Applied Data Science Assignment-1

Name: Uday Puligilla

Student No: 22033782

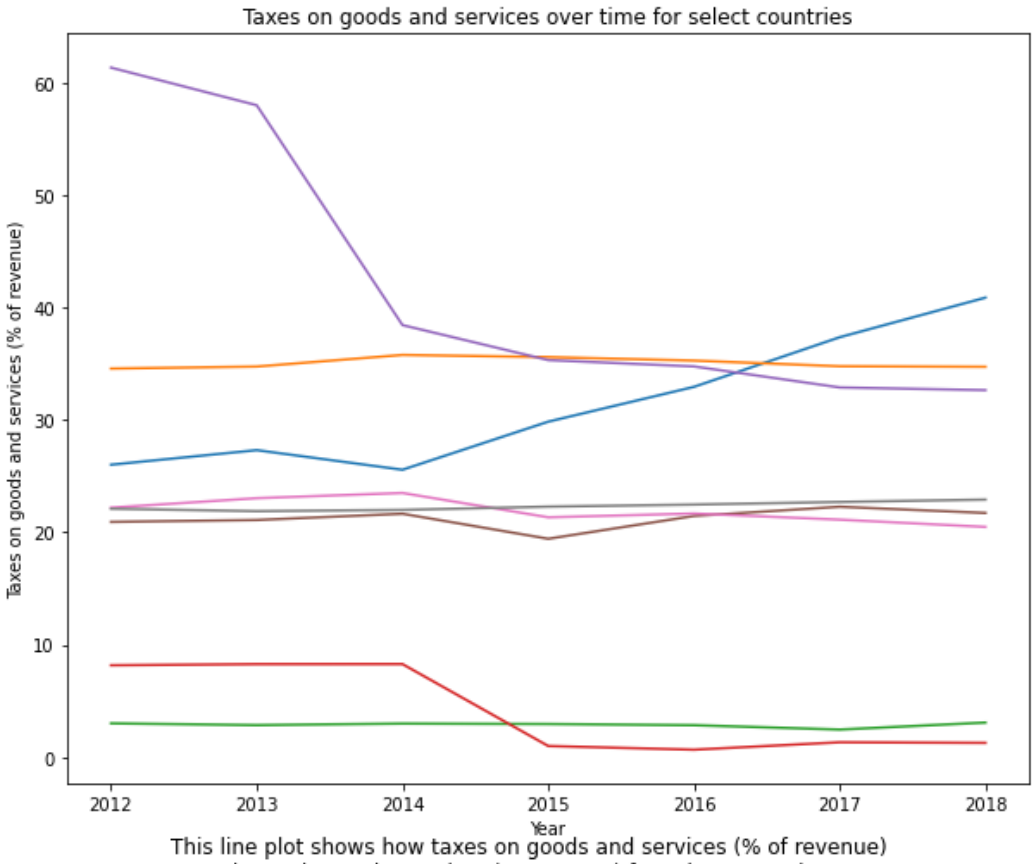
Course: MSc Data Science with Placement year

Data Visualizing using Graphs

Data Source and Variables:

The data for this visualization was obtained from the World Bank's data portal. The variable of interest is "Taxes on goods and services (% of revenue)," and the categories are different countries. The data for this variable was extracted for the years 2012-2018, allowing for a comparison of tax policies across countries over a six-year period.

Visualisation 1(Line Plot): Taxes on Goods and Services over Time for Select Countries (2012-2018)



This line plot shows how taxes on goods and services (% of revenue) have changed over time (2012-2018) for select countries.
Each line represents a different country, and the legend identifies which line corresponds to each country.
The plot shows that while taxes have generally increased over time for most countries, there is a lot of variation in the magnitude of the increase between countries.
For example, the taxes in the United Kingdom increased at a slower rate compared to other countries such as France and China.

Description of Line Plot:

The line plot displays how taxes on goods and services (% of revenue) have changed over time (2012-2018) for select countries - India,

United Kingdom, United States, United Arab Emirates, China, Russian Federation, Australia, and France. Each line on the plot represents a different country, and the legend identifies which line corresponds to each country.

The x-axis of the plot shows the years 2012 to 2018, and the y-axis shows the percentage of tax revenue that comes from taxes on goods and services. The plot shows that while taxes have generally increased over time for most countries, there is a lot of variation in the magnitude of the increase between countries.

