

Mental Disorder Analysis and Prediction:

In this project we are dealing with mental disorder and analysis of different symptoms to diagnose Mania Bipolar Disorder, Depressive Bipolar Disorder, Major Depressive Disorder, and Normal patients. We have several behavioral symptoms as features in our dataset. We dealt with different preprocessing techniques in order to build the final model, We used label encoding to encode our categorical variables and data frame will be able to use in a Machine Learning Model.

Statistical Moments:

	Mean	Median	Standard Deviation	Skewness	Kurtosis
Sadness	1.88333	2	1.07049	-0.599678	-0.87858
Euphoric	1.63333	2	0.849452	0.0330279	-0.666093
Exhausted	1.63333	2	1.14447	-0.265839	-1.34773
Sleep disorder	1.75833	2	1.05317	-0.421472	-1.00357
Mood Swing	0.475	0	0.501468	0.101397	-2.02373
Suicidal thoughts	0.475	0	0.501468	0.101397	-2.02373
Anorxia	0.383333	0	0.488237	0.48601	-1.79398
Authority Respect	0.391667	0	0.490169	0.449516	-1.8287
Try-Explanation	0.475	0	0.501468	0.101397	-2.02373
Aggressive Response	0.483333	0	0.501817	0.0675511	-2.02955
...					
Overthinking	0.541667	1	0.50035	-0.169373	-2.00501
Sexual Activity	3.74167	4	2.01046	0.115509	-0.862765

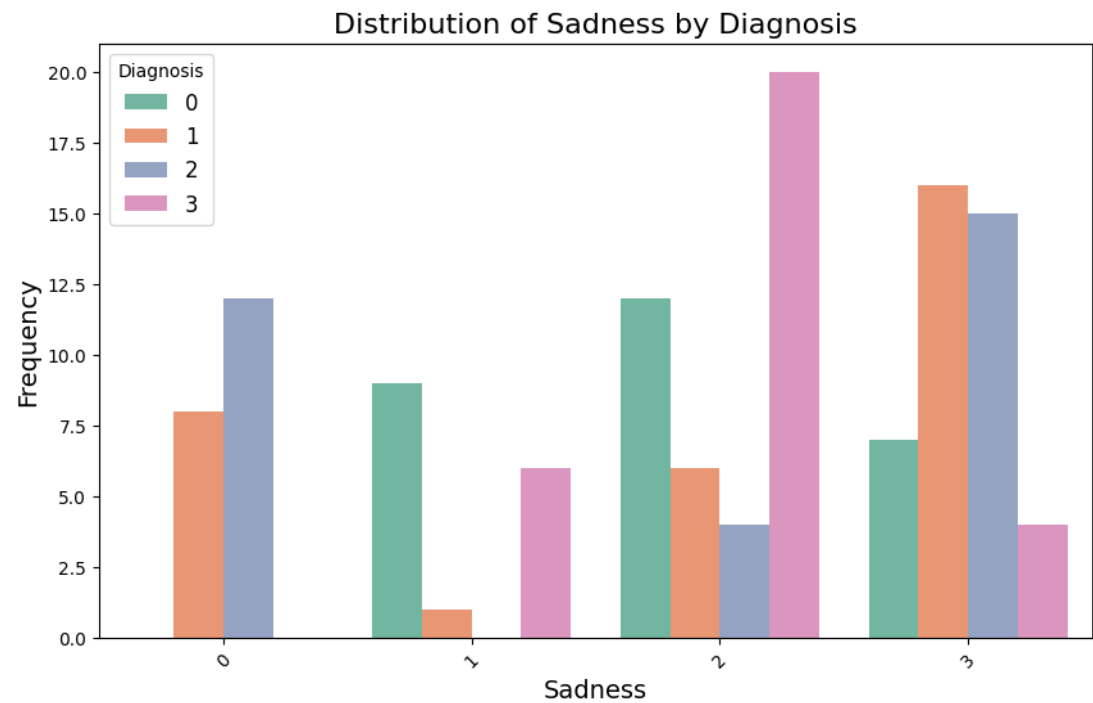
To get to know about distribution of columns and statistical measurement we calculated descriptive data and also statistical moments which tell us different insights:

It seems that all the columns have negative kurtosis which means that it's palykurtic distributed other than normally distributed.

