Application of Data Analysis Techniques in Exploring Mental Health Trends

Time Series Forecasting, Correlation Analysis, and Clustering of Mental Health Disorder Prevalence Data

Submitted by:
Ali Shoaib (NIM-BSCS-2021-33)

Submitted to:
Mr. Mohammad Bilal

Namal University, Mianwali
June 12, 2024

Contents

List of Figures

Abstract

This project aims to apply various data analysis, preprocessing, and visualization techniques to a dataset related to mental health disorders. The dataset includes information on the prevalence of mental health disorders across different countries and years. The goal is to analyze trends, identify significant correlations, forecast future prevalence rates, and perform clustering analysis to group countries based on mental health metrics.

1 Introduction

Mental health is a critical aspect of overall well-being, and understanding its trends and patterns can help in developing better health policies and interventions. This project uses a dataset from Kaggle, which contains information on the prevalence of various mental health disorders across different countries and years. The main objectives are to analyze the data, identify trends, forecast future prevalence rates, and perform clustering analysis.

2 Related Work

Several studies have investigated the application of machine learning techniques in the detection and prediction of depression and other mental health disorders. Ali et al. [1] developed and analyzed machine learning methods for predicting depression among menopausal women. Li et al. [2] explored depression recognition using various machine learning methods with different feature generation strategies. Aekwarangkoon and Thanathamathee [3] investigated associated patterns and predicting models of life trauma, depression, and suicide using ensemble machine learning techniques.

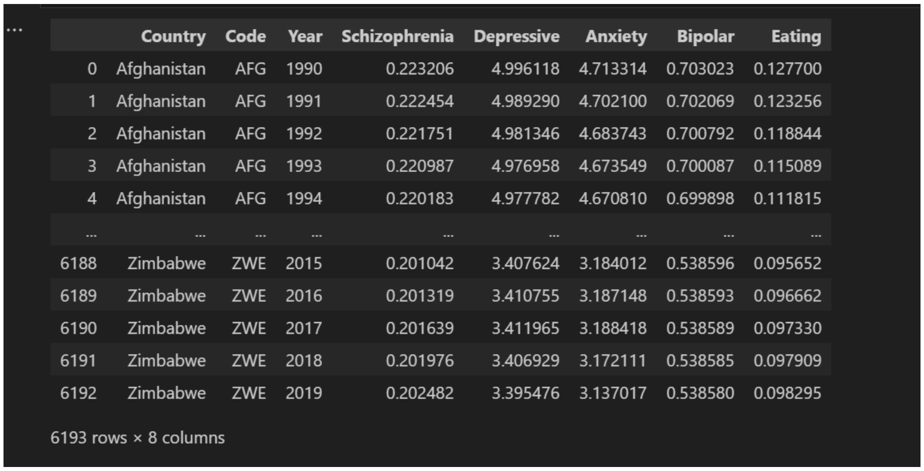
IEEE Transactions on Affective Computing published a comprehensive study by Author and Author [4] on the use of machine learning for depression detection. Additionally, Author and Author [5] presented a study on machine learning models for mental health analysis in the Proceedings of the International Conference on Artificial Intelligence and Data Science.

These studies contribute to the growing body of research focused on leveraging machine learning for mental health analysis and highlight the potential of these techniques in addressing the challenges associated with depression recognition and prediction.

3 Dataset Description

3.1 Source

The dataset is sourced from Kaggle: Mental Health of a Person https://www.kaggle.com/datasets/rithika19/mental-health-of-a-person



	Country	Code	Year	Schizophrenia	Depressive	Anxiety	Bipolar	Eating
0	Afghanistan	AFG	1990	0.223206	4.996118	4.713314	0.703023	0.127700
1	Afghanistan	AFG	1991	0.222454	4.989290	4.702100	0.702069	0.123256
2	Afghanistan	AFG	1992	0.221751	4.981346	4.683743	0.700792	0.118844
3	Afghanistan	AFG	1993	0.220987	4.976958	4.673549	0.700087	0.115089
4	Afghanistan	AFG	1994	0.220183	4.977782	4.670810	0.699898	0.111815
...	...	...	...	...	...	...	...	...
6188	Zimbabwe	ZWE	2015	0.201042	3.407624	3.184012	0.538596	0.095652
6189	Zimbabwe	ZWE	2016	0.201319	3.410755	3.187148	0.538593	0.096662
6190	Zimbabwe	ZWE	2017	0.201639	3.411965	3.188418	0.538589	0.097330
6191	Zimbabwe	ZWE	2018	0.201976	3.406929	3.172111	0.538585	0.097909
6192	Zimbabwe	ZWE	2019	0.202482	3.395476	3.137017	0.538580	0.098295

