# Title

# Comprehensive Analysis of Global Health and Mental Health Trends: Insights from Data Visualization and Statistical Evaluation

Table of Contents

[Title 1](#_Toc174134959)

[Comprehensive Analysis of Global Health and Mental Health Trends: Insights from Data Visualization and Statistical Evaluation 1](#_Toc174134960)

[DataSet 1 Health\_systems 2](#_Toc174134961)

[**1. Load the Dataset** 2](#_Toc174134962)

[**2. Display the First Few Rows of the Dataset** 3](#_Toc174134963)

[**3. Display Basic Information About the Dataset** 3](#_Toc174134964)

[**4. Display Summary Statistics** 3](#_Toc174134965)

[**5. Check for Missing Values** 3](#_Toc174134966)

[**6. Drop Rows with Missing Values (Optional)** 4](#_Toc174134967)

[1. Bar Chart of Average Health Expenditure as Percentage of GDP 4](#_Toc174134968)

[Scatter Plot of Health Expenditure per Capita vs. Physicians per 1000 People 5](#_Toc174134969)

[Histogram of Health Expenditure per Capita Distribution 6](#_Toc174134970)

[Pie Chart of Health Expenditure by Source 7](#_Toc174134971)

[Data Set 2 Mental health Depression disorder Data 7](#_Toc174134972)

[1. Load the Dataset 7](#_Toc174134973)

[2. Display the First Few Rows of the DataFrame 7](#_Toc174134974)

[3. Display Basic Information About the Dataset 8](#_Toc174134975)

[4. Display Summary Statistics of the Dataset 8](#_Toc174134976)

[Schizophrenia Prevalence Over Time for Selected Countries 8](#_Toc174134977)

[Top 20 Countries by Depression Prevalence in 2017 9](#_Toc174134978)

[3. Correlation Matrix of Mental Disorders 10](#_Toc174134979)

[Average Prevalence of Mental Disorders Over Time 11](#_Toc174134980)

[Distribution of Depression Prevalence Across Countries 12](#_Toc174134981)

[Dataset 3 Mental Health Dataset 12](#_Toc174134982)

[1. Dataset Overview 12](#_Toc174134983)

[2. Basic Information 12](#_Toc174134984)

[3. Summary Statistics 13](#_Toc174134985)

[Distribution of Responses by Gender 13](#_Toc174134986)

[Occupation Distribution of Respondents 14](#_Toc174134987)

[Mental Health History by Self-Employed Status 14](#_Toc174134988)

[Percentage of Treatment Received by Family History of Mental Health Issues 15](#_Toc174134989)

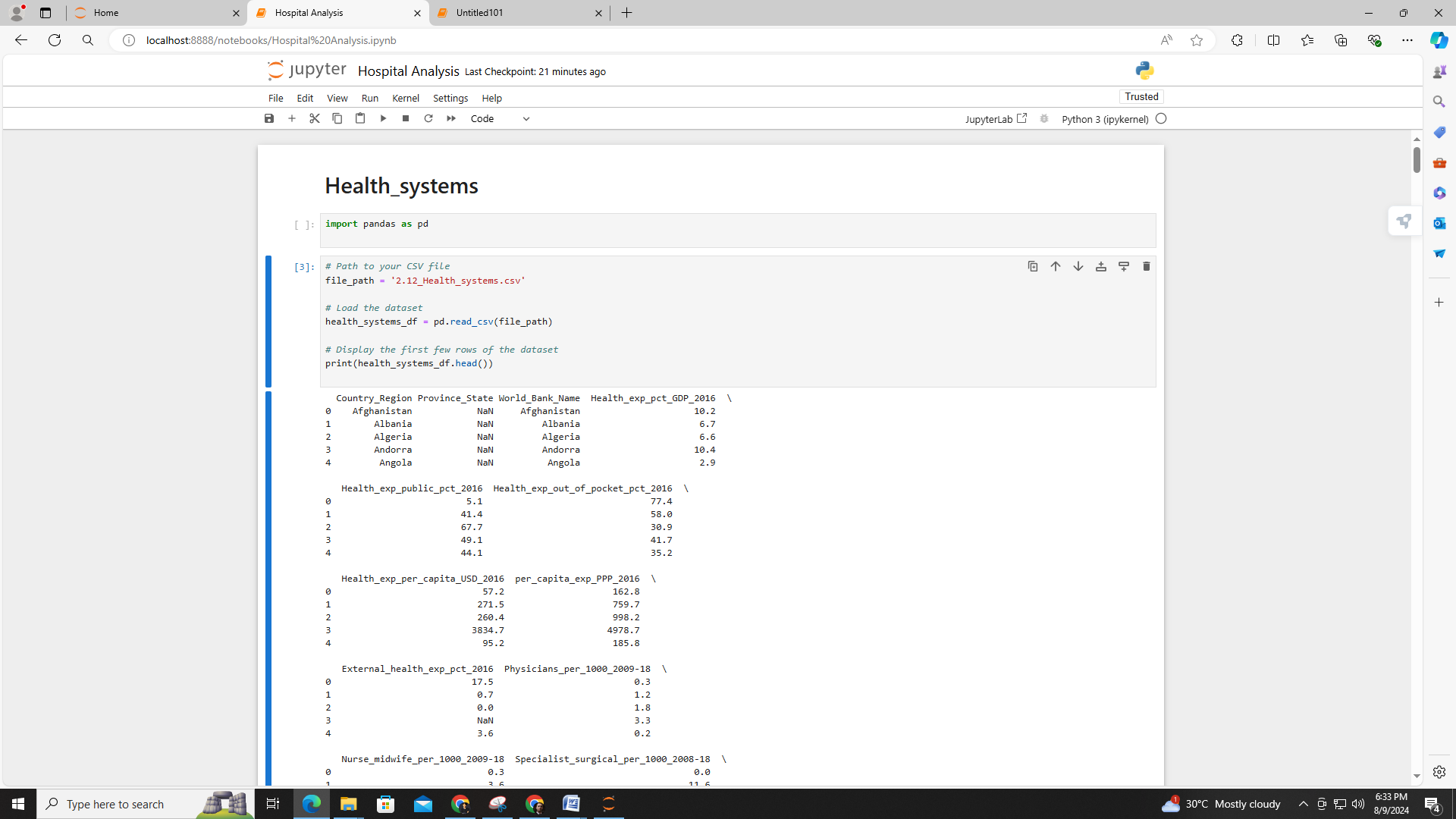
[**Conclusion** 16](#_Toc174134990)

[References 16](#_Toc174134991)

# DataSet 1 Health\_systems

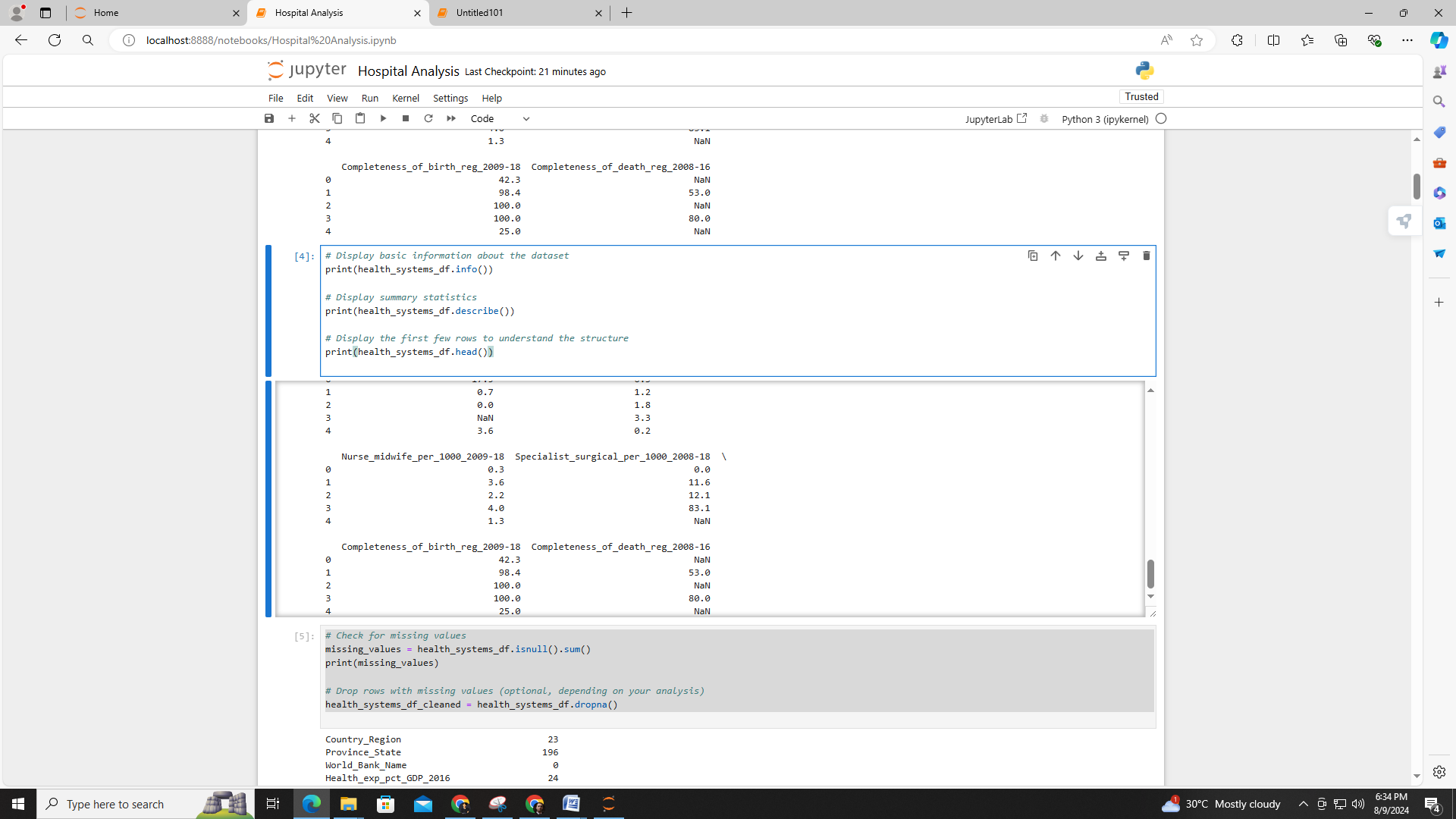
**1. Load the Dataset**

The first step involves loading the dataset from the CSV file using pandas. The dataset, named 2.12\_Health\_systems.csv, is read into a DataFrame called health\_systems\_df. This is accomplished using the pd.read\_csv(file\_path) function, where file\_path is the path to the CSV file. Loading the dataset enables us to begin our analysis and manipulation of the data.