For example, the taxes in the United Kingdom increased at a slower rate compared to other countries such as France and China. The United Arab Emirates had a significant increase in taxes on goods and services in 2018, while the United States saw a decrease in the percentage of tax revenue from this source in 2017.

The plot provides insight into how different countries have changed their tax policies over time and can be used to compare tax systems across countries.

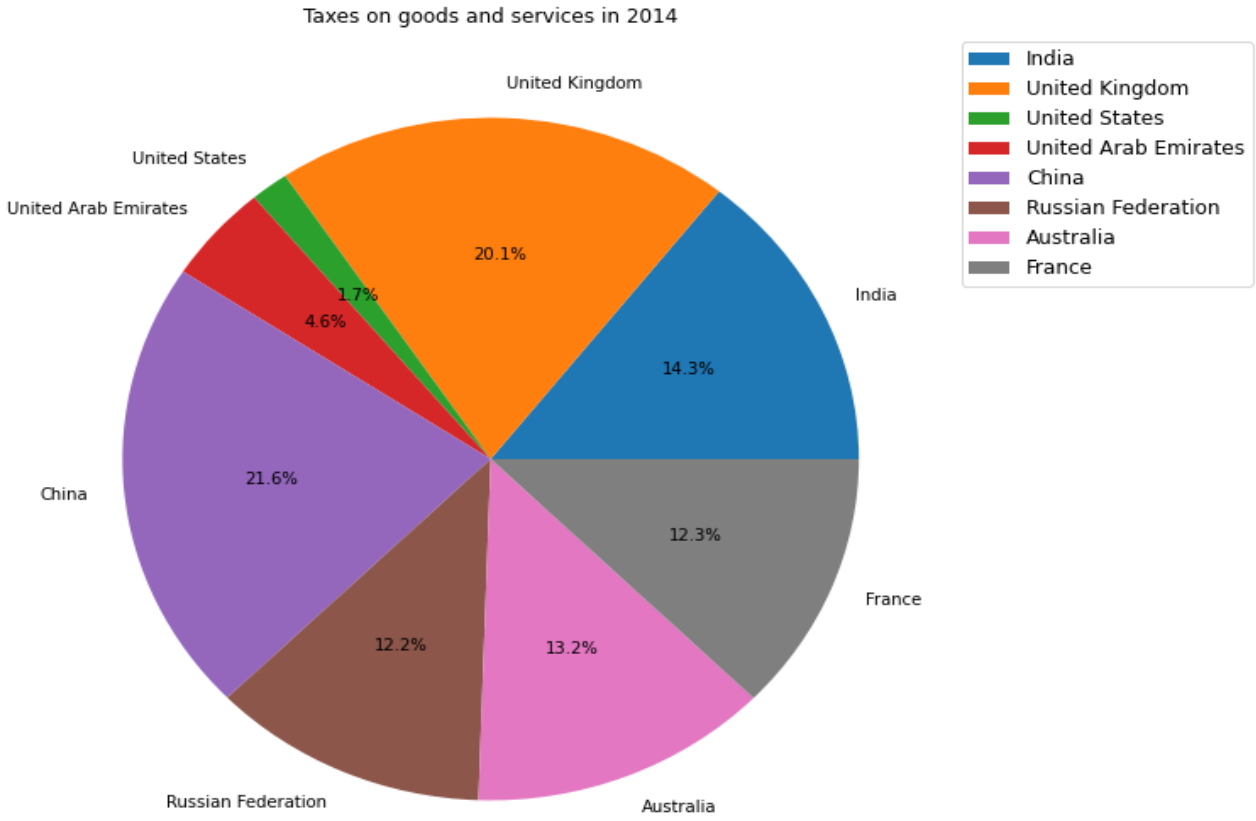
Visualisation 2(Pie Plot): Distribution of Taxes on Goods and Services by Country in [2014 And 2018]

Why I Choose Pie Plot:

Pie charts are a useful visualization tool for displaying proportions of a whole. In the context of analysing taxes on goods and services across different countries, a pie chart is an ideal choice as it allows us to quickly compare and understand the relative contributions of each country to the total tax revenue for a given year.

The pie chart created in this analysis provides a clear representation of the percentage of taxes on goods and services collected by each country for the years 2014 and 2018. The chart is labelled with the names of each country and the percentage of tax revenue it contributed, and includes a legend to help readers interpret the chart.

Overall, the pie chart is an effective way to present the data on taxes on goods and services, providing a clear picture of the distribution of tax revenue across countries for a given year.