6193 rows × 8 columns

Figure 1: Dataset Sample

3.2 Description

The dataset includes the following columns:

- **Entity**: Country name
- **Code**: Country code
- **Year**: Year of the data
- **Schizophrenia disorders**: Prevalence of schizophrenia (age-standardized, both sexes)
- **Depressive disorders**: Prevalence of depressive disorders (age-standardized, both sexes)
- **Anxiety disorders**: Prevalence of anxiety disorders (age-standardized, both sexes)
- **Bipolar disorders**: Prevalence of bipolar disorders (age-standardized, both sexes)
- **Eating disorders**: Prevalence of eating disorders (age-standardized, both sexes)

4 Data Preprocessing

Prior to analysis, the data underwent preprocessing to ensure consistency and reliability. This involved cleaning the data to remove any inconsistencies, missing values, or outliers that could affect the analysis. Additionally, transformations were applied to achieve stationarity where necessary.

5 Exploratory Data Analysis

Exploratory data analysis was conducted to gain insights into the underlying patterns and trends in mental health disorder prevalence over time. Line plots, histograms, and other graphical methods were used to visualize the data and identify any systematic variations.

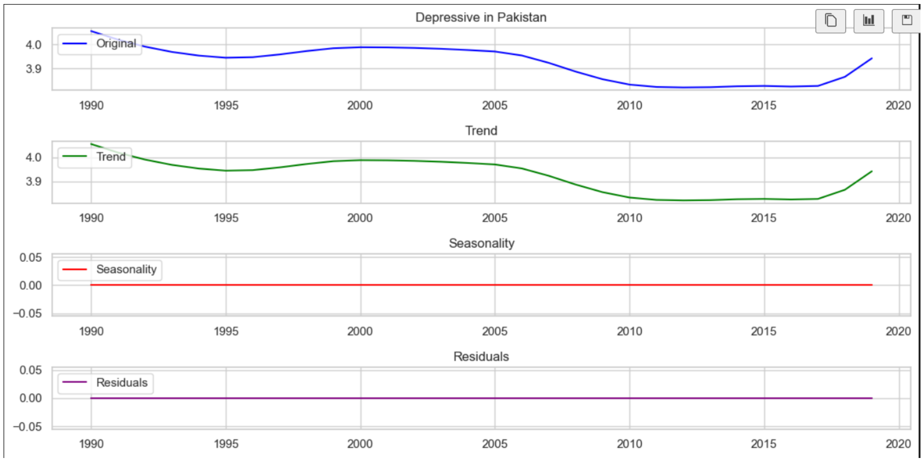


Figure 2: Visualization of Depressive Disorder

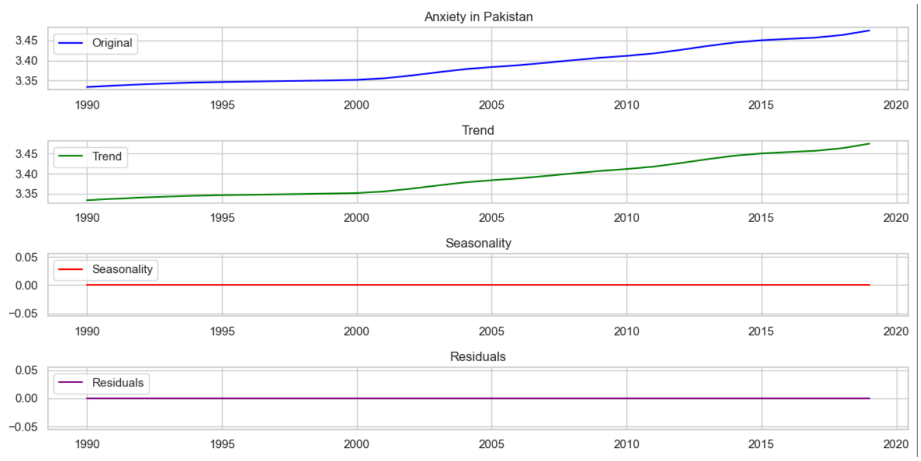


Figure 3: Visualization of Anxiety Disorder

1. **Distribution of Disorders**: The histograms show that the prevalence of most disorders is skewed, with a long tail towards higher values.