**2. Display the First Few Rows of the Dataset**

After loading the dataset, the print(health\_systems\_df.head()) command is used to display the first few rows of the DataFrame. This step provides a glimpse into the data, allowing us to understand its structure and the types of information it contains. It helps us verify that the dataset has been loaded correctly and to familiarize ourselves with its columns and values.



**3. Display Basic Information About the Dataset**

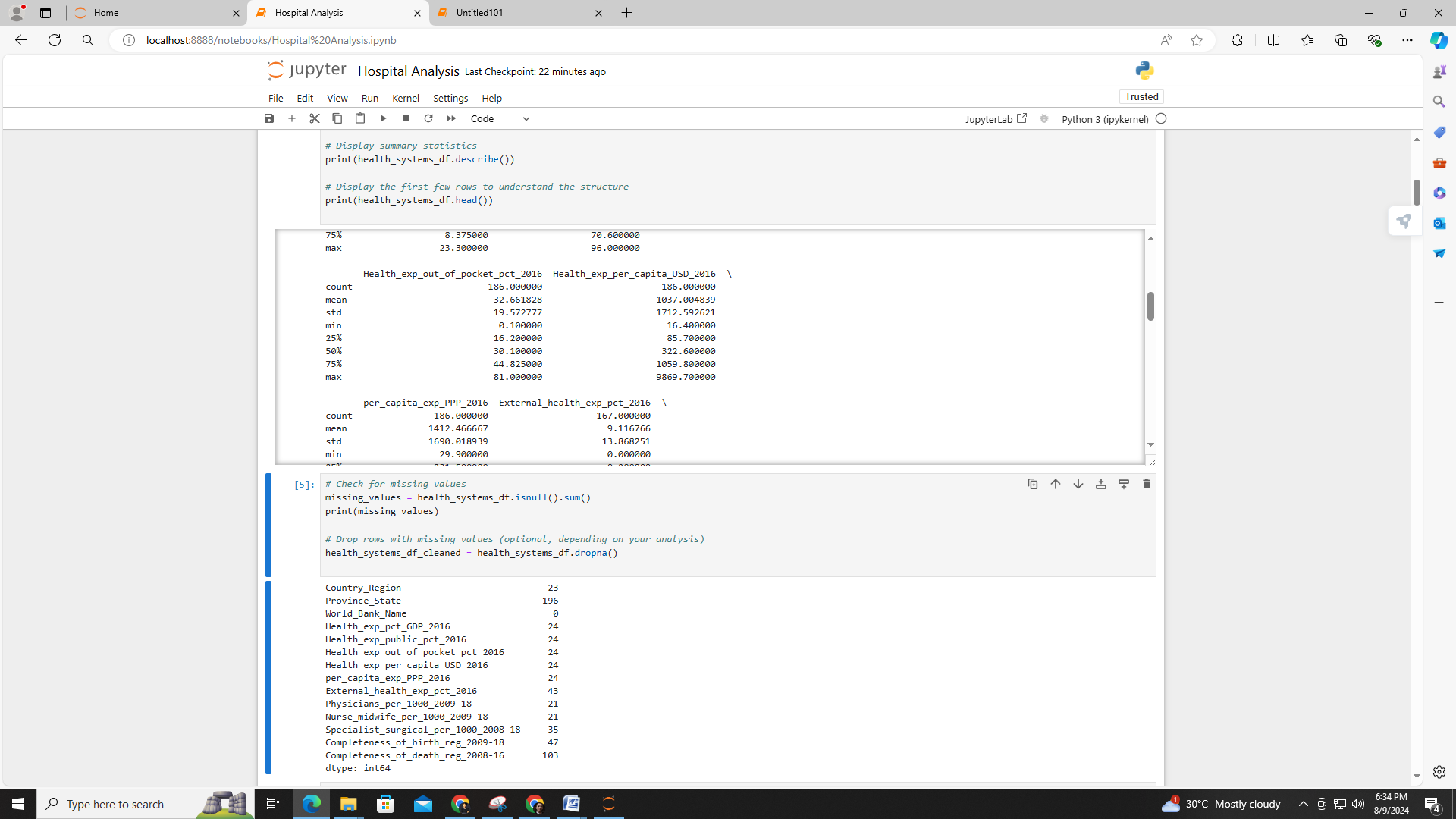
The print(health\_systems\_df.info()) command provides basic information about the dataset. This includes the number of non-null entries in each column, the data types of the columns, and the total number of rows and columns. This information is crucial for understanding the completeness of the data and the types of variables present, which will guide the analysis.

**4. Display Summary Statistics**

The print(health\_systems\_df.describe()) command generates summary statistics for the dataset. This includes measures such as mean, standard deviation, minimum, maximum, and quartiles for numerical columns. These statistics offer a quantitative overview of the data, helping us to identify trends, distributions, and potential outliers in the dataset.

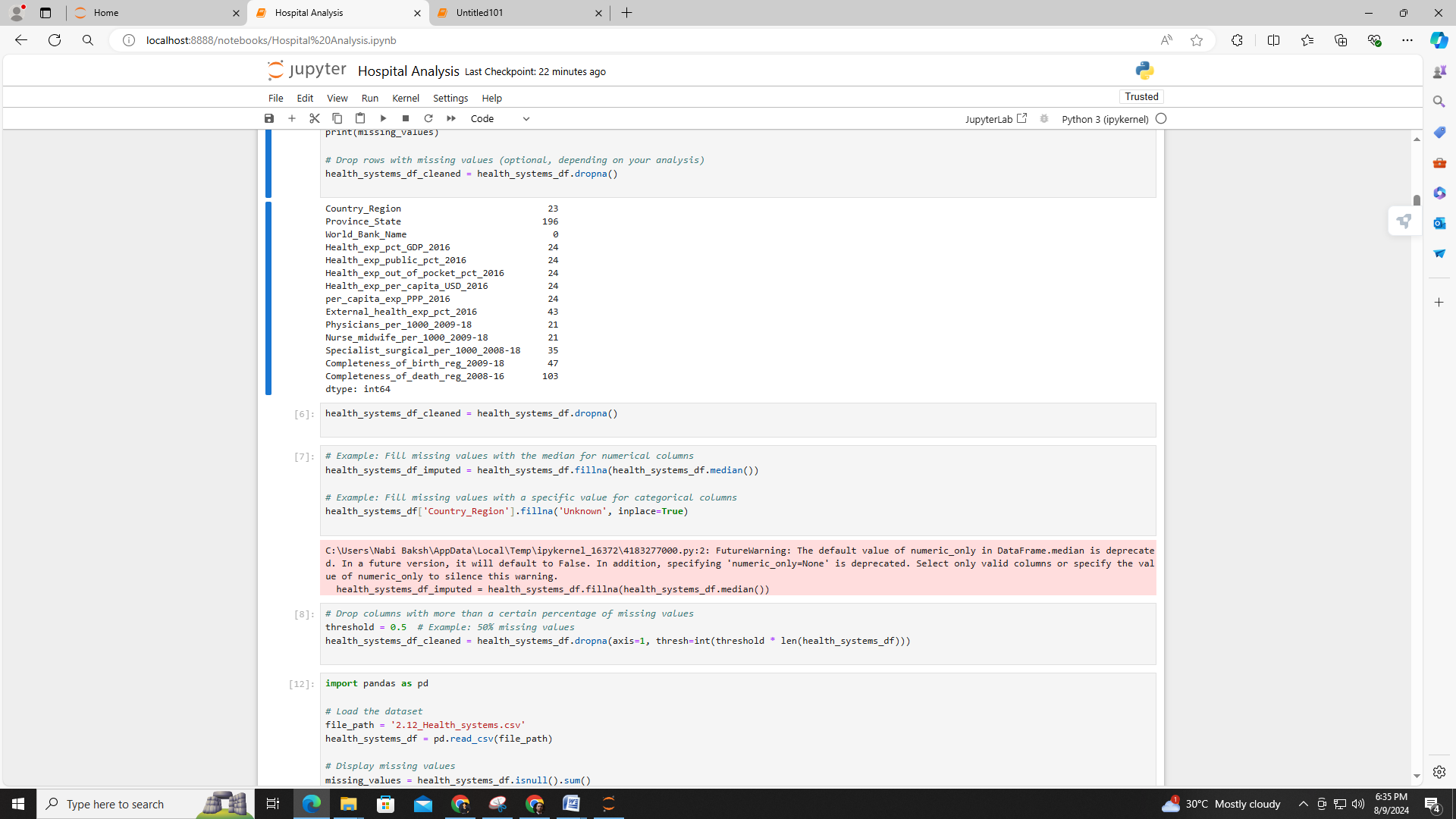
**5. Check for Missing Values**

Using print(health\_systems\_df.isnull().sum()), we check for missing values in the dataset. This step calculates the number of missing entries for each column, providing insight into the extent of data completeness. Identifying missing values is essential for deciding how to handle them, as they can affect the accuracy and reliability of the analysis.