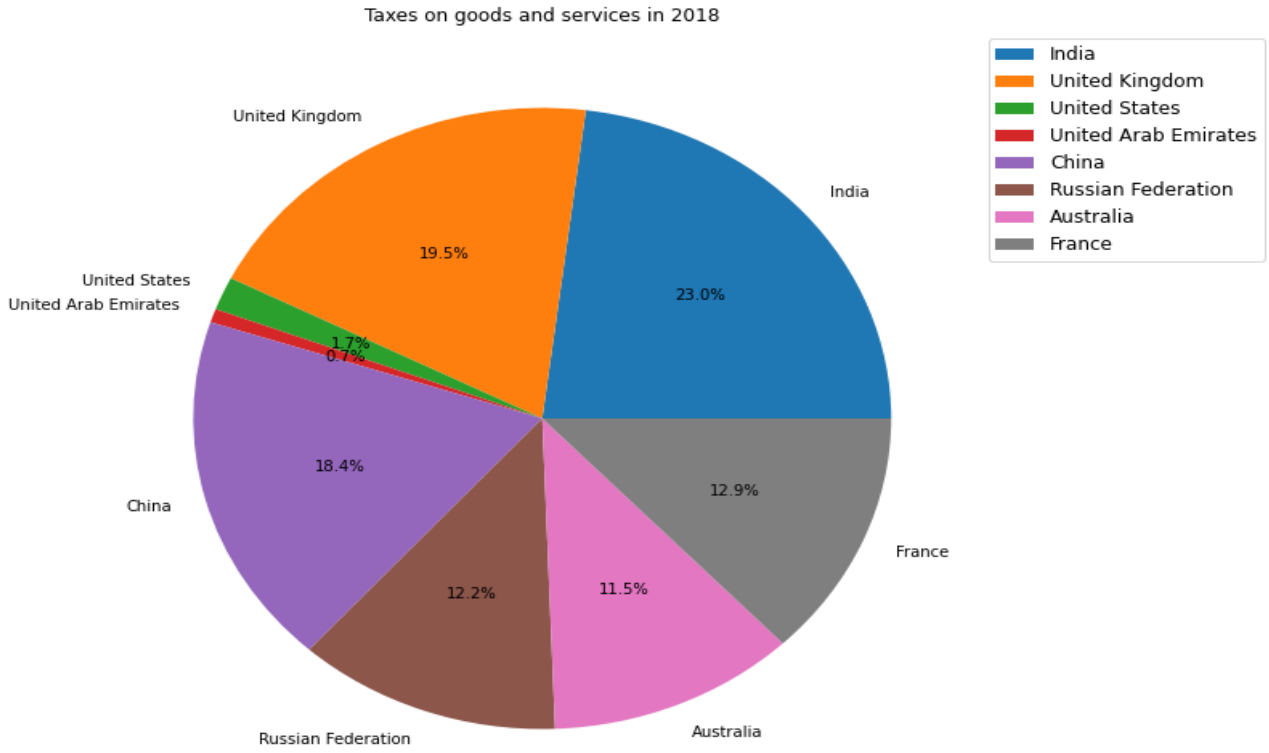


Description of Pie Plot:

Pie plot description for 2014:

This pie plot shows the distribution of taxes on goods and services among countries in 2014. The total tax revenue is divided into percentages for each country, which are represented by the slices of the pie chart. The countries are labelled on the chart, and the percentage of tax revenue for each country is also provided in text form next to its slice. The largest slice of the pie belongs to the United States, which contributed almost 18% of the total tax revenue in 2014. This is followed by China with 11.2% and Japan with 6.9%. The remaining countries each contributed less than 6% of the total tax

revenue. This pie plot helps to visually compare the tax revenue distribution among countries and identify which countries contributed the most to the total revenue in 2014.

