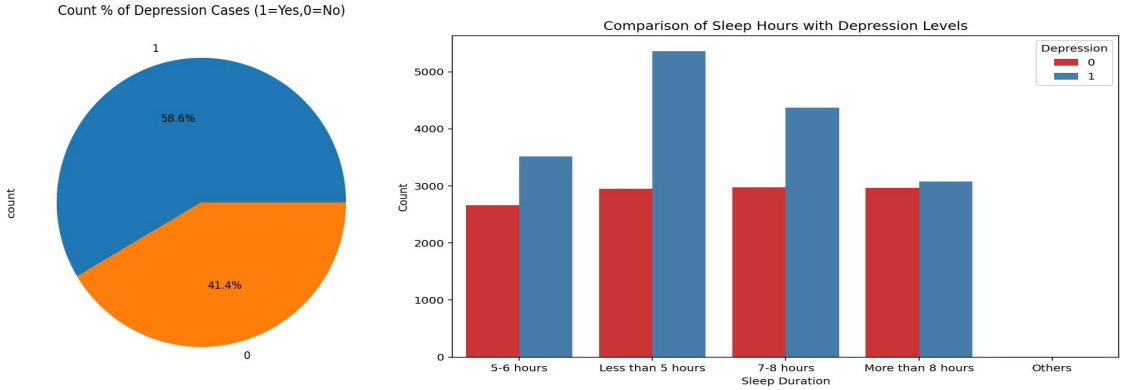
Clustering and Fitting: Student Depression dataset Analysis Report
Name: Praveen Kumar Savariraj
Student Number: 23089117
GitHub Repository: https://github.com/PKSR-DS/Student-depression-analysis

Introduction: -
In this report I have used Statistical methods, Clustering, Linear regression fitting techniques to analyse various factors that causing depression among students. I have considered following key columns and its values from dataset to analysis this report such as Age, Academic Pressure, CGPA, Study Satisfaction, Work/Study Hours, Financial Stress, Depression, Work Pressure, Job Satisfaction. I have downloaded the dataset from- https://www.kaggle.com/datasets/hopesb/student-depression-dataset

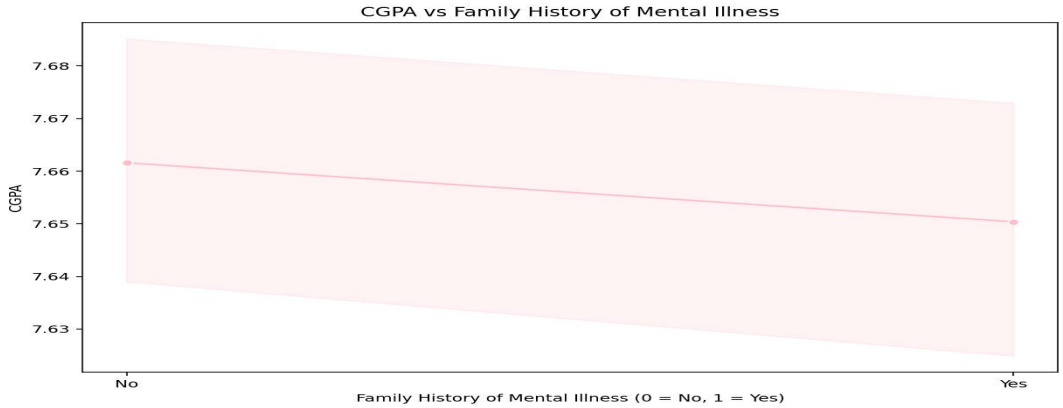
Dataset Overview: -
I have cleaned the dataset and dropped invaluable columns such as ID and City. Key metrics of dataset columns are age, academic pressure, CGPA, Study Satisfaction, work/study hours, financial stress and depression. I have considered depression column as target. From the data set average age of students are 25.8 year with median 25. The age data distribution shows minimal skewness of 0.132 and Kurtosis value -0.846 indicates that age data relatively normal distribution. Similarly mean, median, skewness and Kurtosis provided for other important columns in below table.

Metric	Mean	Median	Std Dev	Skewness	Kurtosis
Age	25.822	25.00	4.905	0.132	-0.846
Academic Pressure	3.141	3.00	1.381	-0.135	-1.161
CGPA	7.656	7.77	1.470	-0.113	-1.023
Study Satisfaction	2.944	3.00	1.361	0.010	-1.223
Work/Study Hours	7.157	8.00	3.707	-0.455	-1.000
Financial Stress	3.140	3.00	1.437	-0.130	-1.325
Depression	0.586	1.00	0.493	-0.347	-1.880
Work Pressure	0.0004	0.00	0.044	108.589	12109.334
Job Satisfaction	0.0007	0.00	0.044	74.102	5926.360