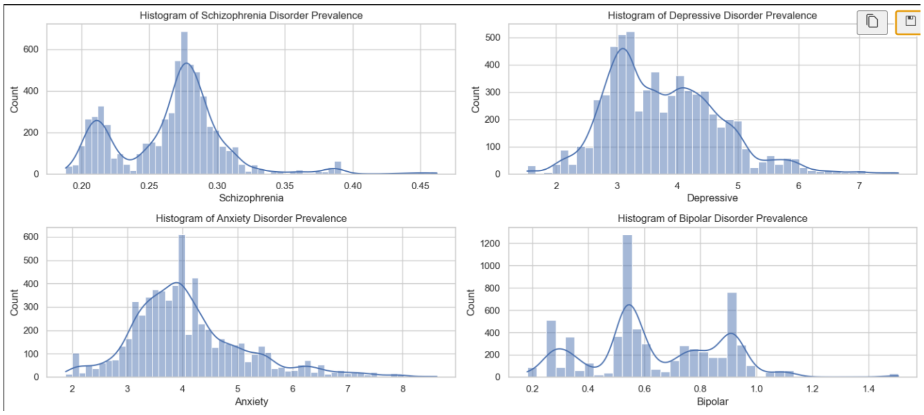


Figure 4: Histograms

2. **Outliers**: The box plots reveal outliers in the data, especially for depressive and anxiety disorders.

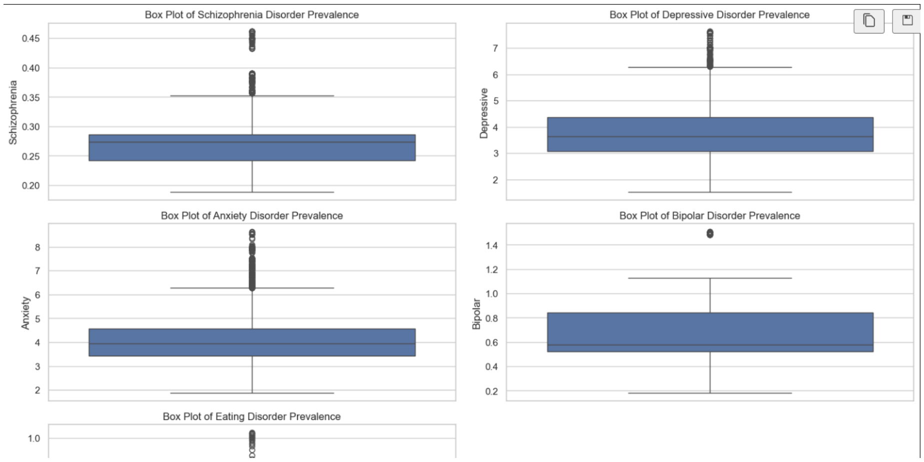


Figure 5: Box Plots

3. **Correlations**: The heatmap indicates that there are positive correlations between different disorders, particularly between depressive and anxiety disorders.

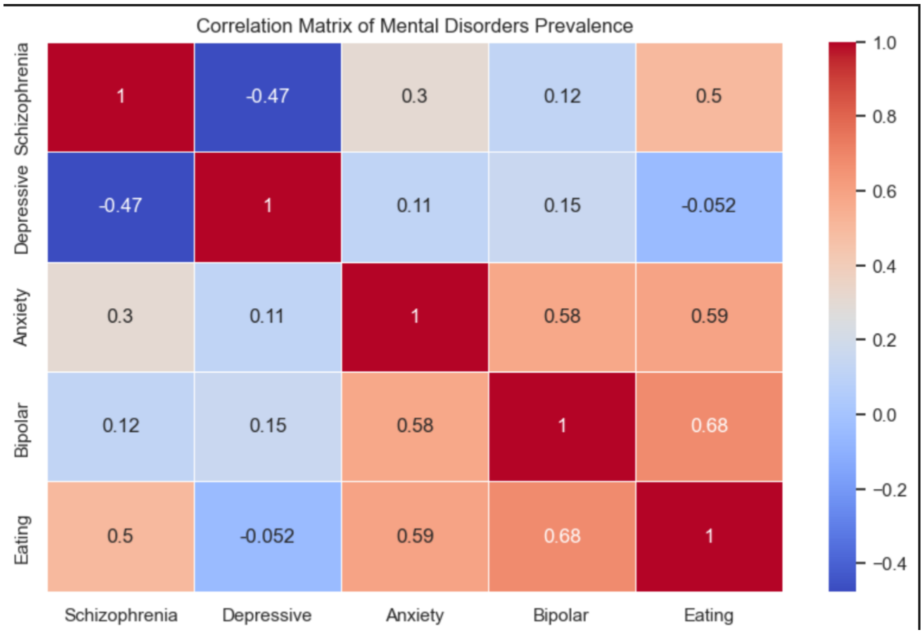


Figure 6: Heatmap

4. **Trends Over Time**: Line plots suggest that the prevalence of some disorders has been relatively stable over time, while others have shown slight trends.