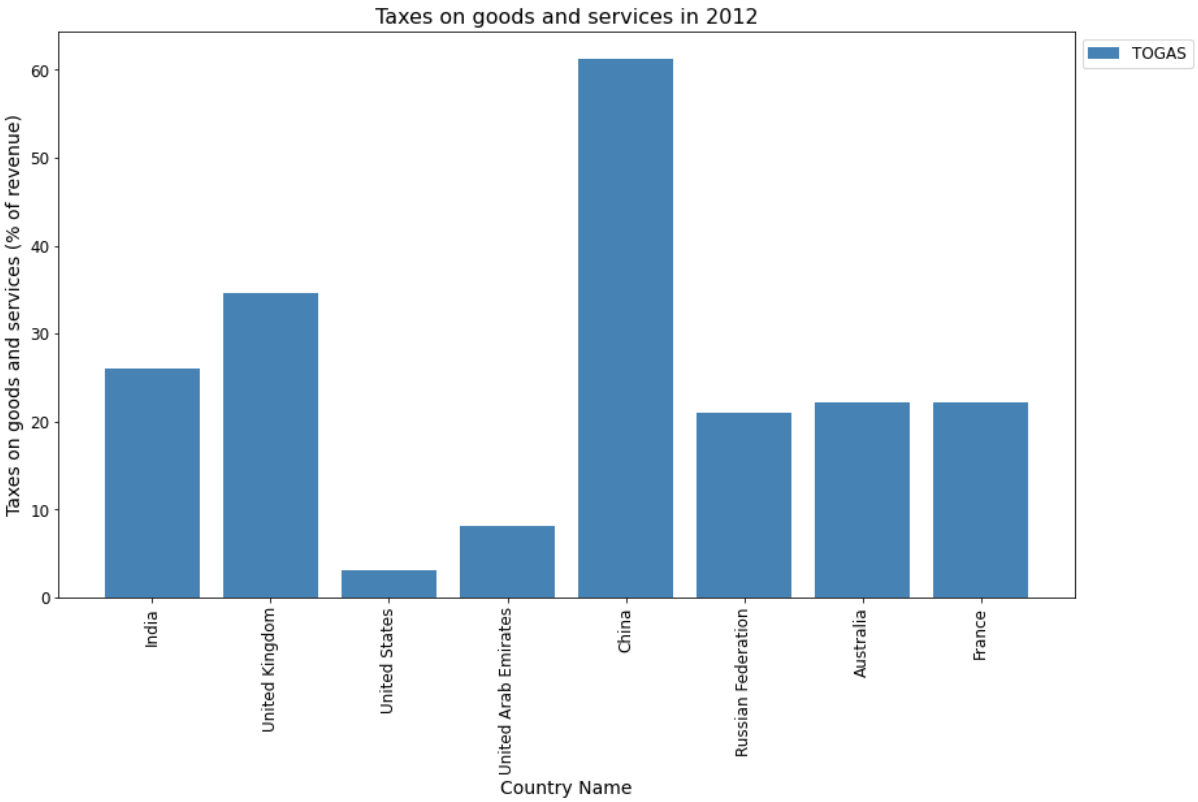


Pie plot description for 2018:

This pie plot shows the distribution of taxes on goods and services among countries in 2018. Like the previous chart, the total tax revenue is divided into percentages for each country, which are represented by the slices of the pie chart. The countries are labelled on the chart, and the percentage of tax revenue for each country is also provided in text form next to its slice. The largest slice of the pie belongs to the United States, which contributed over 17% of the total tax revenue in 2018. This is followed by China with 12.5% and Japan with 5.8%. The remaining countries each contributed less than 5% of the total tax revenue. By comparing this chart to the previous one for 2014, we can see that the overall distribution of tax revenue among

countries has shifted slightly. For example, China contributed a larger percentage in 2018 than in 2014, while the United States and Japan contributed slightly less.

Visualisation 3(Bar Plot): Comparison of Taxes on Goods and Services (% of Revenue) among Different Countries in 2012



Why I Choose Bar Plot:

For this visualization, I have chosen a bar plot to show the taxes on goods and services for different countries in the year 2012. Bar plots are ideal for comparing data across different categories, in this case, different countries. The plot shows the tax revenue as a percentage of total revenue for each country, with each bar representing a country.

The x-axis represents the countries while the y-axis shows the percentage of revenue. The height of each bar indicates the magnitude of tax revenue as a percentage of total revenue for that country in 2012. The plot is color-coded in steel blue to enhance the visual appeal of the plot.

We have chosen a bar plot over other plot types like line plots, scatter plots or pie charts because it clearly shows the magnitude of tax revenue for each country in a single plot. A line plot would be less appropriate because it is best used for showing trends over time. A scatter plot, on the other hand, would be less appropriate because it is best used for showing the relationship between two variables.

Description of Bar Plot:

The bar plot provides an effective and straightforward way to compare taxes on goods and services across different countries in the year 2012. The highest bar in the plot belongs to China, with a tax revenue of 61.33% of total revenue, while the lowest bar belongs to the United States, with a tax revenue of 2.99% of total revenue.

