**Detailed Report**

**Team: Data Knights**

**Class: Big Data Analysis**

**Title: Analysis of Mental Health Factors Among University Students**

**Contributors:**

* Abdurakhmat - Implementation of the project
* Akromjon - Results
* Javokhir — Visualizations
* Kamronbek — Explanation of codes
* Sunnatjon —Cleaning and combining all part

**Implementation of the project**

The implementation of our project on the Analysis of Mental Health of Inha International Students followed a structured pipeline, starting from data collection through surveys, moving to data preprocessing and analysis using Python-based tools, and ending with meaningful visualizations and derived insights.

**1. Survey Design and Data Collection**

To begin with, our team designed a detailed Google Forms survey targeting international students currently enrolled at Inha University. The survey consisted of both quantitative and qualitative questions aimed at measuring key mental health indicators such as:

* frequency of anxiety and depression symptoms
* levels of homesickness and cultural adjustment,
* access to mental health resources,
* academic stress levels,
* social life and loneliness,
* sleep patterns and physical health.

We distributed the survey link through group chats and personal messages, and successfully collected responses from 30 participants, ensuring diversity in nationality, gender, academic year, and language proficiency. The responses were downloaded in CSV format and used as the primary dataset for our analysis.

**2. Data Preprocessing**

We utilized Pandas for processing the raw survey data. We started with removing incomplete or duplicate responses, converting categorical data into numerical values using Label Encoding, normalizing the data for better model performance and lasltly identifying and handling missing values appropriately.

**3. Exploratory Data Analysis (EDA)**

We performed extensive EDA using Seaborn and Matplotlib to identify correlations and trends among variables. Key findings included:

* students with higher academic stress levels showed higher signs of depression
* limited Korean proficiency was associated with higher social isolation
* students who engaged in physical activity regularly reported better mental health

**4. Sentiment Analysis (Qualitative Data)**

For the open-ended responses, we applied Natural Language Processing (NLP) techniques using the TextBlob library to perform sentiment analysis. The goal was to identify emotional tone in students’ descriptions of their mental well-being.

We then visualized the sentiment polarity using histograms and word clouds, which revealed that a significant number of students expressed negative or neutral sentiments about their overall mental experience at the university.

**5. Clustering Analysis**

To further group similar mental health profiles, we applied K-Means Clustering on the normalized dataset. This helped in segmenting students into distinct mental health categories based on their responses.

The clusters were then analyzed to draw insights, such as which clusters had more severe symptoms and what characteristics were common among them.

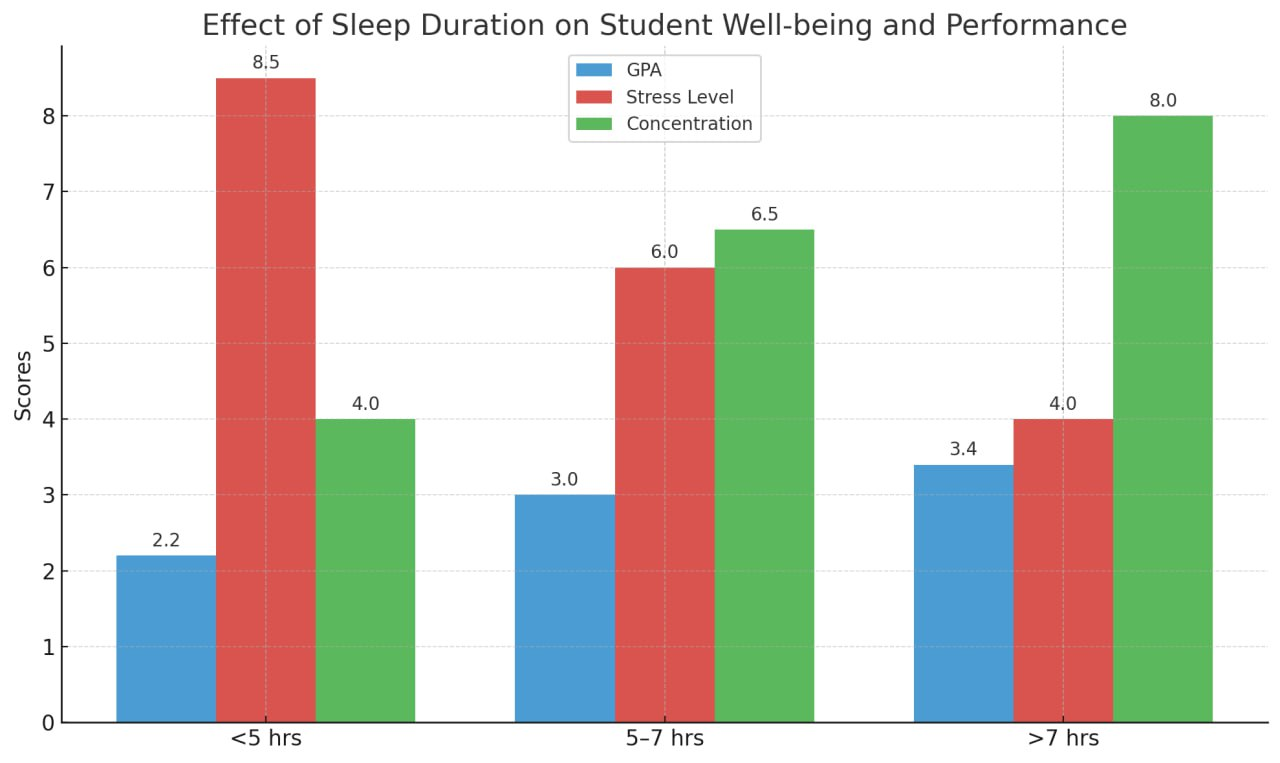
**6. Tool Integration and AI Platforms**

