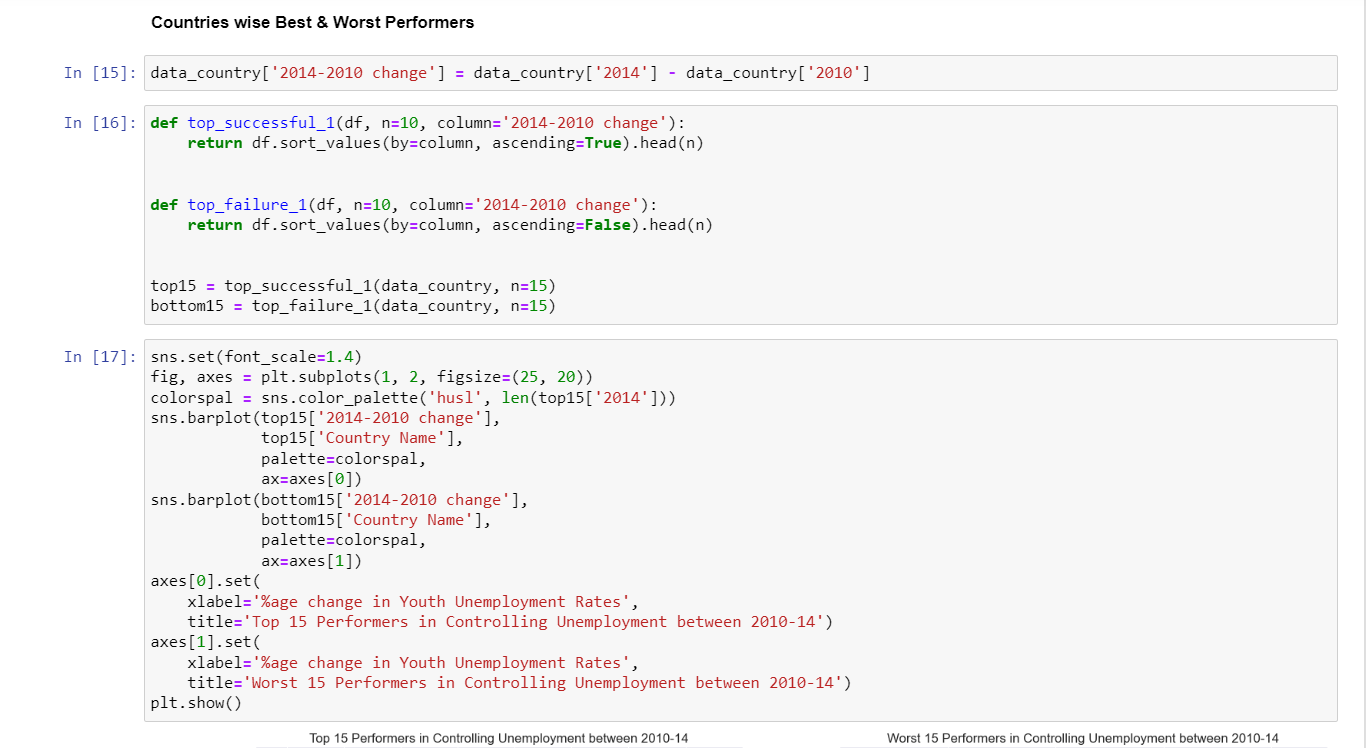
4) Comparison of the countries having unemployment rate:

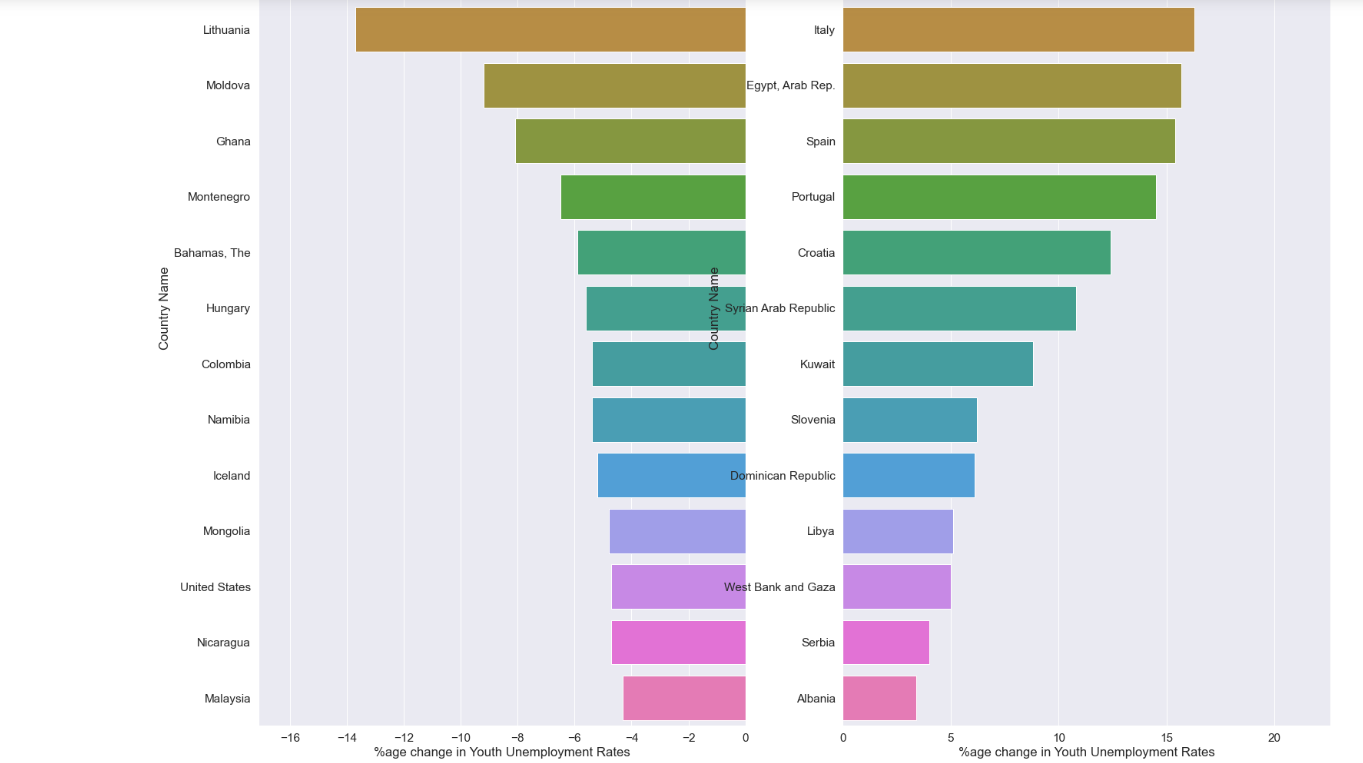




For displaying the comparison of countries having the unemployment rate visualization we have used the scatter plot. Several nations, such as South Africa and Bosnia and Herzegovina, which were already experiencing high unemployment rates in 2010 are still among the countries with the highest jobless rates in 2014. The situation in several European nations, on the other hand, has worsened with time.

5) Countries wise Best and Worst performers:

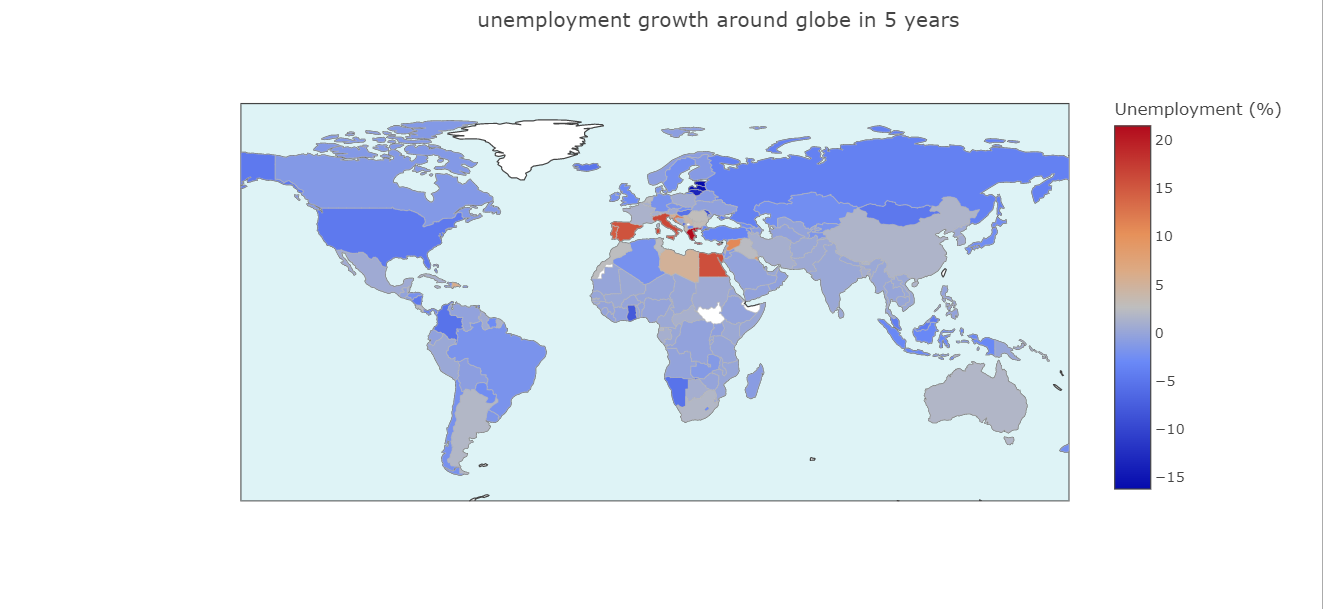




For visualizing the countries wise best and worst performers, we have used the bar graph with country names and the percentage of age change in youth with unemployment. A bar graph on the left displays the top 15 nations in terms of how well they dealt with unemployment between 2010 and 2014. Most of these nations are located in Europe, with the exception of Estonia, which had a negative growth rate of -16.29 percent despite having an unemployment rate of 17 percent, which is still greater than most other countries in the region. Estonia, Latvia, Lithuania, Moldova, and Ghana round out the top five.

6) Unemployment growth around globe from 2010 to 2014:





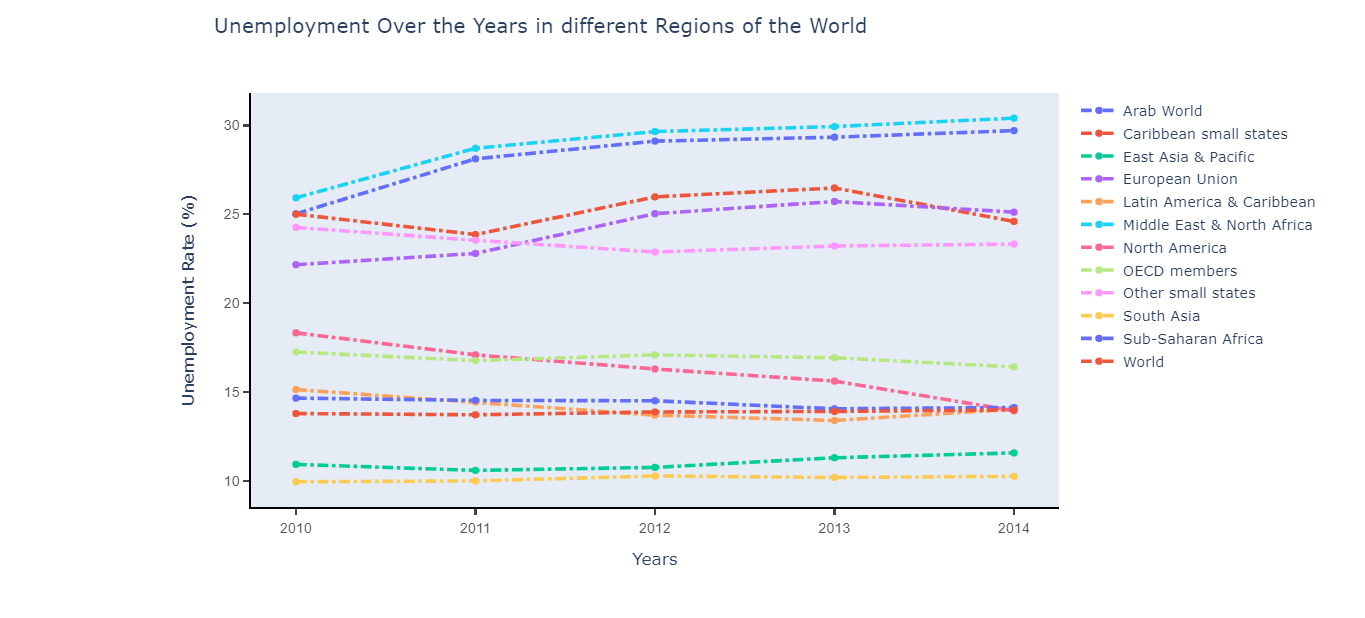
For visualizing the unemployment growth around the globe from 2010 to 2014, we have used the Choropleth map. Choropleth map is the one of the geospatial data visualization techniques.