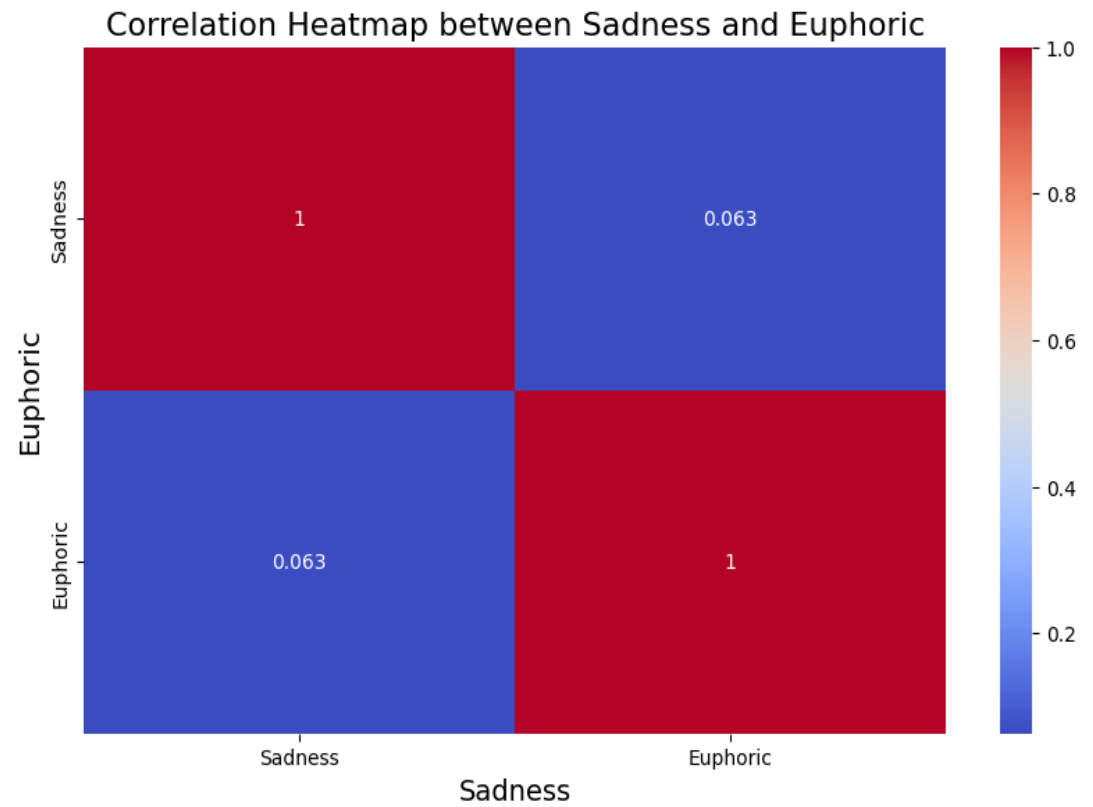
df.describe()

	Sadness	Euphoric	Exhausted	Sleep disorder	Mood Swing	Suicidal thoughts	Anorxia	Authority Respect	Try-Explanation	Aggressive Response	Ignore & Move-On	Nervous Break-down	Admit Mistakes
count	120	120	120	120	120	120	120	120	120	120	120	120	120
unique	4	4	4	4	2	3	2	2	2	2	2	2	2
top	Usually	Seldom	Sometimes	Sometimes	NO	NO	NO	NO	NO	NO	NO	YES	NO
freq	42	46	38	44	63	63	74	73	63	62	70	62	61