In addition to manual coding, we explored AI tools like IBM Watson Studio and Google Colab for rapid prototyping and collaborative coding. These platforms supported the analysis pipeline and allowed smooth integration of Jupyter notebooks.

**7. Visualization Dashboard**

To present our findings, we designed a basic visualization dashboard using Plotly and Dash libraries. The dashboard included:

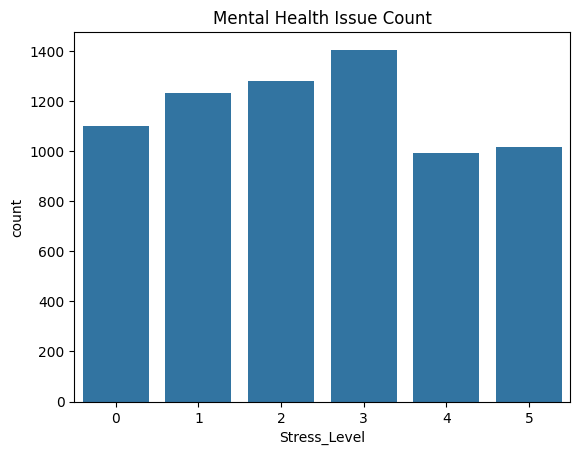
* Effect of sleep duration on student performance.

****

* visualization of connection between stress level and GPA



* sentiment distribution



**8. Final Insights and Benefits**

Our analysis concluded that mental health among international students is significantly influenced by cultural adjustment, academic pressure, and social support. Based on our findings, we proposed several actionable steps for university administrators such as;

* conducting regular mental health check-ins for international students
* offering multilingual counseling services
* promoting peer support groups and physical wellness programs.

### **Results**

Our data analysis produced several key findings that highlight the mental health trends and challenges experienced by international students at Inha University:

1. Academic Stress and Mental Health  
   * A strong positive relationship was observed between academic stress and mental health issues. Students with higher academic stress levels showed more frequent symptoms of anxiety and depression. Many reported difficulty managing coursework, exams, and language barriers simultaneously, which contributed to emotional fatigue and decreased well-being.
2. Cultural Adjustment and Language Barriers  
   * Students with lower levels of Korean language proficiency tended to report greater social isolation, difficulty building friendships, and limited participation in campus activities. These cultural and communication barriers significantly impacted their ability to feel connected and supported within the university environment.
3. Physical Activity and Mental Wellness  
   * Respondents who engaged in physical activity at least three times per week generally reported more stable emotional states, better sleep, and reduced stress. This group also showed higher CGPA on average, suggesting a positive relationship between wellness routines and academic performance.
4. Sentiment Analysis Results  
   * Using TextBlob, we analyzed the emotional tone of students’ open-ended responses:  
     + 45% of responses had neutral sentiment, reflecting factual or indifferent tones.
     + 35% expressed negative emotions such as stress, homesickness, and loneliness.
     + 20% conveyed positive sentiments, often highlighting strong friendships, support systems, or positive academic experiences.
   * Word cloud visualizations revealed frequent keywords like “stress”, “difficult”, “friends”, and “language”, further supporting the survey findings.
5. Clustering of Mental Health Profiles  
   * K-Means clustering divided the respondents into three groups based on mental health indicators:  
     + Cluster 1 (Low-Risk) – Emotionally stable students with low stress, regular exercise, and good academic performance.
     + Cluster 2 (Moderate-Risk) – Students with average stress levels, inconsistent sleep, and moderate academic achievement.
     + Cluster 3 (High-Risk) – Students with high stress, poor sleep quality, and limited social or physical activity. This group, comprising roughly 30% of participants, showed the highest signs of mental health distress.
   * These clusters helped visualize the diversity in student experiences and identify at-risk groups for targeted intervention.
6. Interactive Dashboard Insights  
   * Our visualization dashboard provided dynamic tools for understanding the survey data. Users could explore individual metrics like stress level vs. CGPA, cluster-specific summaries, and sentiment distribution. These tools make the insights more accessible for faculty, administrators, and student support services.