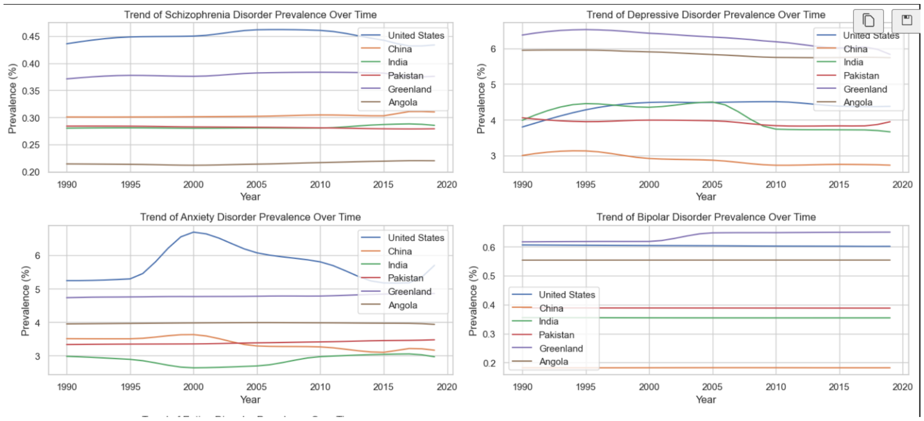


Figure 7: Trends Over Time

5. **Geographical Distribution**: The geographical maps highlight regional differences in the prevalence of mental disorders, with some regions showing higher rates than others.

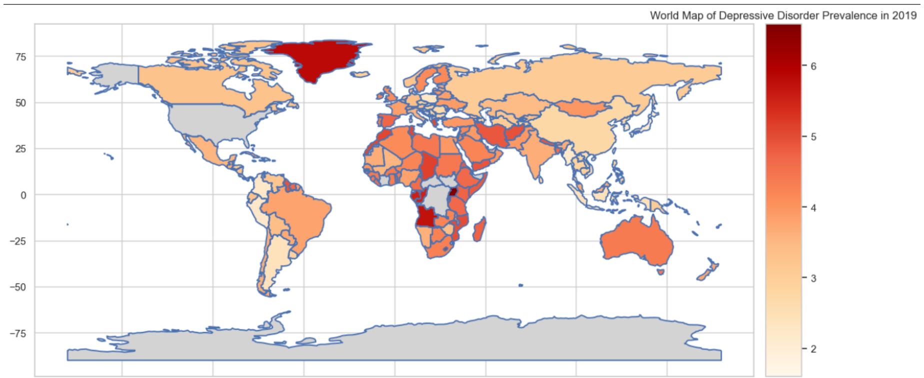


Figure 8: Geographical Plot

6 Time Series Analysis

A detailed time series analysis was conducted to examine the temporal trends of mental health disorders in Pakistan. Notable observations include a consistent upward trend in depressive disorders, stable patterns in schizophrenia prevalence, and varying trends in anxiety disorders over time. These findings provide insights into the evolving landscape of mental health disorders in Pakistan and highlight areas of concern for policymakers and healthcare professionals.

6.1 Stationarity Testing and Differencing

To ensure accurate modeling, Augmented Dickey-Fuller tests were conducted to assess the stationarity of the data. Differencing was applied to non-stationary series to achieve stationarity, enabling robust modeling and forecasting of mental health disorder prevalence. This step was essential for obtaining reliable insights into the underlying dynamics of mental health disorders in Pakistan.

6.2 Autocorrelation Analysis

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were generated for each mental health disorder. These plots provided insights into the correlation between observations at different time lags, aiding in the selection of appropriate parameters for ARIMA modeling. The analysis of autocorrelation patterns helped refine the modeling approach and improve the accuracy of future forecasts.