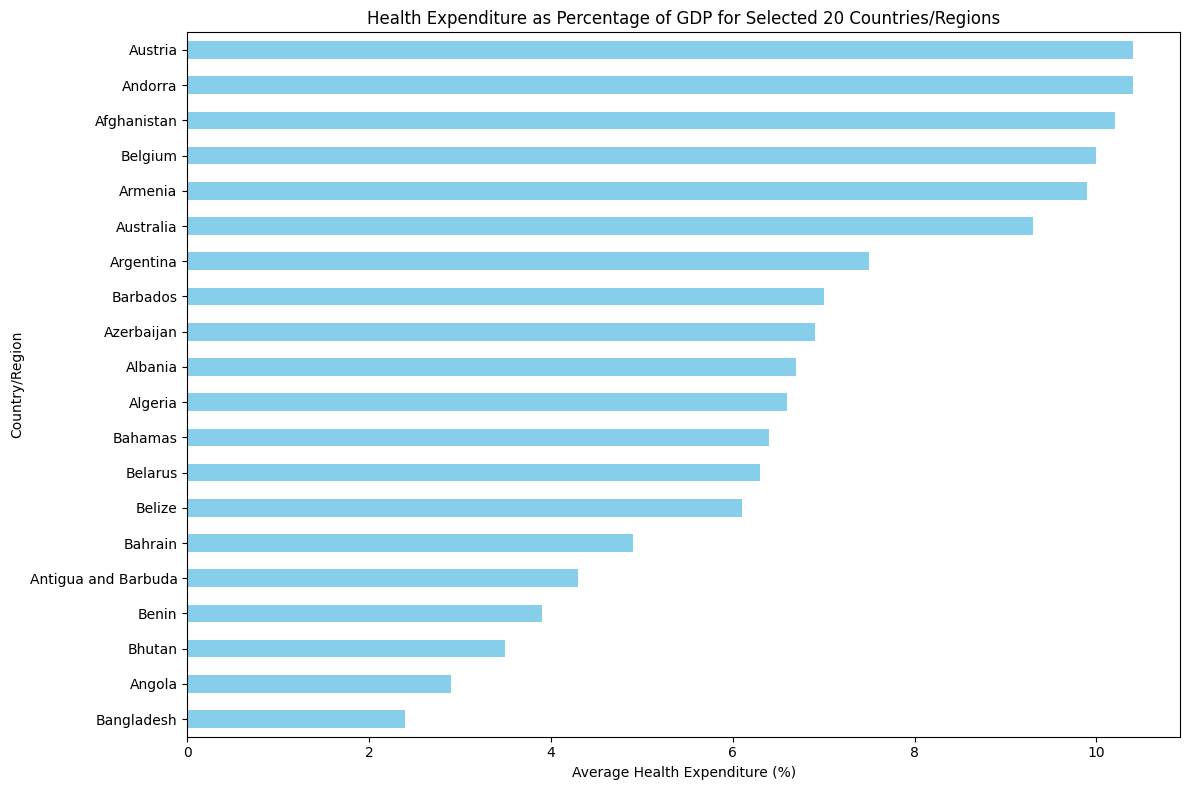


**6. Drop Rows with Missing Values (Optional)**

Depending on the analysis needs, we can use health\_systems\_df\_cleaned = health\_systems\_df.dropna() to remove rows with missing values. This step is optional and should be considered based on the extent of missing data and its impact on the analysis. Dropping missing values ensures that the dataset used for analysis is complete, but it may lead to loss of data if a significant amount is missing.

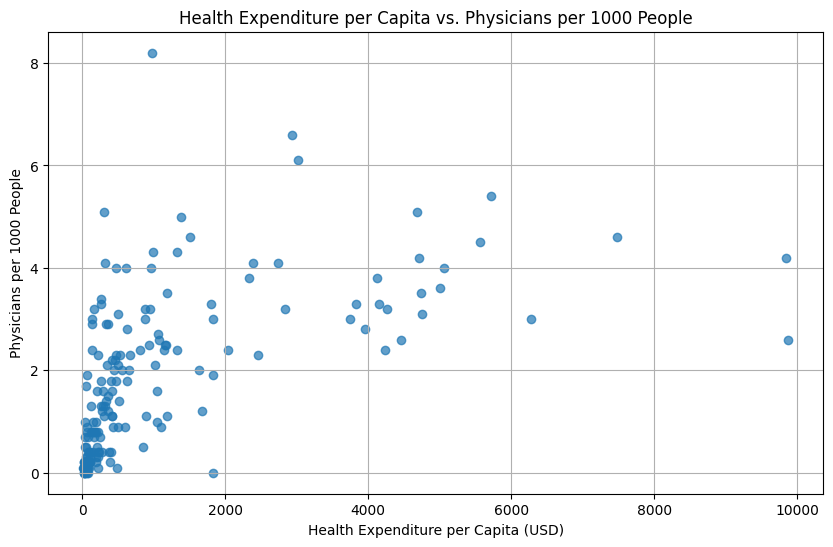


### 1. ****Bar Chart of Average Health Expenditure as Percentage of GDP****



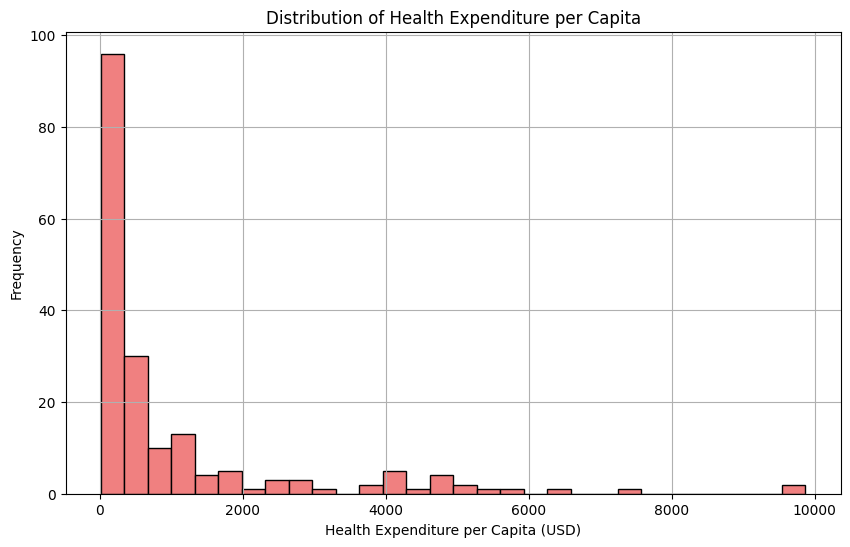
This bar chart visualizes the average health expenditure as a percentage of GDP for the top 20 countries or regions. Each bar represents a country or region, with the length indicating the average percentage of GDP spent on health. The chart highlights differences in health spending priorities among these selected countries, providing a clear comparison of their commitment to health expenditure relative to their GDP.

### ****Scatter Plot of Health Expenditure per Capita vs. Physicians per 1000 People****



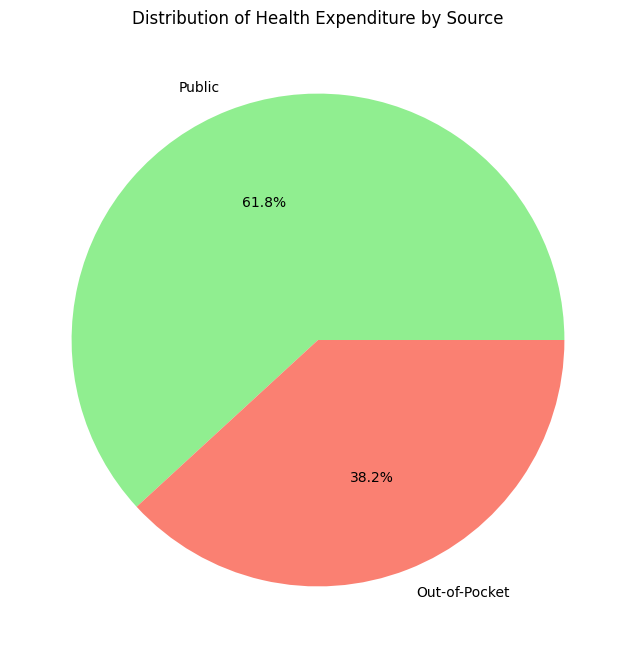
The scatter plot illustrates the relationship between health expenditure per capita and the number of physicians per 1000 people. Each point represents a country, with its position determined by its health expenditure per capita (on the x-axis) and the number of physicians available (on the y-axis). This plot helps identify trends or correlations between higher spending on health and the availability of medical professionals.

# Histogram of Health Expenditure per Capita Distribution



The histogram depicts the distribution of health expenditure per capita across different countries. The x-axis represents the expenditure per capita, while the y-axis shows the frequency of countries falling within each expenditure range. This visualization reveals the spread and concentration of health spending per capita, highlighting how common or rare different expenditure levels are among the countries analyzed.