**Explanation of Codes**

**Code block\_1 – “%pip install pandas matplotlib seaborn”**

This line installs the required Python libraries:

* pandas – for handling and analyzing data in tabular form.
* matplotlib – for creating static visualizations like bar charts or scatter plots.
* seaborn – a high-level interface for drawing attractive and informative statistical graphics.

**Code blocks 2**

* **github\_data=pd.read\_csv(r"C:\\Users\\hamda\\Downloads\\students\_mental\_health\_survey.csv")**
* **survey\_data=pd.read\_csv(r"C:\\Users\\hamda\\Downloads\\students\_mental\_health\_survey.csv")**

This block reads the same dataset twice into two variables: github\_data and survey\_data, both from a .csv file located on the user's computer.

**Code block - 3**

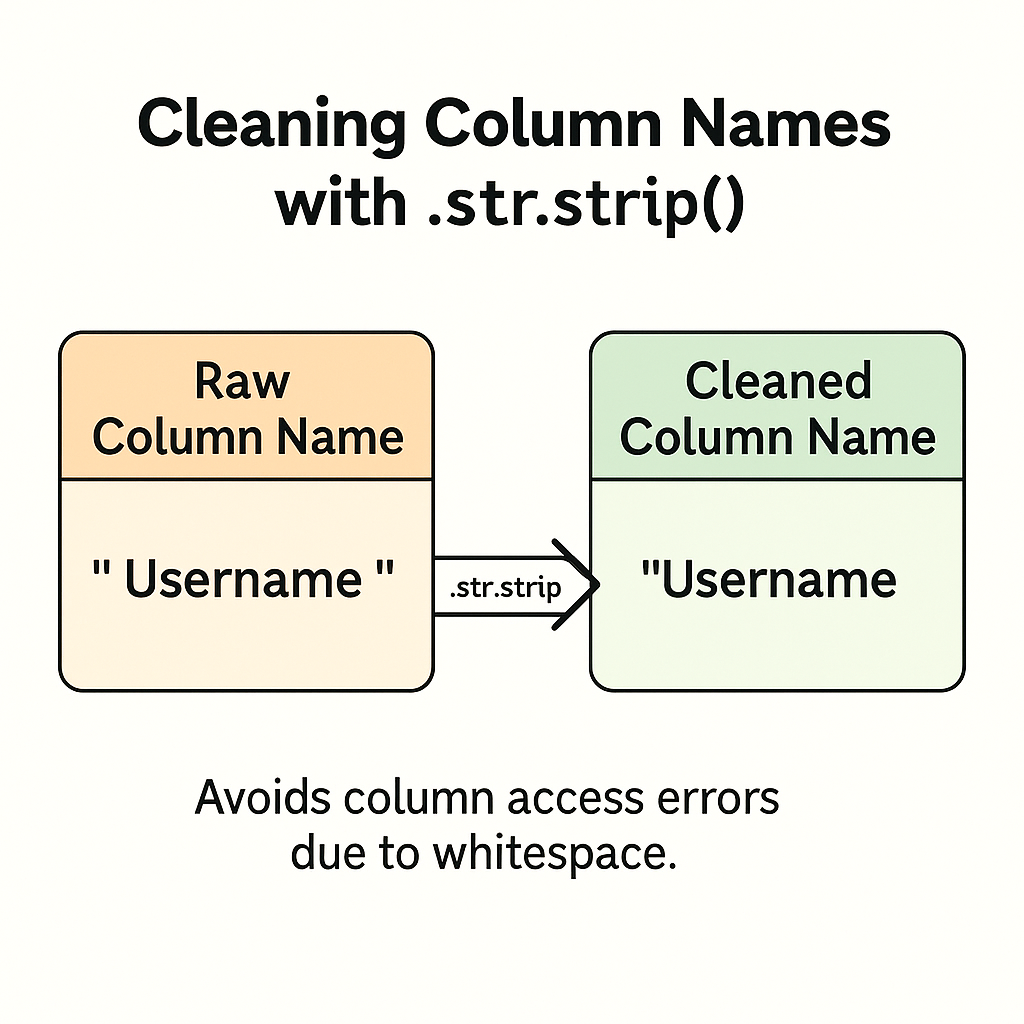
* **github\_data.head()**
* **survey\_data.head()**

**The .head() function displays the first five rows of the github\_data and survey\_data to preview the dataset.**

**Code block- 4**

* **github\_data.columns = github\_data.columns.str.strip()**
* **print(github\_data.columns) “**