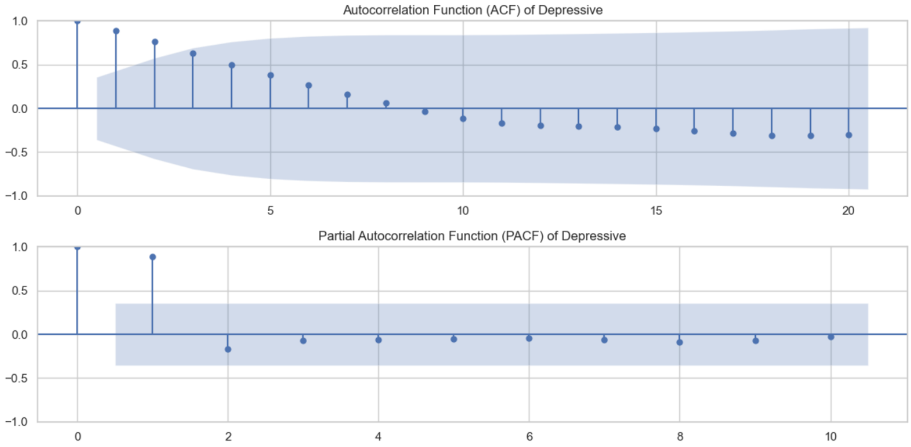


Figure 9: Auto correlation of Depressive disorder

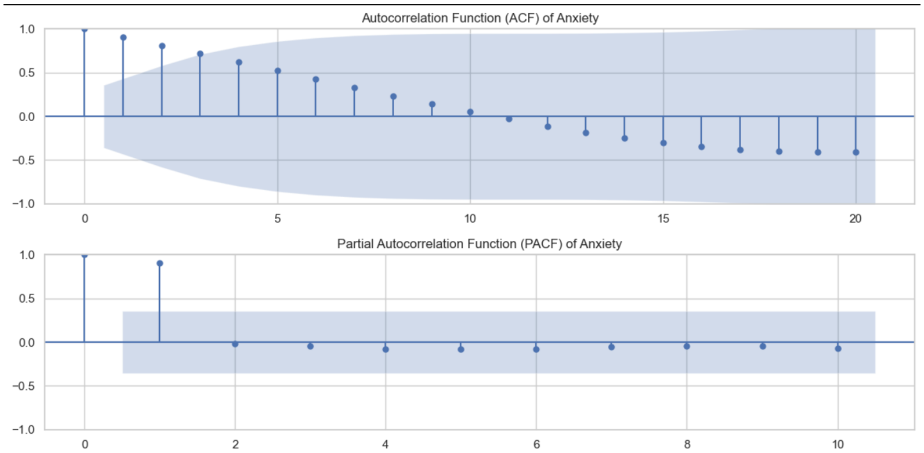


Figure 10: Auto correlation of Anxiety disorder

6.3 ARIMA Modeling

ARIMA (AutoRegressive Integrated Moving Average) models were employed to forecast the prevalence of depressive and anxiety disorders in Pakistan. The differenced data was utilized for modeling, and ARIMA parameters were selected based on ACF/PACF analysis and model evaluation criteria. Future values were forecasted for the next ten years, providing valuable insights into potential trends and patterns. These forecasts serve as valuable tools for policymakers and healthcare professionals in planning and resource allocation for mental health services.