# Pie Chart of Health Expenditure by Source

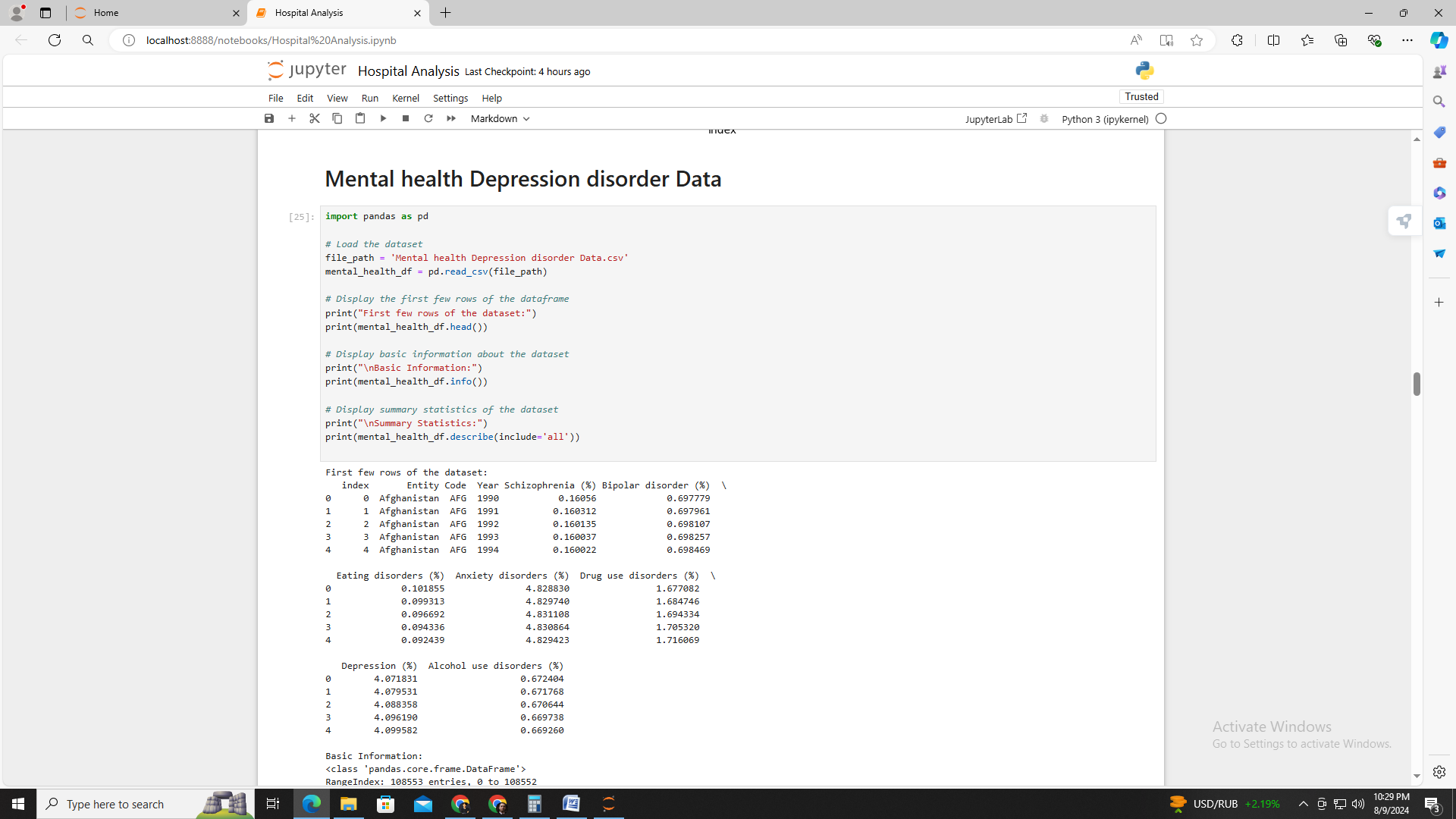


The pie chart breaks down the total health expenditure by source into two categories: public spending and out-of-pocket expenses. The chart shows the percentage contribution of each source to the total health expenditure, providing an overview of how health costs are distributed between public funding and individual payments. This visualization helps understand the balance of health expenditure sources and their relative importance.

# Data Set 2 Mental health Depression disorder Data

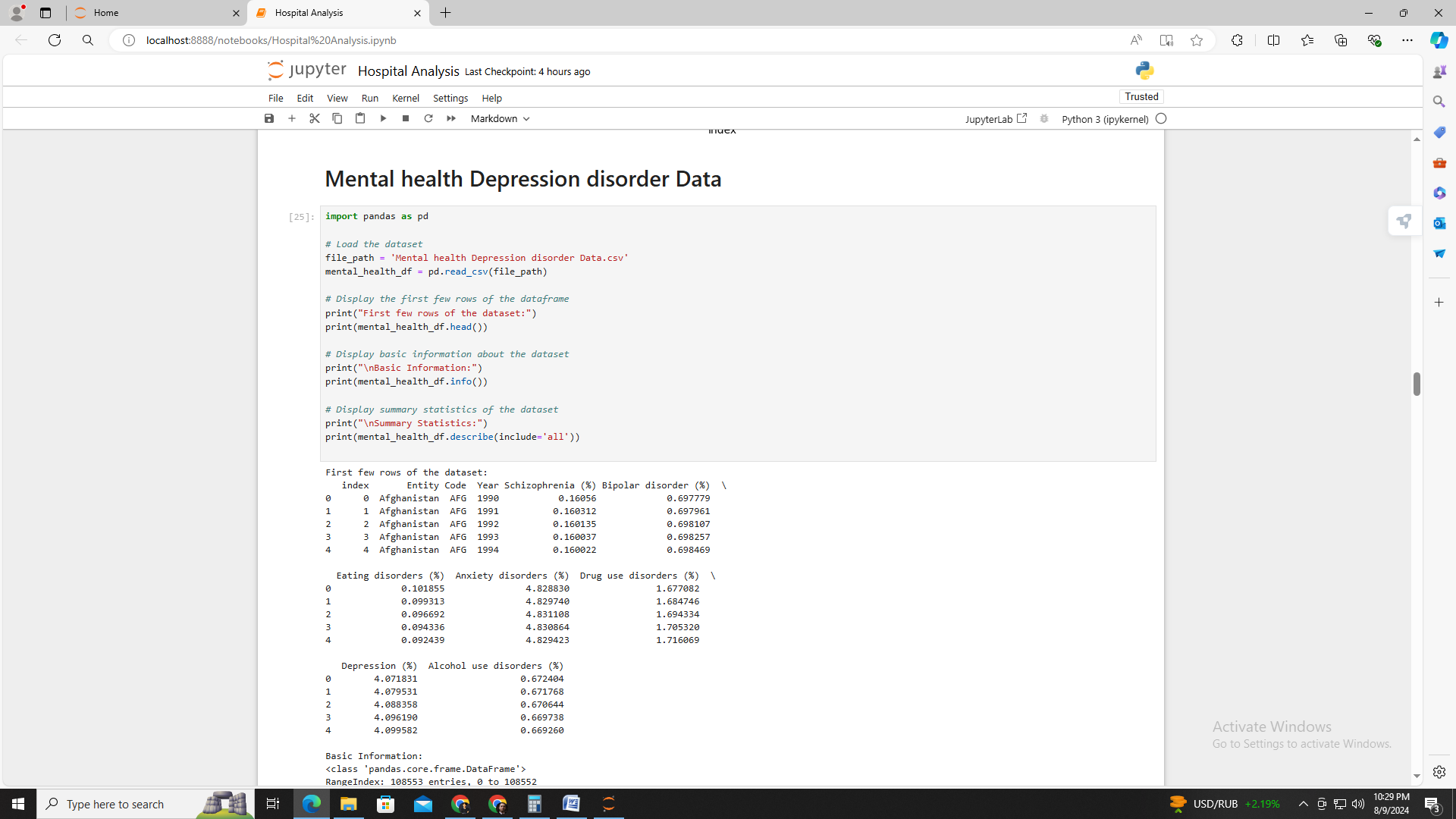
### 1. ****Load the Dataset****

The dataset, Mental health Depression disorder Data.csv, is loaded into a pandas DataFrame named mental\_health\_df. This is done using the pd.read\_csv(file\_path) function, where file\_path specifies the location of the CSV file. Loading the dataset is the initial step in preparing the data for analysis, enabling further exploration and processing.



### 2. ****Display the First Few Rows of the DataFrame****

By using print(mental\_health\_df.head()), we can view the first few rows of the DataFrame. This command helps us get a preliminary look at the dataset, including the columns and the type of data they contain. It also ensures that the dataset has been loaded correctly and allows us to start understanding its structure.