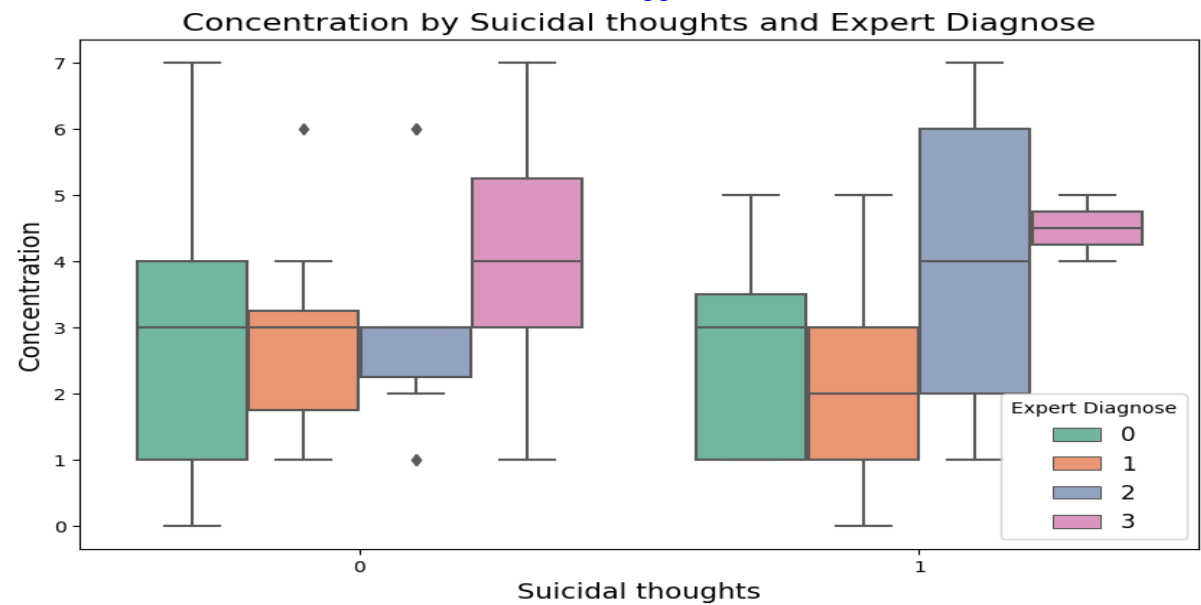
We also able to see description of categorical columns where we are able to see the frequency of each column variable and top variable. We are also able to find our unique values.



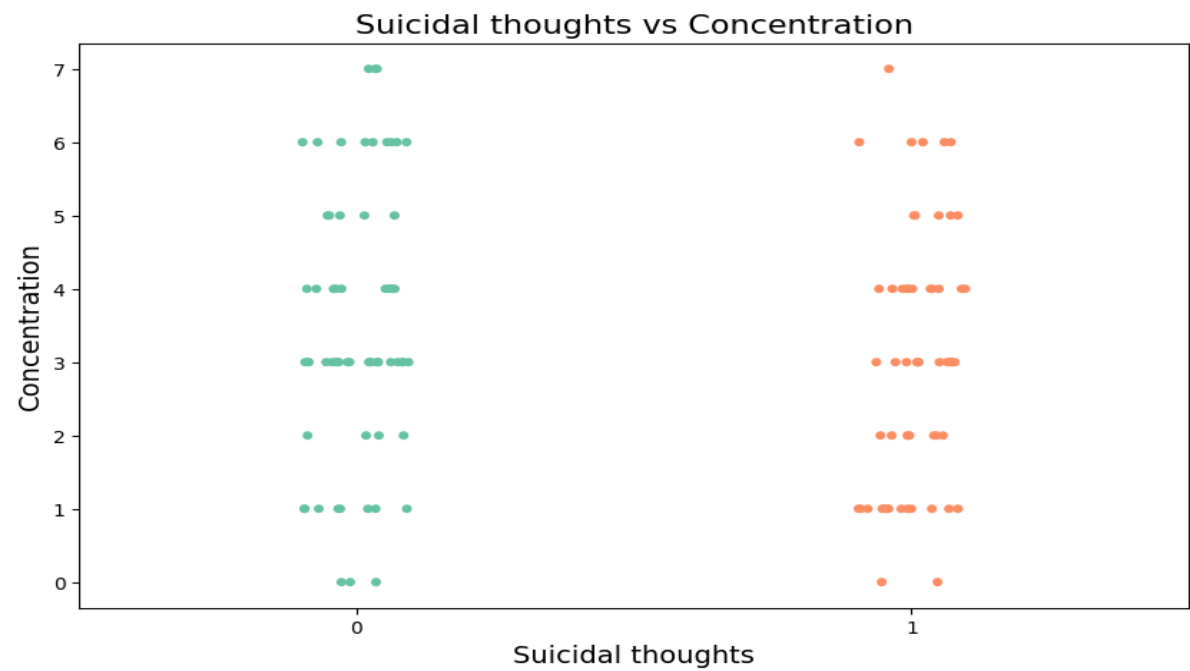
We plotted barplot to get to know about sadness distribution by diagnosis. We are able to see that people having most-often sadness having bi-polar type 1 very much also we plotted for usual and sometime sadness, We check having no depression to people who having sometimes sadness.