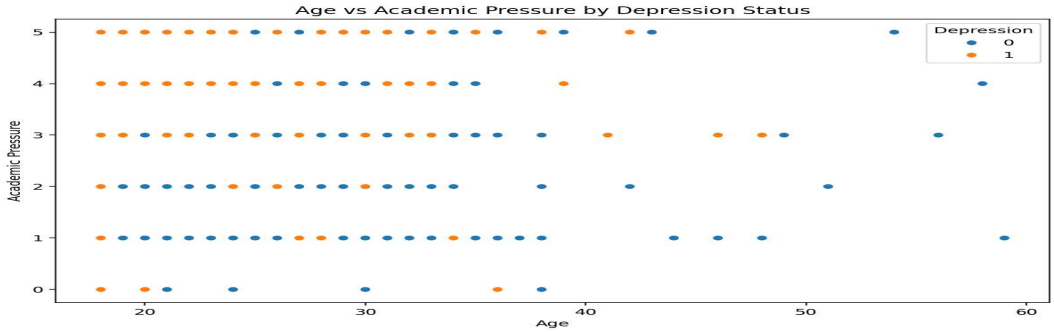
Statistical Analysis and Visualizations with Plots: -
Pie chart (Count % of Depression cases) and Bar Chart: Sleep Duration vs Depression Levels
Pie Chart: From the pie chart we could understand that **58.6%** of students affected by depression and **41.4%** of students have good mental health without any depression.
Bar Chart: The students with less than 5 hours sleep pattern, have direct correlation with depression. **60% of students** with fewer sleep hours have affected by depression. On the other hand, the students with proper sleep patterns have less depression cases**. Hence sleep pattern important factor for students mental health.**



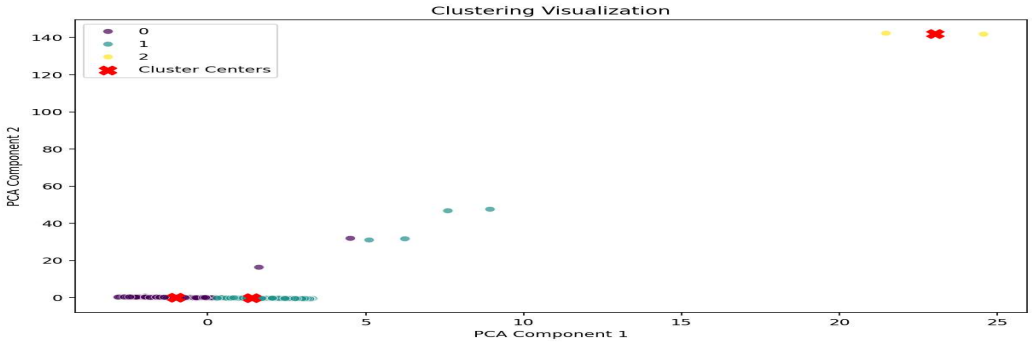
Line Plot: Family History of Mental Illness vs CGPA shows impact of a family history of mental illness on student CGPA scores. Students with family history of mental illness obtained slightly less CGPA score an average as 6.5 and Students without any family history of mental illness scored higher CGPA as average of 7.2.



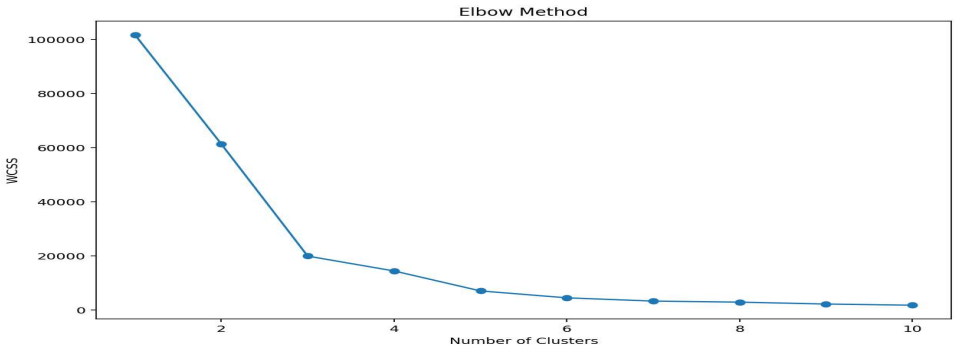
Scatterplot to analyse Age vs Academic Pressure by Depression Status:- This plot explains that students with high academic pressure affected by depression. Students with less academic pressure have good mental health.