It is created by dividing the area to be mapped into sections, for as by geographic or political boundaries, and then filling each piece with a different hue or shade. Each color or shade represents a separate variable, as well as a distinct value or range for a single variable. Choropleth maps are effective for showing data clusters across a geographic area while keeping regional borders in view. Most nations studied had either a minor positive increase or a net drop in unemployment rates over the five-year period from 2010 to 2014.

The remainder of the globe seems to be doing quite well, other from few North African and Southern European countries that have suffered a significant rise in unemployment. Just 0.19 percent, 1.5 percent, and 1.7 percent growth in India, China, and Australia over the last five years shows that the South Asian and Oceania area is performing well.

7) Unemployment over the years in different regions of the world:

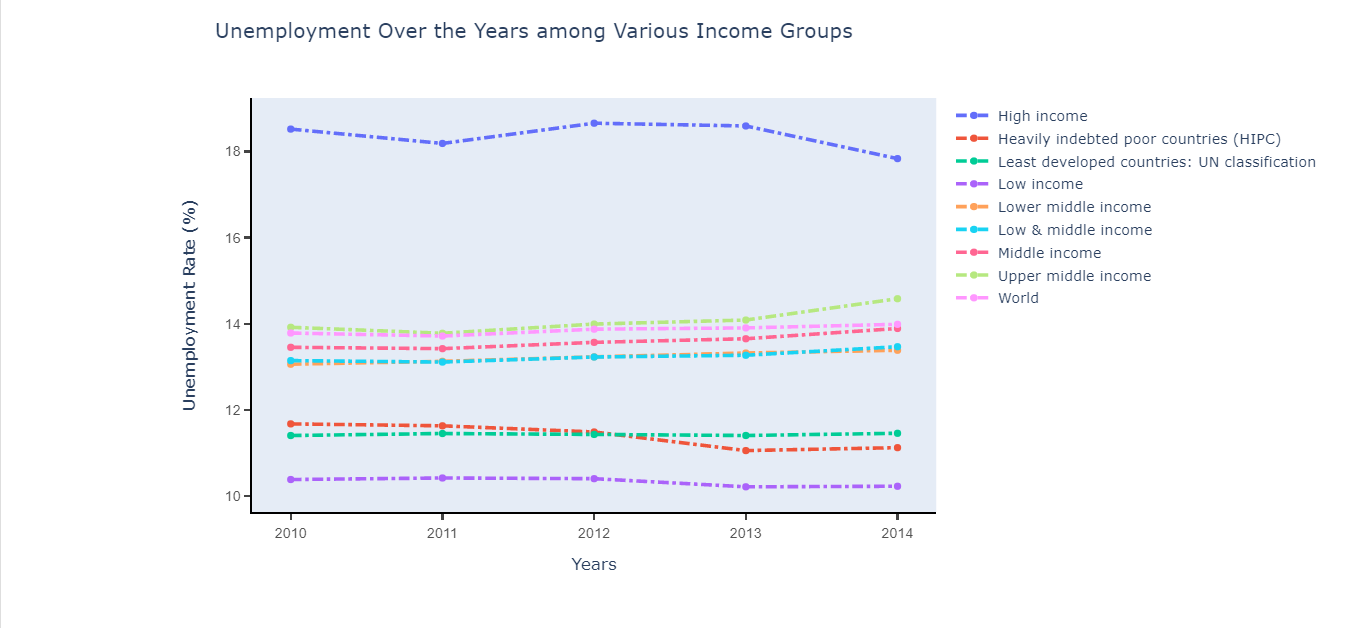




For visualizing the unemployment over the years in different regions of the world we have used the multiple line graph. Multiple line graph is a line graph which shows the multiple variables for comparing trends between two different attributes of a quantity. In this we have taken the countries combined with respect to their region and plotted the unemployment rate of them over the years. As can be seen from the graphs above, Asian countries (East Asia & Pacific, South Asia & Latin America) have lower unemployment rates than the global average. This serves as a basis for comparison. Economic crisis that has been particularly painful for countries in the European Union and the Middle East appears to be strengthening the argument that emerging countries are propelling the global economy ahead rather than their wealthier counterparts.