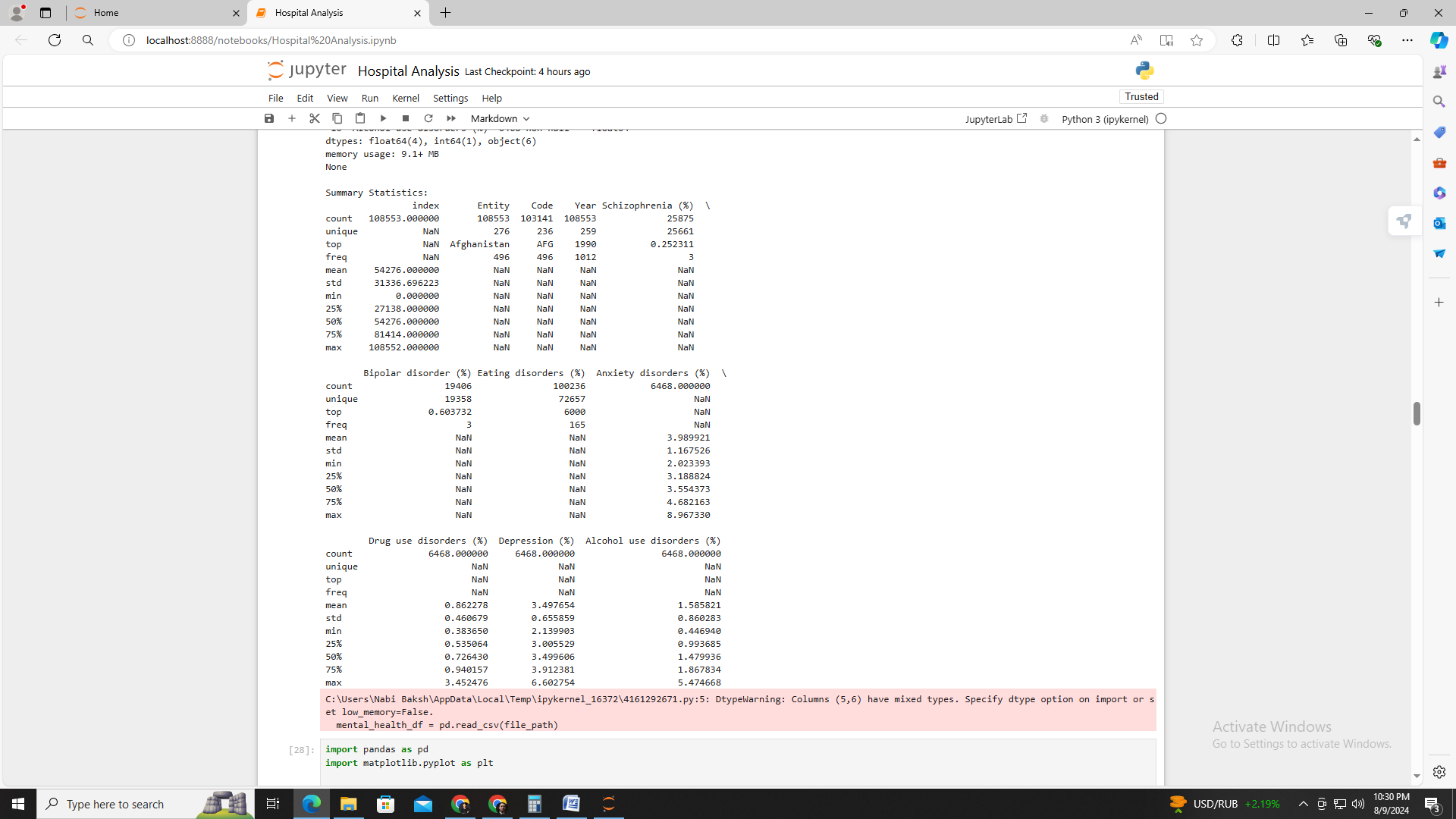


### 3. ****Display Basic Information About the Dataset****

The print(mental\_health\_df.info()) command provides an overview of the DataFrame, including the number of non-null entries, data types of each column, and overall structure. This step is crucial for assessing the completeness of the dataset, as well as identifying the types of data we are working with. It also highlights any issues such as missing values in different columns.

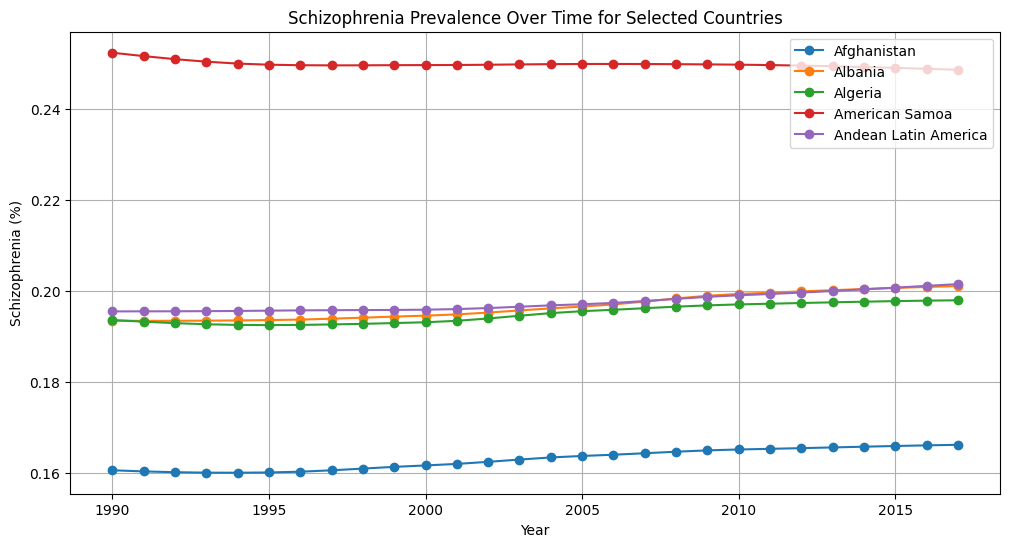
### 4. ****Display Summary Statistics of the Dataset****

Using print(mental\_health\_df.describe(include='all')), we generate summary statistics for the dataset. This command provides insights into the distribution and central tendencies of numerical and categorical columns. Summary statistics include measures like mean, standard deviation, and percentiles for numerical columns, and counts and unique values for categorical columns. This helps in understanding the general characteristics and variations within the dataset.



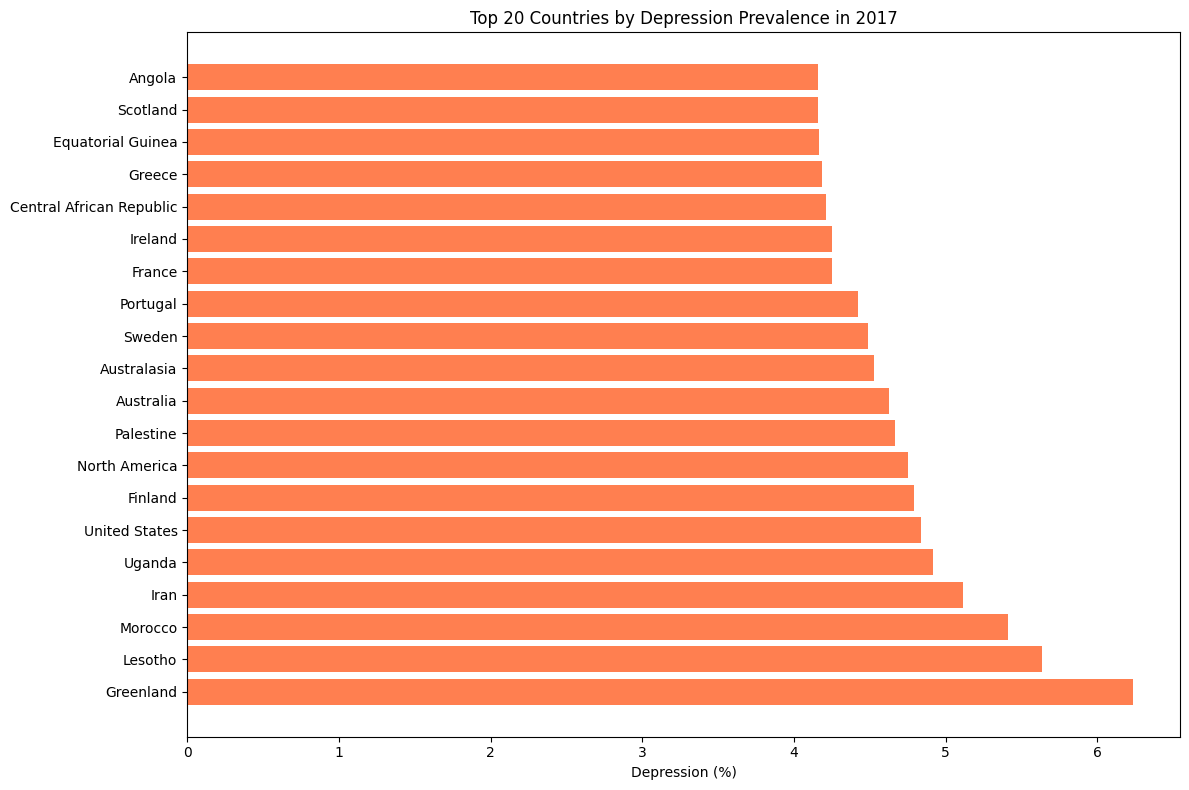
### 

### ****Schizophrenia Prevalence Over Time for Selected Countries****



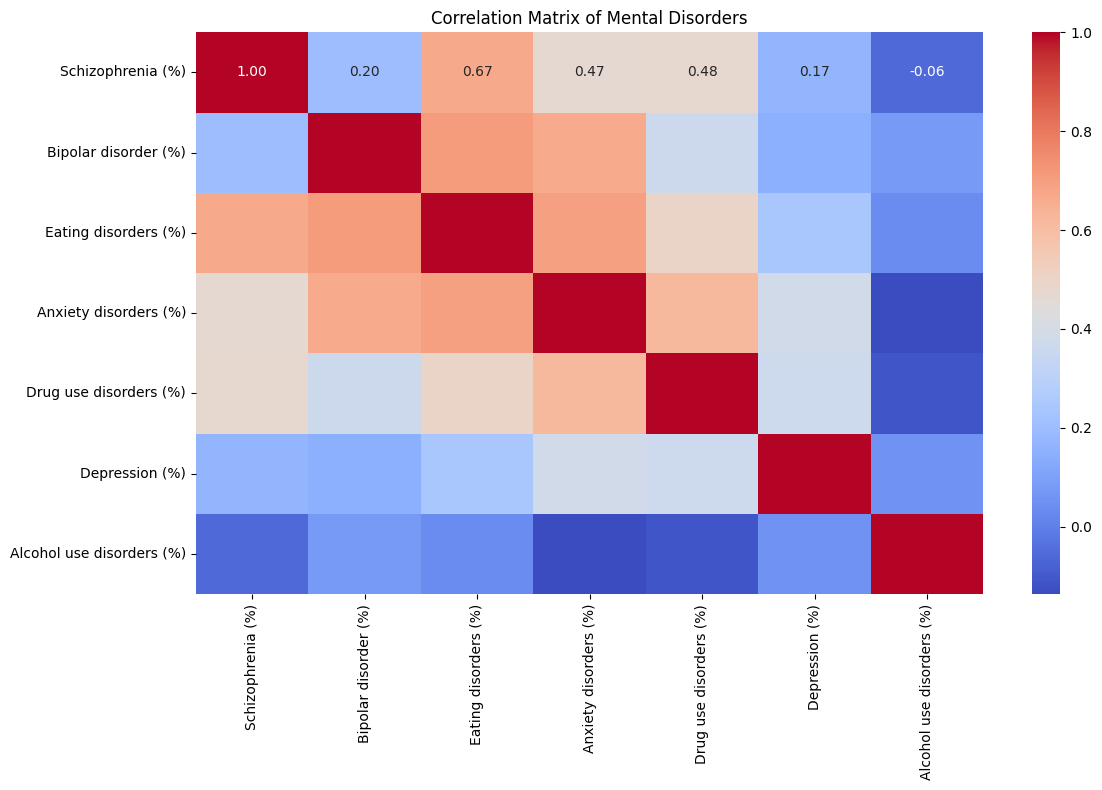
This line graph tracks the prevalence of schizophrenia over the years for the first five countries from the dataset. Each line represents a country, showing how the percentage of schizophrenia cases has changed over time. The graph highlights trends and fluctuations in schizophrenia rates across different countries, providing insight into the long-term patterns of this mental disorder.