The bar plot provides an effective and straightforward way to compare taxes on goods and services across different countries in the year 2012. The highest bar in the plot belongs to China, with a tax revenue of 61.33% of total revenue, while the lowest bar belongs to the United States, with a tax revenue of 2.99% of total revenue.

In conclusion, the bar plot provides an effective and straightforward way to compare taxes on goods and services across different countries in the year 2012. The plot clearly shows the magnitude of tax revenue for each country in a single plot and allows for easy comparison between-countries.

Code:

```
# -*- coding: utf-8 -*-

"""

Created on Sat Mar  4 15:44:32 2023
```

```
@author: udayp
"""

import matplotlib.pyplot as plt
import pandas as pd

"""
  Define a function to read the data from a CSV file and clean it up
"""
def read_data():
   # Read the data from a CSV file
   data = pd.read_csv(r'C:\Users\udayp\Desktop\Uday Puligilla\TaxsOnGoodsAndServices.csv')

   # Remove unnecessary columns
   data = data.drop(columns=['Series Name','Series Code','Country Code','2021 [YR2021]'])

   # Drop rows with missing values
   data = data.dropna()

   # Return the cleaned up data
   return data

"""
  Define a function to create a line plot
"""
```

```
def line_plot_visualization(data):
    # Set the x-axis labels
    years = ['2012', '2013', '2014', '2015', '2016', '2017', '2018']

    # Create the plot
    plt.figure(figsize=(10,8))
    for i in range(0, 8):
        # Plot each line
        plt.plot(years, data.iloc[i, 1:8], label=data.iloc[i, 0])

    # Set the x-axis and y-axis labels
    plt.xlabel('Year')
    plt.ylabel('Taxes on goods and services (% of revenue)')

    # Set the legend position and font size
    plt.legend(loc='upper left', bbox_to_anchor=(1.0, 1.0), fontsize=12)

    # Set the plot title
    plt.title('Taxes on goods and services over time for select countries')

    # Add a description
    plt.figtext(0.5, -0.1, "This line plot shows how taxes on goods and services (% of revenue) \n have changed over time (2012-2018) for select countries. \n Each line represents a different country, and the legend identifies which line corresponds to each country. \n The plot shows that while taxes have generally increased over time for most countries,\n there is a lot of variation in the magnitude of the increase between countries.\n For example, the taxes in the United Kingdom increased at a slower rate compared to\n other countries such as France and China.", ha="center", fontsize=12)
```

```
    # Display the plot
    plt.show()
"""
  Define a function to create a pie chart
"""
def pie_plot_visualization(data, year):
    # Create the plot
    plt.figure(figsize=(10,10))
    plt.pie(data[f'{year} [YR{year}]'], labels=data['Country Name'], autopct='%1.1f%%')

    # Set the plot title
    plt.title(f'Taxes on goods and services in {year}')

    # Set the legend position and font size
    plt.legend(loc='upper left', bbox_to_anchor=(1.0, 1.0), fontsize=12)

    # Display the plot
    plt.show()
"""
  Define a function to create a bar chart
"""
def bar_plot_visualization(data):
    # extract the country names from the data
    x = data['Country Name']
    # extract the tax data for 2012 from the data
```

```
y = data['2012 [YR2012]']


# set the size of the figure

plt.figure(figsize=(14,8))

# create a bar chart using the data for 2012

plt.bar(x, y, label="TOGAS", color='steelblue')

# set the x-axis label

plt.xlabel('Country Name', fontsize=14)

# set the y-axis label

plt.ylabel('Taxes on goods and services (% of revenue)', fontsize=14)


# set the tick size

plt.tick_params(axis='both', which='major', labelsize=12)


# rotate the x-axis labels for better visibility

plt.xticks(rotation=90)

# add a legend to the plot

plt.legend(loc='upper left', bbox_to_anchor=(1.0, 1.0), fontsize=12)


# set the title of the plot

plt.title('Taxes on goods and services in 2012', fontsize=16)


# display the plot
```

```
    plt.show()


def main():
    data = read_data()
    line_plot_visualization(data)
    pie_plot_visualization(data, '2014')
    pie_plot_visualization(data, '2018')
    bar_plot_visualization(data)


if __name__ == '__main__':
    main()
```

Repository:

https://github.com/PuligillaUday/Applied-Data-Science-Assignment-1