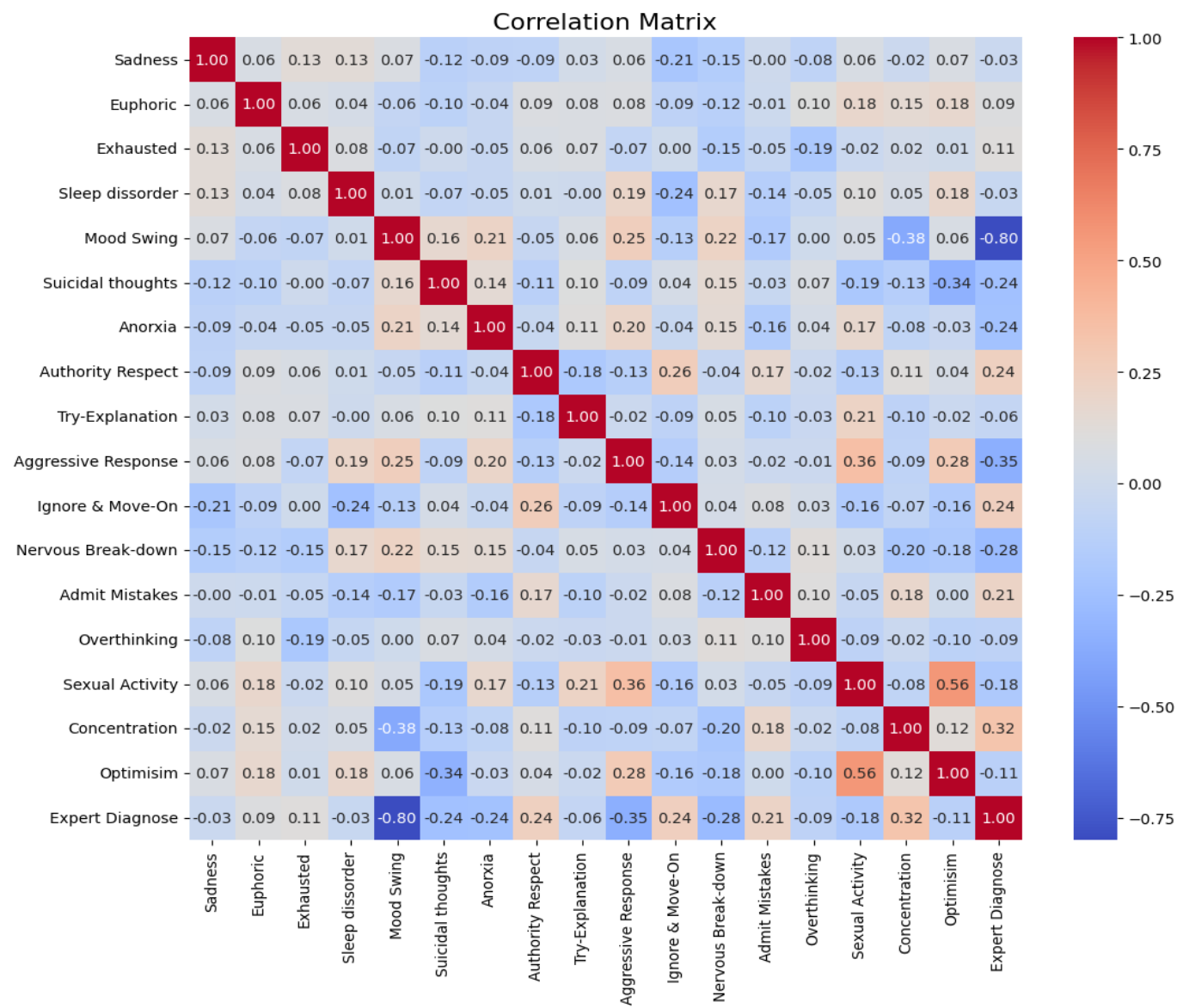
The correlation heatmap give us this insight that Europhobic and Sadness are very weak positive correlated in our dataset. They both are weak positively correlated means it will not on each other.