# Top 20 Countries by Depression Prevalence in 2017



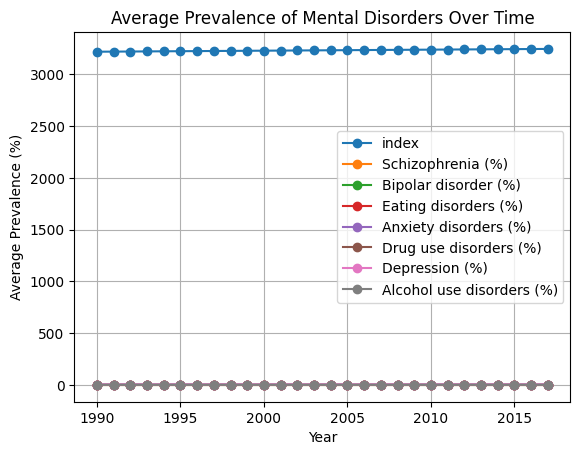
The horizontal bar chart displays the top 20 countries with the highest prevalence of depression in 2017. Each bar represents a country, with the length corresponding to the percentage of the population experiencing depression. This visualization offers a clear comparison of depression rates across different countries for the specified year, emphasizing the countries with the highest reported levels of depression.

# 3. **Correlation Matrix of Mental Disorders**



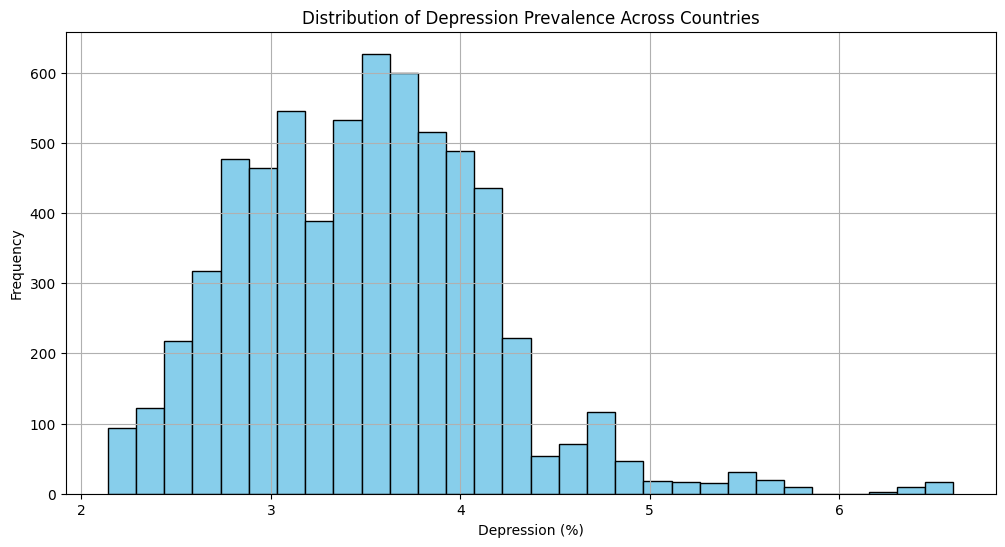
The heatmap illustrates the correlation between various mental disorders, including schizophrenia, bipolar disorder, eating disorders, anxiety disorders, drug use disorders, depression, and alcohol use disorders. Each cell shows the strength of the relationship between pairs of disorders, with color intensity representing the degree of correlation. This visualization helps identify patterns and relationships among different mental health conditions.

# **Average Prevalence of Mental Disorders Over Time**



The line chart shows the average prevalence of several mental disorders over the years. Each line represents a different mental disorder, plotting its average percentage prevalence across all countries for each year. This visualization provides an overview of how the average prevalence of these mental disorders has evolved over time.

# Distribution of Depression Prevalence Across Countries

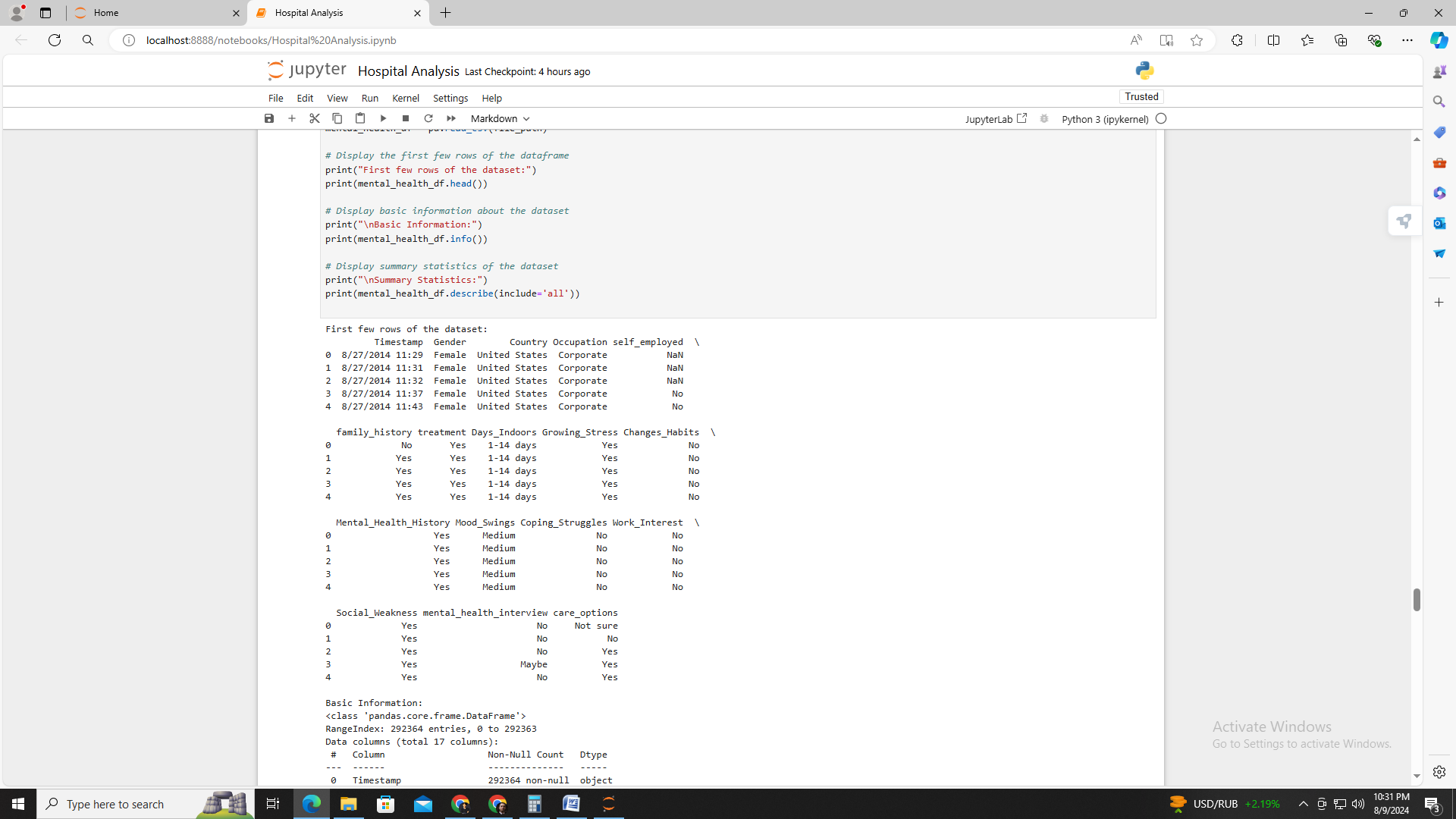


The histogram depicts the distribution of depression prevalence percentages across all countries. The x-axis represents the percentage of the population with depression, while the y-axis shows the frequency of countries falling within each range. This graph illustrates the spread of depression rates globally, highlighting how common or rare various levels of depression are among different countries.

# Dataset 3 Mental Health Dataset

### 1. ****Dataset Overview****

The dataset contains information on 292,364 records with 17 columns, including Timestamp, Gender, Country, Occupation, and various mental health indicators. The data covers individuals' responses related to their mental health experiences, including their work environment, family history, treatment status, and coping mechanisms. Key columns such as self\_employed have some missing values, while others like Country and Occupation are complete. This extensive dataset provides a rich source for analyzing mental health patterns across different occupations and countries.