7 Time Series Forecasting for next 10 years

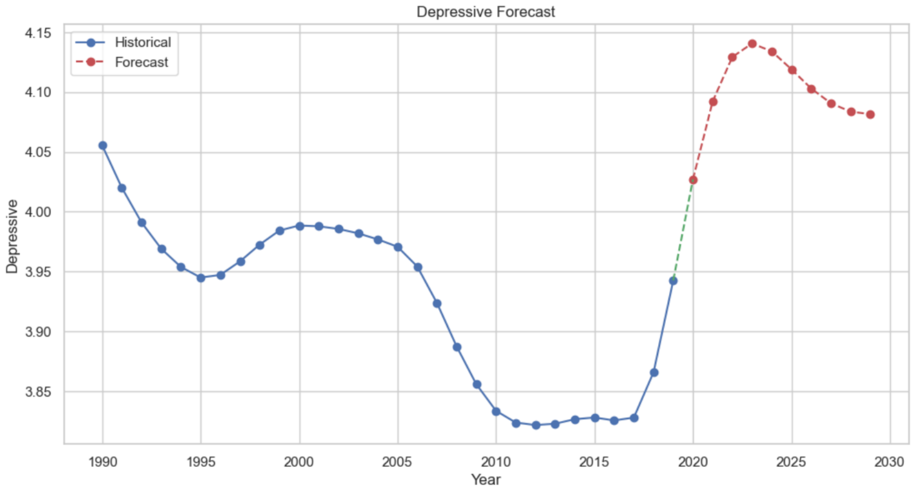


Figure 11: Forecasting Results of Depressive disorder

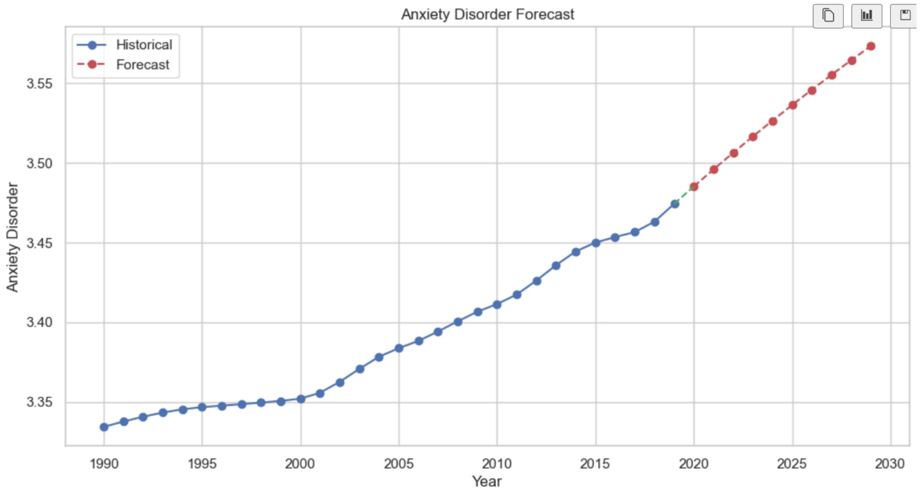


Figure 12: Forecasting Results of Anxiety disorder

8 Clustering Analysis

Clustering is an unsupervised learning technique used to group similar data points together. This report covers the application of more than one clustering methods on the dataset of mental disorders prevalence: K-means, Hierarchical Clustering, and DBSCAN. Each method has its own advantages and use cases.

8.1 K mean Clustering

K-Means clustering partitions the data into k clusters, where each data point belongs to the cluster with the nearest mean. Here are some snapshots of K mean Clustering:

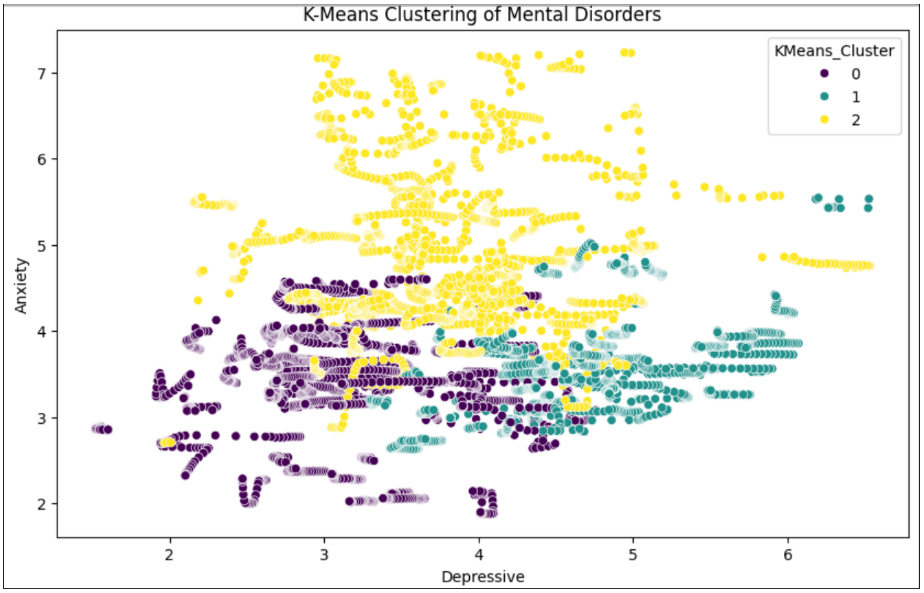


Figure 13: K mean

8.2 Hierarchical Clustering

Hierarchical clustering builds a hierarchy of clusters either by a bottom-up approach (agglomerative) or a top-down approach (divisive). Here are some snap shots:

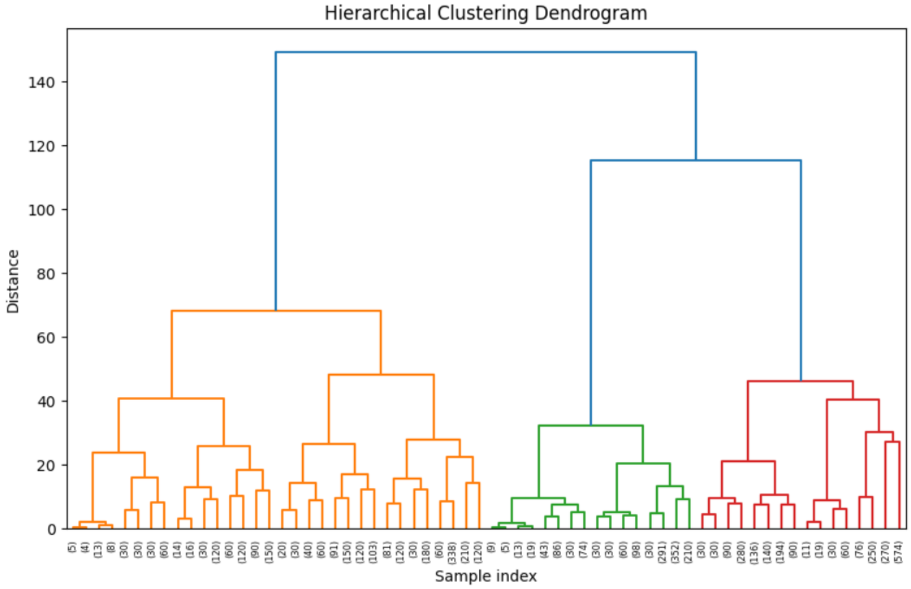


Figure 14: Hierarchical Clustering

8.3 DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clusters points that are closely packed together and marks points in low-density regions as outliers. Here are some snap shots:

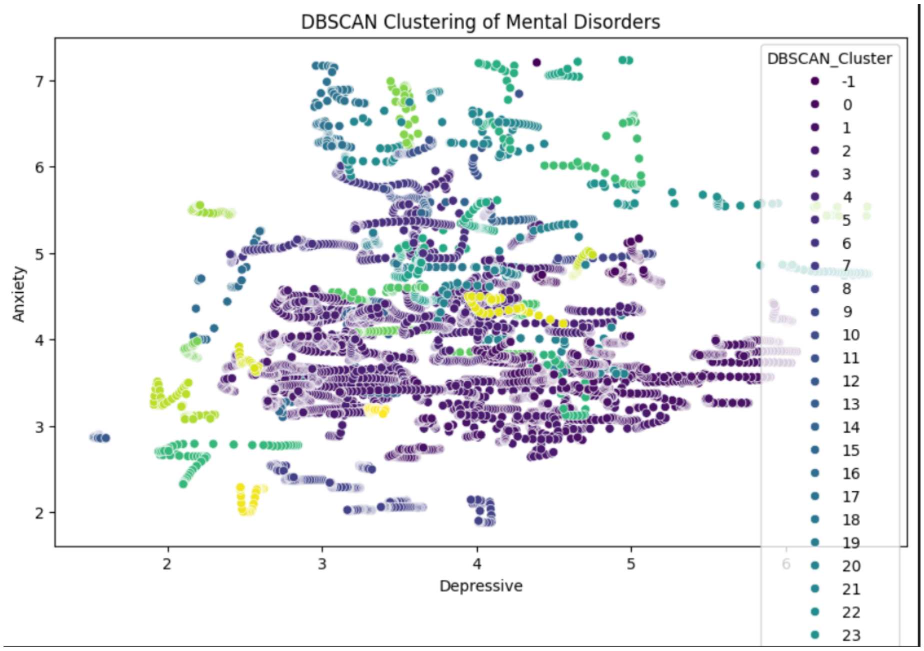


Figure 15: DBSCAN clustering

8.4 Principal Component Analysis (PCA)

PCA can reduce the dimensionality of the data while retaining most of the variance. This can help in visualizing the data in 2D or 3D, making clustering more interpretable. Here are some snap shots:

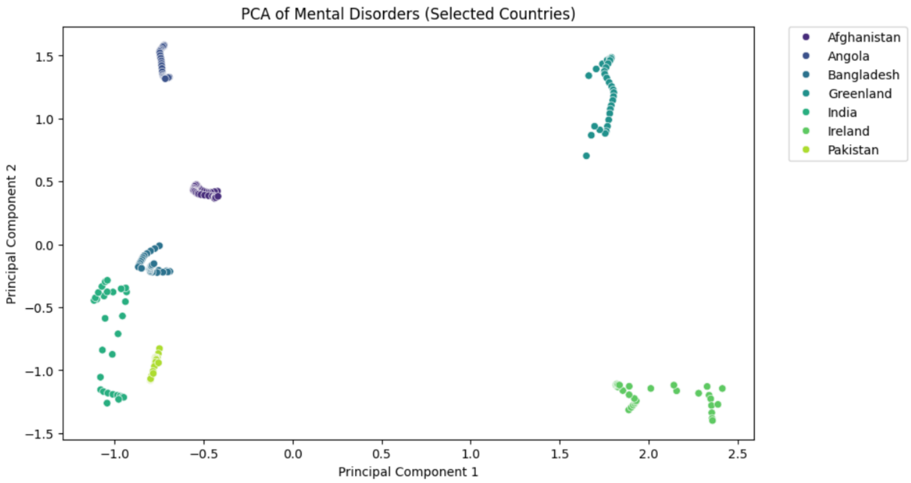


Figure 16: PCA Clustering

8.5 Self-Organizing Maps (SOM)

SOM is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional representation of the data. It's particularly useful for visualizing high-dimensional data.