8) Unemployment over the years among various income groups:





For visualizing the unemployment over the years among the various income groups, we have used the multiple line graph. We have created the data for different income groups among the countries and plotted a multiple line graph using the unemployment rates over the years. Based on the nation's income levels, the world's unemployment rates are compared to those in each country. While less developed and emerging nations are able to provide more job chances and limit unemployment rates from growing, high-income countries suffer from greater unemployment rates. This is evident from the scatter plot.

1. Webpage:

Chart, box and whisker chart

Description automatically generated

We have created html page for displaying the visualizations that we have plotted for the unemployment rate in the world. HTML page is submitted you can access it. You can make interactions with data in the visualization displayed in the html page.

**Story Telling:**

**Chapter 1: Describing real-world problems facing the targeted users or audience.**

**• Who are the people or communities in need of help?**

Unemployment among young people is an ever-increasing monster staring us in the face, and nothing is being done to combat it.

**• What problem happened to them?**

The Unemployment situation has worsened recently as a result of the economic recession, which has resulted in the loss of many employments and the drying up of fresh prospects for those who remain unemployed. In the absence of a strong hand and adequate guidance, our youngsters will face adversities, which will have a negative influence on their future.

**• When did the problem take place?**

During the five years from 2010 to 2014, the Globe Bank gathered unemployment data from across the world and compiled it into this little study.

**• Where means two things: 1) The environment and settings that the people or the community is living in, and 2) the place/location where the problem take place.**

The data includes individual nations as well as nations grouped according to income level and supranational groups such as Arab World, EU etc.

**• Why means the possible causes and/or origin of the problem?**

The economy is not increasing at a fast rate because of a lack of investment and a slow expansion in the manufacturing sector, which produces the most employment.

**Chapter 2: Data.**

Unemployment rate dataset contains the data about the unemployment data in the world of five years from 2010 to 2015. All the countries in the world are sampled in this data set. Data of all the countries from 2010 to 2014 are under sampled. Yes, the data is the representative of the main characters in the assignment 1 in the form of Users. Yes, there is an identifiable information in the data set like the data from 2010 to 2014 of all the countries in the world with their unemployment data rate which is non-disclosed identifiable information. When we combine a box plot and a scatter plot, we can better grasp the data. The data shows a definite upward trend from 2010 to 2014, with a mean of 18.285 percent and a range of 18.2546 percent. statistical displays of data from certain nations, such as box plots, scatter plots, and so on, for a quick glance at how various countries compare in terms of statistical analysis.

Here are some events, activities, behaviors and observations are recorded by the dataset like Country Name, Country code and the years 2010, 2011, 2012, 2013,2014. These observations are taken from the time period of years 2010 to 2014. This is cross-sectional and real time data. Several countries with high unemployment rates in 2010, such as South Africa and Bosnia and Herzegovina, are still among the countries with the highest unemployment rates in 2014. On the other hand, the situation in numerous European countries has deteriorated over time. Between 2010 and 2014, a bar graph depicts the top 15 countries in terms of how successfully they dealt with unemployment. Except for Estonia, which had a negative growth rate of -16.29 percent while having a 17 percent unemployment rate, which is still higher than most other countries in the area, most of these countries are in Europe. The top five are Estonia, Latvia, Lithuania, Moldova, and Ghana.

Over the five-year period from 2010 to 2014, most countries surveyed saw either a modest positive rise or a net decrease in unemployment rates. Except for a few North African and Southern European nations that have seen major increases in unemployment, the rest of the world appears to be doing rather well. Over the previous five years, India, China, and Australia have grown at rates of 0.19 percent, 1.5 percent, and 1.7 percent, respectively, demonstrating that the South Asian and Oceania region is doing well.