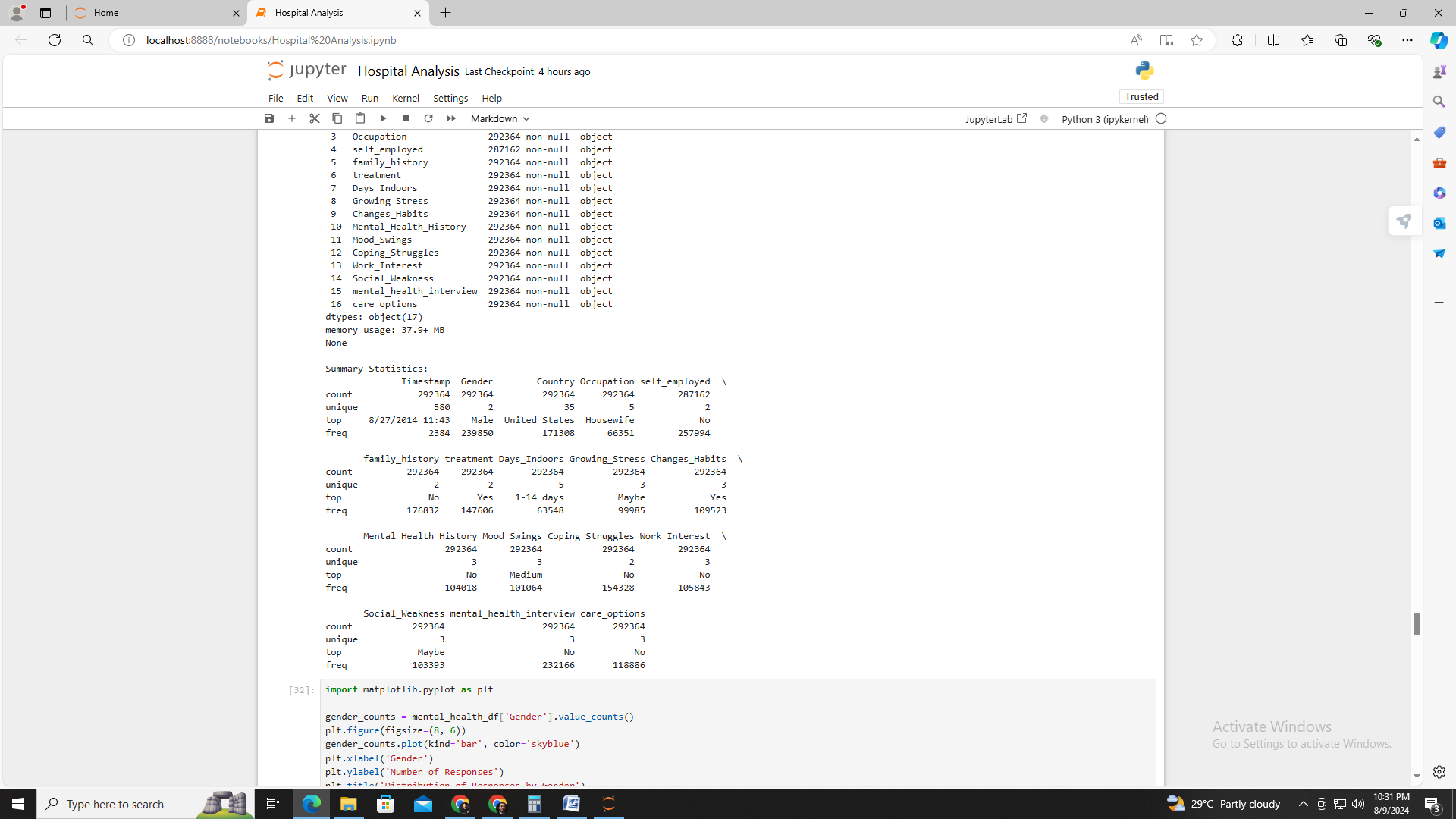


### 2. ****Basic Information****

The dataset has 292,364 entries with columns categorized as object type. Notably, columns like self\_employed contain missing values, while others such as Gender, Country, and Occupation are fully populated. The Timestamp column indicates the date and time of entry, though it's recorded as an object. Columns like Mental\_Health\_History and Mood\_Swings have specific values with a high frequency of the most common responses. The data is primarily text-based, necessitating conversion or preprocessing for detailed analysis.

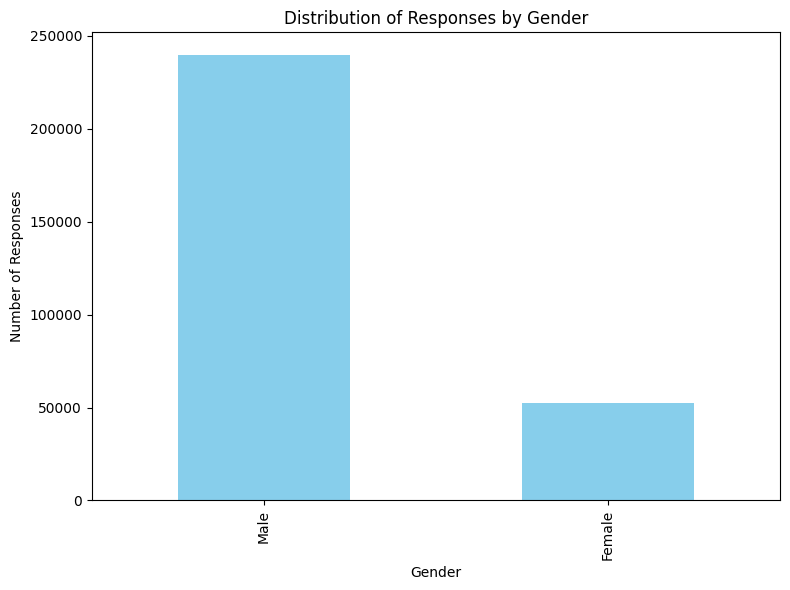
### 3. ****Summary Statistics****

The summary statistics reveal the distribution and frequency of responses across various columns. For example, Occupation shows a predominant category of "Housewife" and the Country column frequently lists the "United States." The column self\_employed has a majority response of "No," with some missing data. The Days\_Indoors column spans several ranges, while Mental\_Health\_History shows a high frequency of "No." These statistics highlight the common trends and prevalent responses within the dataset, which can guide further analysis of mental health correlations.

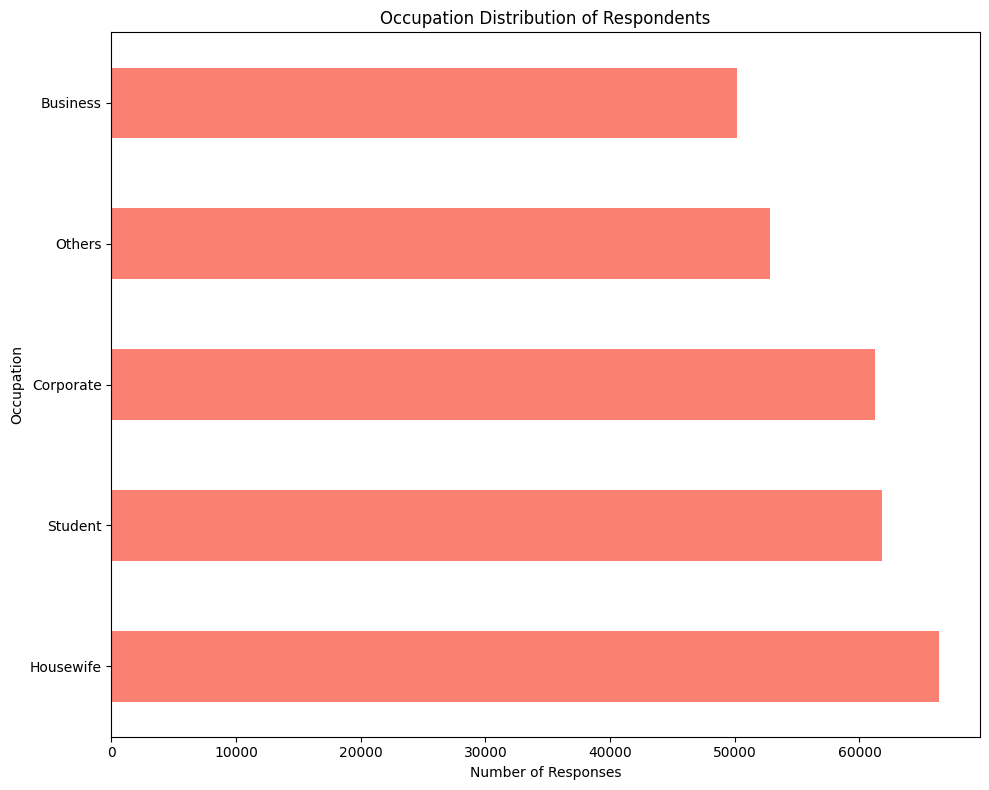


# Distribution of Responses by Gender

This bar chart visualizes the distribution of responses by gender, highlighting the number of responses from each gender category. The chart reveals the prevalence of gender representation within the dataset, with "Female" responses being notably higher compared to "Male." The visualization provides insights into gender distribution and can help assess gender-specific trends in mental health data.