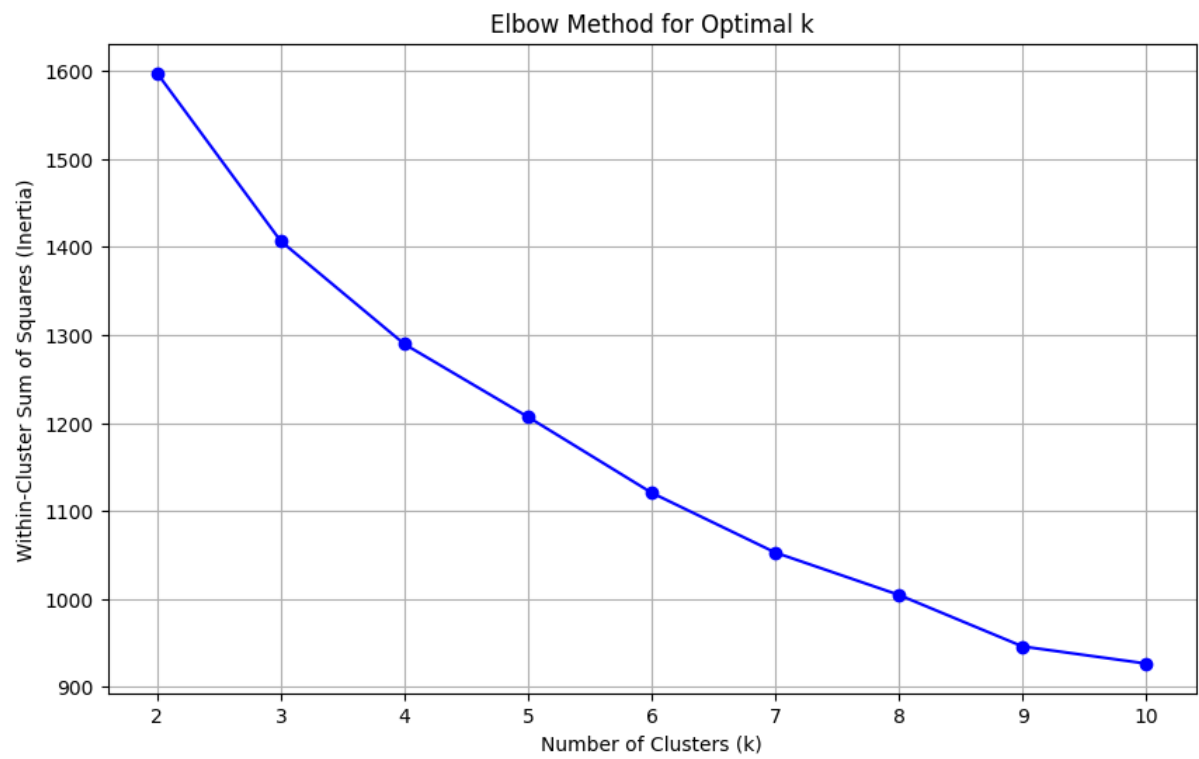


Here we found boxplot we got to know interesting facts, like people having suicidal thoughts and having much concentration are to like to be diagnosed with depression. Just like this we also found that people having no suicidal thoughts and having medium concentration are likely to be diagnosed with bipolar type-2 disease