Cluster Analysis: -I used K-means cluster and identified distinct groups between students based on normalized metrics. The best number of clusters was identified using the elbow method and silhouette score. A scatter plot indicates students with similar characteristics were successfully grouped. 2 Clusters: Silhouette Score = 0.652**, 3 Clusters: Silhouette Score = 0.654,** 4 Clusters: Silhouette Score = 0.550, 5 Clusters Score 0.552

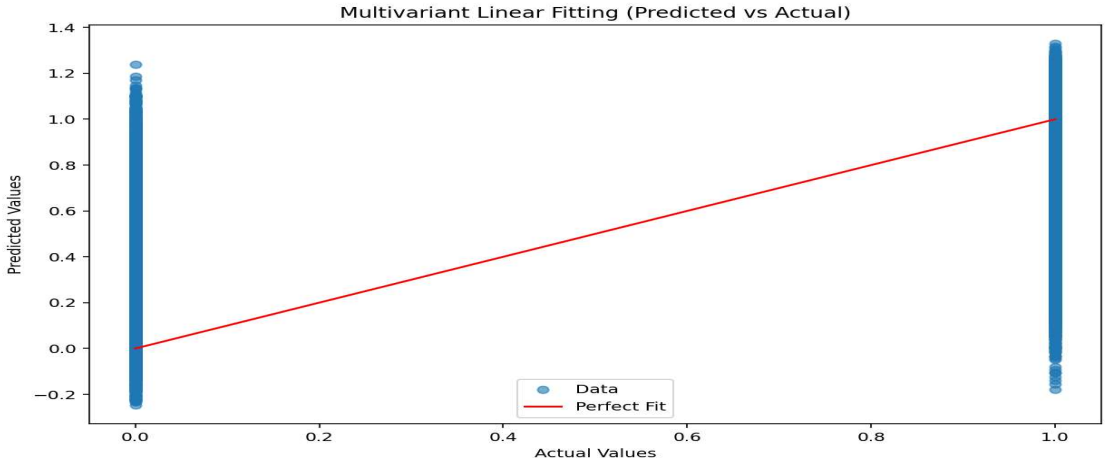


Elbow Method Plot:- The elbow method plot showed that WCSS decreased significantly up to 3 clusters, so considered this as optimal cluster number.



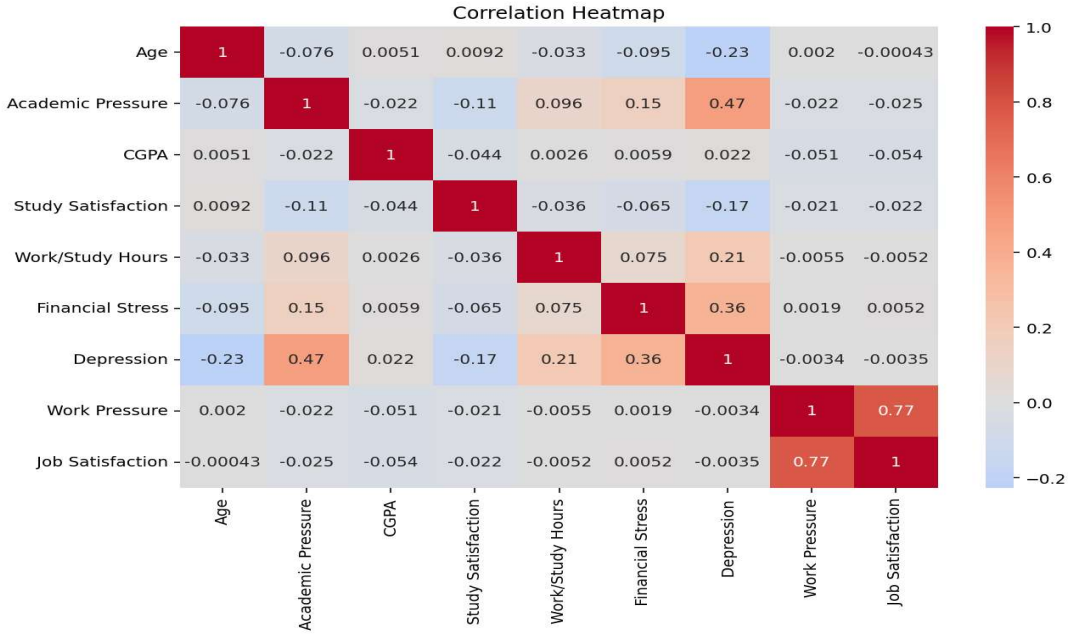
Linear Regression Fitting Analysis: -

I used linear regression model fitting to predict actual values vs predicted values. As per below fitting plot, it's perfectly fitting. Model Coefficients values for key metrics: Age: -0.081403, Academic Pressure: 0.195811, CGPA:0.01278, Study Satisfaction: -0.048439, Work/StudyHours:0.069474, Financial Stress: 0.133207, Work Pressure: 0.001860, Job Satisfaction: 0.000989, Intercept: 0.585526. Positive coefficient values indicating that when those positive coefficient values increase it also increasing the depression level for students.



Correlation Heatmap:-

A correlation heatmap revealed that Financial stress (0.47), Academic pressure (0.47) having strong corelation with depression.



Conclusion: -

We found based on above analysis that fewer sleep duration, academic pressure and financial stress are the significant factors in students depression. Clustering effectively able to groups students with similar stress and satisfaction levels and Linear regression fitting provided predictive insights and emphasizing the importance of balanced workloads and academic support systems as per dataset.