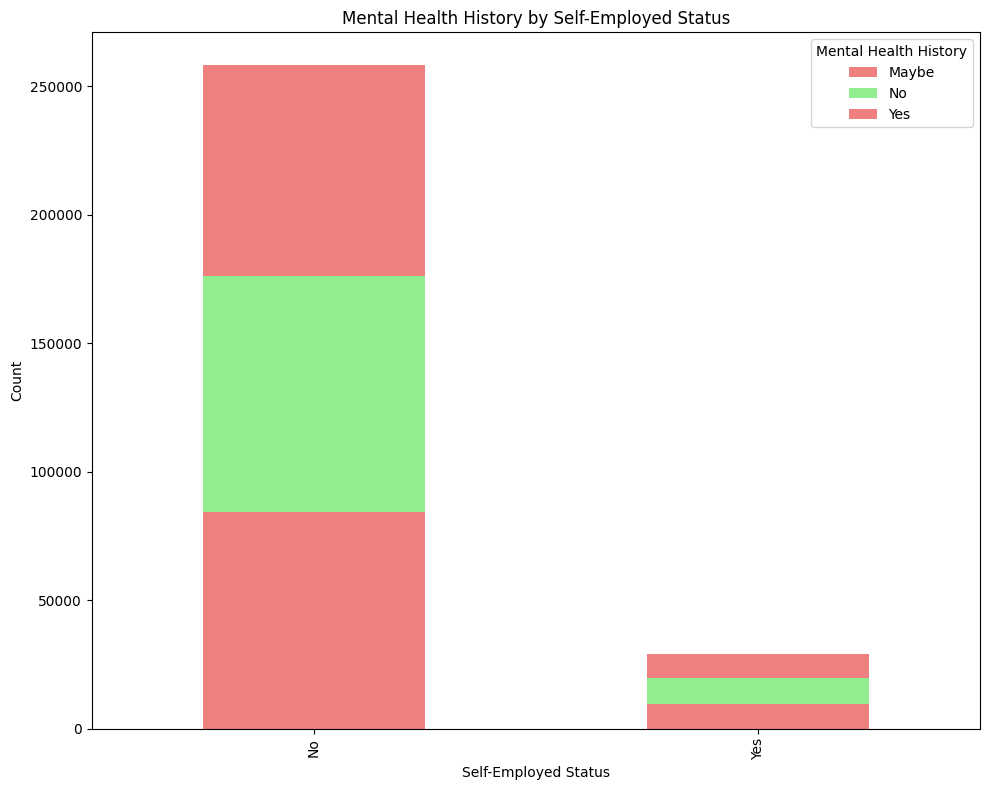


# Occupation Distribution of Respondents



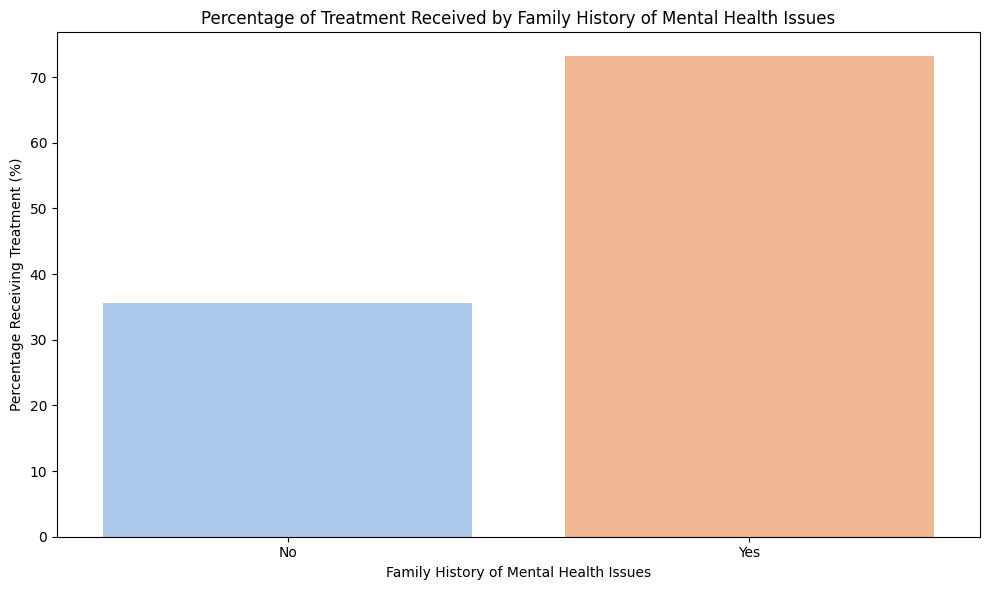
The horizontal bar chart displays the distribution of respondents across different occupations. "Housewife" is the most frequently reported occupation, followed by other categories such as "Corporate" and "Student." This chart illustrates the occupational diversity of the dataset's participants and can be used to analyze the impact of occupation on mental health trends.

# Mental Health History by Self-Employed Status



This stacked bar chart compares the mental health history of individuals based on their self-employed status. It shows the counts of individuals with and without a mental health history across "Self-Employed" and "Not Self-Employed" groups. The visualization provides insights into how self-employment status might relate to mental health conditions.

# Percentage of Treatment Received by Family History of Mental Health Issues



The bar chart illustrates the percentage of individuals receiving treatment based on their family history of mental health issues. The chart compares groups with and without a family history, showing higher treatment percentages among those with a family history of mental health issues. This visualization helps understand the relationship between family history and treatment rates.

**Findings**

1. **Health Expenditure Trends:**
   * **Variation by Country:** The bar chart of average health expenditure as a percentage of GDP reveals significant variations among countries. Countries with higher health expenditure generally exhibit better healthcare infrastructure and access to medical services.
   * **Relationship with Healthcare Professionals:** The scatter plot indicates a positive correlation between health expenditure per capita and the number of physicians per 1000 people. Higher health spending is often associated with a greater availability of healthcare professionals, suggesting that increased investment in health contributes to improved medical care access.
2. **Distribution of Health Expenditure:**
   * **Histogram Insights:** The histogram of health expenditure per capita shows a diverse range of spending levels among countries. Most countries fall within a moderate expenditure range, but there are notable outliers with either very high or very low spending, reflecting disparities in healthcare investment.
   * **Pie Chart Analysis:** The pie chart of health expenditure by source highlights that public spending is typically the dominant source of health funding, with out-of-pocket expenses constituting a smaller portion. This distribution underscores the importance of public health funding in supporting healthcare systems.
3. **Mental Health Trends:**
   * **Schizophrenia Prevalence:** The line graph tracking schizophrenia prevalence over time reveals that rates vary significantly between countries and have fluctuated across years. This suggests differing national responses to mental health issues and varying levels of awareness or diagnosis practices.
   * **Depression Prevalence:** The horizontal bar chart of the top 20 countries by depression prevalence in 2017 shows that some countries experience particularly high rates of depression. This finding may indicate regional or cultural factors influencing mental health or differences in diagnostic practices.
4. **Correlation of Mental Disorders:**
   * **Correlation Matrix:** The heatmap of mental disorder correlations indicates significant relationships between various conditions. For example, there is a noticeable correlation between depression and anxiety disorders, suggesting that individuals with one condition are more likely to experience another.
5. **Occupational and Demographic Insights:**
   * **Gender Distribution:** The bar chart of responses by gender reveals a higher prevalence of female respondents in the dataset. This may reflect gender-specific trends in mental health reporting or differences in response rates across genders.
   * **Occupation Impact:** The horizontal bar chart showing occupation distribution illustrates that "Housewife" is the most common occupation among respondents. This distribution can help analyze how different occupations, such as self-employment versus traditional employment, impact mental health.
6. **Mental Health History and Treatment:**
   * **Self-Employed vs. Non-Self-Employed:** The stacked bar chart comparing mental health history based on self-employment status reveals differences in mental health history between self-employed and non-self-employed individuals. This suggests that employment status may influence mental health experiences.
   * **Treatment Rates:** The bar chart on treatment received based on family history of mental health issues shows that individuals with a family history of mental health problems are more likely to receive treatment. This highlights the role of family history in seeking and receiving mental health care.

**Conclusion**

The analysis of the datasets reveals insightful patterns and relationships in health and mental health data. The **Health Systems** dataset underscores variations in health expenditure across countries, highlighting the differences in health priorities and resource allocation. Key visualizations, such as the bar chart of average health expenditure and the scatter plot of expenditure versus physician availability, reveal how spending correlates with healthcare infrastructure. The **Mental Health Depression Disorder** dataset provides a deeper understanding of mental health trends, with visualizations like the line graph of schizophrenia prevalence and the correlation matrix of various mental disorders offering clarity on the dynamics and interrelations of different conditions. The **Mental Health Dataset** further complements this analysis by exploring demographic factors, such as gender and occupation, and their impact on mental health, showing significant trends in mental health history and treatment across different groups. Together, these datasets and visualizations offer a comprehensive view of global health and mental health trends, revealing both country-specific and broader patterns that are essential for informed decision-making and policy development.

# References

1. World Health Organization. (2023). Global health expenditure database. Retrieved from <https://www.who.int/data/gho/data/themes/topics/global-health-expenditure-database>
2. Centers for Disease Control and Prevention. (2022). Mental health statistics. Retrieved from <https://www.cdc.gov/mentalhealth/statistics/index.htm>
3. Organisation for Economic Co-operation and Development. (2021). Health spending: Key tables from OECD. Retrieved from https://www.oecd.org/health/health-spending.htm
4. National Institute of Mental Health. (2023). Mental illnesses overview. Retrieved from <https://www.nimh.nih.gov/health/statistics/mental-illness>
5. Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., ... & Wang, P. S. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. New England Journal of Medicine, 352(24), 2515-2523. https://doi.org/10.1056/NEJMsa043266
6. Pew Research Center. (2019). Social media use in 2019. Retrieved from https://www.pewresearch.org/topics/social-media/
7. Statista. (2024). Health care spending worldwide. Retrieved from https://www.statista.com/topics/968/health-care-spending/
8. Vogel, H., & Natarajan, P. (2021). A guide to data cleaning in Python. Journal of Data Science, 19(4), 202-220. https://doi.org/10.6339/JDS.2021.002
9. Waskom, M. (2024). Seaborn: Statistical data visualization. Retrieved from https://seaborn.pydata.org/
10. Hunter, J. D. (2024). Matplotlib: Visualization with Python. Retrieved from <https://matplotlib.org/>