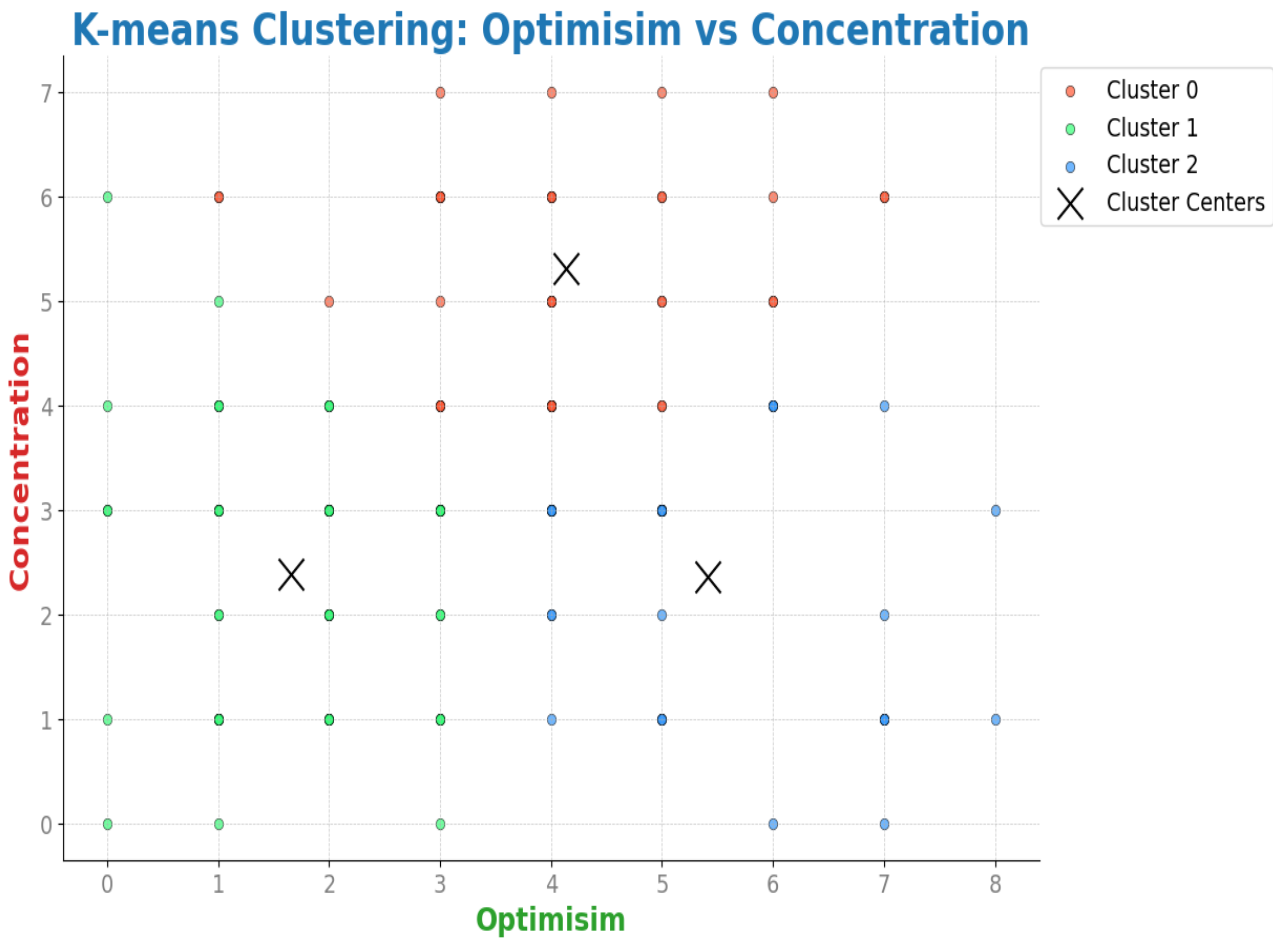


The scatterplot distribution shows very balanced distribution between people having no suicidal thoughts and having suicidal thoughts both have rising concentration.