The unemployment rate in Asia (East Asia & Pacific, South Asia & Latin America) is lower than the world average. This may be used to make comparisons. The economic crisis, which has been particularly painful for nations in the European Union and the Middle East, appears to be bolstering the claim that developing economies, rather than their wealthier counterparts, are driving the global economy forward.

The world's unemployment rates are compared to those in each country based on income levels. While less developed and rising countries may provide more work opportunities while keeping unemployment rates low, high-income countries face higher unemployment rates.

**Chapter 3: “Users” is a story about the user using the application/visualization.**

**• Who: the main character is the targeted user or audience?**

Targeted user is the main character. The analysis showing clear understanding and insights into the unemployment across the world.

**• What can the application do? What does the visualization show?**

An exploratory analysis of the data by producing numerous simple statistical plots such as box plots, scatter plots etc. of the data from specific nations and taking a quick glimpse at how the various countries compare against one another.

**• When can the user use the application/visualization?**

User can use this visualization whenever user wants to understand the problem of unemployment and how to solve the problem. This visualization will help to get more insights into the data.

**• Where will the visualization and applications be deployed, for example, mobile phones, the web etc.?**

The application can be deployed as web application so that the dynamic graphs will give good insights on the data. It will give the access to try different types of analysis.

**• Why is the visualization or application useful to the user?**

Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information.

**• How will the people/the community use this application or visualization to make changes?**

Users can use this application for understanding the problem of unemployment in the world. They will get to know about the various scenarios of the unemployment in the world like what is the unemployment rate in the world among the years 2010 to 2014, Which countries have the most unemployment rate, how it is increasing or decreasing.

**Project Management:**

Implementation status report:

**Work completed:**

**Description:**

Our aim is to visualize the unemployment data analysis in the world and we have decided to visualize the different categories of analysis of the data like unemployment of all the individual countries for the 5-year duration, comparison of the countries having unemployment rate, comparison of best and worst performers in controlling the unemployment in all the countries, unemployment growth around the globe in those 5 years, unemployment rate in the different regions of the world and unemployment over the years among various income groups. Then we completed the user interface for the visualization on webpage.

**Responsibility:**

We are visualizing the different types of data analysis of unemployment rate in the world in which we have distributed the work to complete for each among the team. We divided the tasks as following.

* Data extraction for unemployment data analysis
* Data preprocessing
* Visualizing the Unemployment of all countries from 2010 to 2014.
* Visualizing the comparison of the countries having unemployment rate.
* Visualizing the comparison of best and worst performers in controlling the unemployment in all the countries.
* Visualizing the unemployment growth around the globe in those 5 years.
* Visualizing the unemployment rate in the different regions of the world.
* Visualizing the unemployment over the years among various income groups.
* Creating the webpage for visualizations.
* Documentation for Increment 1.
* Documentation for Increment 2.

**Contributions:**

1. Venkata Naga Bitra – (25% of the project)

* Data Preprocessing
* Visualizing the Unemployment of all countries from 2010 to 2014.
* Documentation for increment 1.
* Documentation of Final report and Presentation video.

2. Prasanth Damarla - (25% of the project)

* Visualizing the comparison of the countries having unemployment rate.
* Visualizing the comparison of best and worst performers in controlling the unemployment in all the countries.
* Documentation of Final report and Presentation video.

3. Sravanthi Kande - (25% of the project)

* Visualizing the unemployment growth around the globe in those 5 years.
* Visualizing the unemployment rate in the different regions of the world.
* Creating the webpage for visualizations in html.

4. Yamuna Bollepalli - (25% of the project)

* Visualizing the unemployment rate in the different regions of the world.
* Visualizing the unemployment over the years among various income groups.
* Creating the webpage for visualizations in html.

**GitHub Link:**

<https://github.com/damarlaprasanth/UnEmployement-Data-Analysis>

**References:**

* <https://github.com/SFUStatgen/SFUStat341/blob/master/Exercises2018/API_ILO_country_YU.csv>
* <https://en.wikipedia.org/wiki/Unemployment>
* <https://github.com/SFUStatgen/SFUStat341/blob/master/Exercises2018/API_ILO_country_YU.csv>