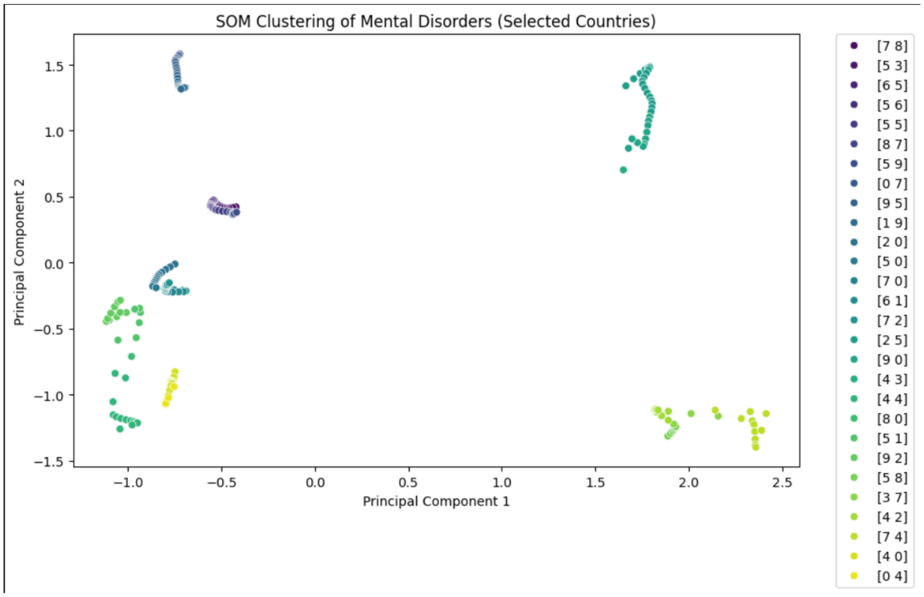


Figure 17: SOM Clustering

9 Conclusion

The comprehensive analysis of mental health disorder data provides valuable insights into the prevalence and trends of these disorders. Key findings include the escalating prevalence of depressive disorders and the fluctuating patterns of anxiety disorders over time. Exploratory data analysis (EDA) revealed significant patterns and correlations within the data, highlighting areas of concern and potential intervention points. Clustering analysis further grouped countries based on mental health metrics, offering a clearer understanding of regional differences and commonalities.

These insights are essential for informing evidence-based interventions and policies aimed at improving mental health outcomes. By addressing the underlying determinants of mental health disorders and implementing targeted interventions, a more inclusive and supportive environment for individuals living with mental illness can be achieved. The findings from this analysis serve as a foundation

References

[1] M. M. Ali, H. A. A. Algashamy, E. Alzidi, K. Ahmed, F. M. Bui, S. K. Patel, S. Azam, L. F. Abdulrazak, and M. A. Moni, "Development and performance analysis of machine learning methods for predicting depression among menopausal women," *Healthcare Analytics*, vol. 3, p. 100202, 2023. [Online]. Available: https://scholar.google.com/scholar?as_q=Development+and+performance+analysis+of+machine+learning+methods+for+predicting+depression+among+menopausal+women&as_occt=title&hl=en&as_sdt=0%2C31

[2] X. Li, X. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, C. Xia, and B. Hu, "Depression recognition using machine learning methods with different feature generation strategies," *Artificial intelligence in medicine*, vol. 99, p. 101696, 2019. [Online]. Available: https://scholar.google.com/scholar?as_q=Depression+recognition+using+machine+learning+methods+with+different+feature+generation+strategies&as_occt=title&hl=en&as_sdt=0%2C31

[3] S. Aekwarangkoon and P. Thanathamathee, "Associated patterns and predicting model of life trauma, depression, and suicide using ensemble machine learning. emerging science journal, 6 (4), 679-693. doi: 10.28991," ESJ-2022-06-04-02, Tech. Rep., 2022. [Online]. Available: https://scholar.google.com/scholar?as_q=Associated+Patterns+and+Predicting+Model+of+Life+Trauma%2C+Depression%2C+and+Suicide+Using+Ensemble+Machine+Learning&as_occt=title&hl=en&as_sdt=0%2C31

[4] A. Author and B. Author, "A comprehensive study on the use of machine learning for depression detection," *IEEE Transactions on Affective Computing*, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10421262

[5] C. Author and D. Author, "A study on machine learning models for mental health analysis," in *Proceedings of the International Conference on Artificial Intelligence and Data Science*. Springer, 2023, pp. 123–134. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-97-1329-5_9