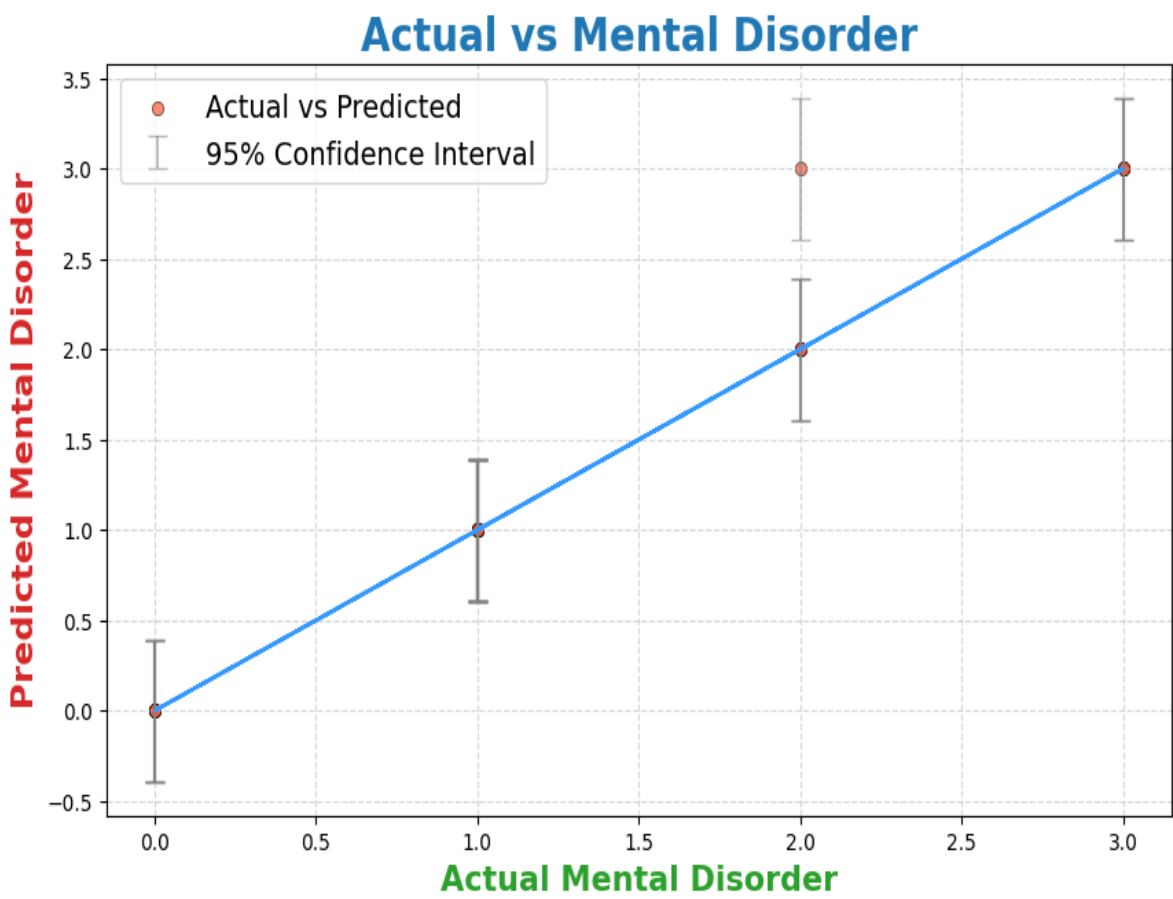




The elbow point in the plot, which appears to be around 3 or 4 clusters, represents the point where the rate of decrease in inertia starts to slow down significantly. Therefore, we look for the value of k where the elbow is formed. In this case, we can see that the elbow is formed at k=3 but there is not hard and fast rule behind it, indicating the optimal number of clusters for the dataset.



The plot shows the clusters produced by the K-Means Clustering algorithm applied to the data. KMeans is the algorithm to find clustering and relationship between variables to find out the actual numbers of ground in target variable, We can very linear relation concentration and optimism defining target variable (Expert Diagnose).



The regression line is displayed through graphs, x-axis displays the number of mental disorder that are expected to develop, while the y-axis displays the number of mental disorders that are now in existence. The data points shown on the line show the relationship between a person actual number of mental disorder and the number of disorders that were anticipated for them. The solid line, which is displayed in the graph's center, is the optimum fit line. This line represents the ideal situation where a person has exactly the predicted number of mental disorders and real mental diseases. The solid line is surrounded by a black circle, which represents the 95% confidence interval. The range of anticipated values related to mental disorders