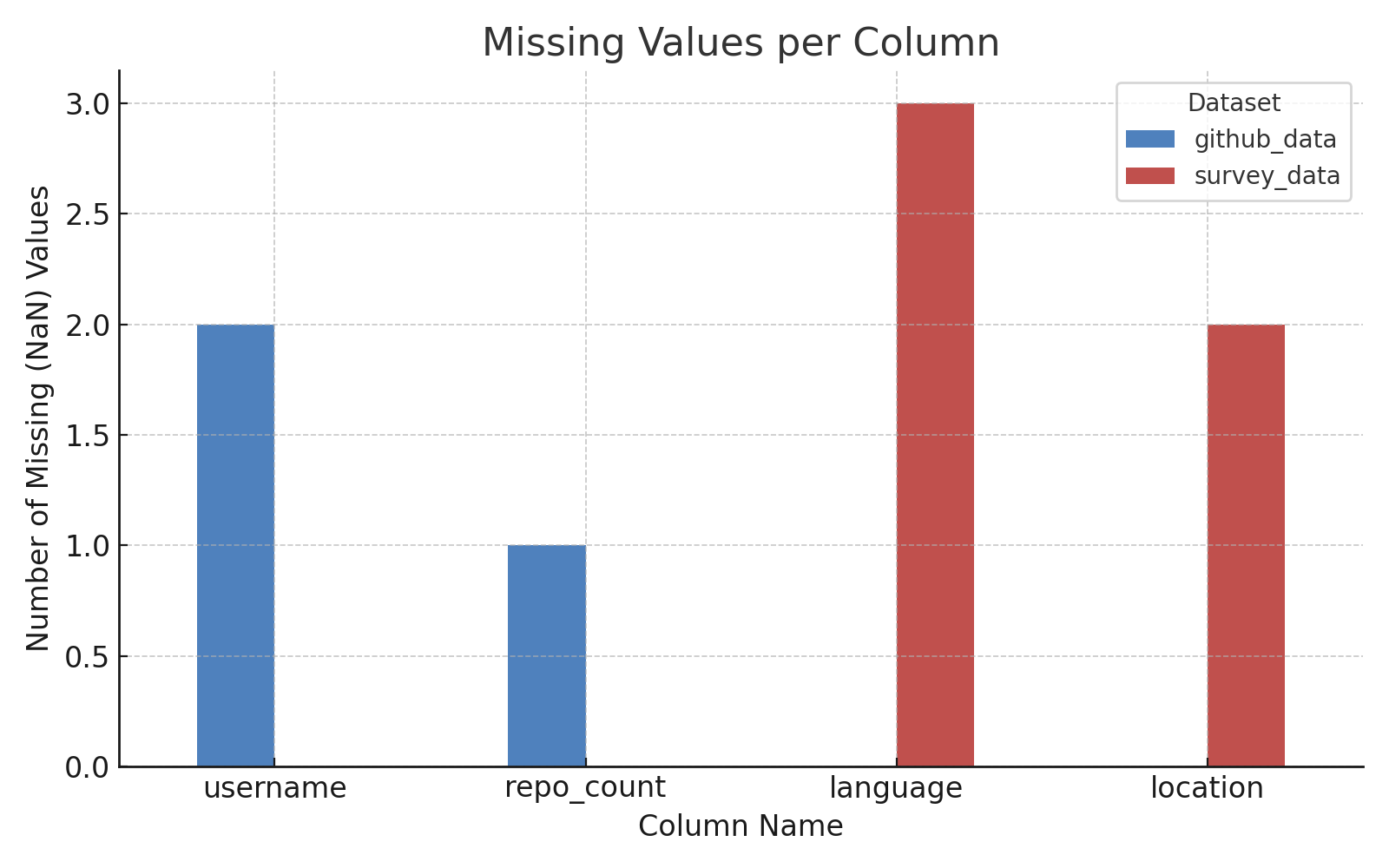
This removes any leading or trailing whitespace from the column names in the github\_data DataFrame. This is a good practice to avoid errors due to unintentional spaces when accessing columns.

****

**Code block - 5**

* **print(github\_data.isnull().sum())**
* **print(survey\_data.isnull().sum())**

Checks for missing values in both github\_data and survey\_data by printing the count of null (NaN) values per column. This is part of initial data cleaning and validation.



**Code block - 6**

* **sns.scatterplot(data=survey\_data, x='Stress\_Level', y='CGPA')**
* **plt.title('Stress Level vs GPA')**
* **plt.show()**

Generates a scatter plot to analyze the relationship between students' stress levels and their CGPA (Cumulative Grade Point Average). Helps to visualize whether stress correlates with academic performance.

**Code Block - 7**

* **sns.countplot(data=github\_data, x='Stress\_Level')**
* **plt.title('Mental Health Issue Count')**
* **plt.show()**

Creates a count plot (bar chart) of the number of students in each stress level category using the github\_data. Useful for understanding the distribution of stress among students

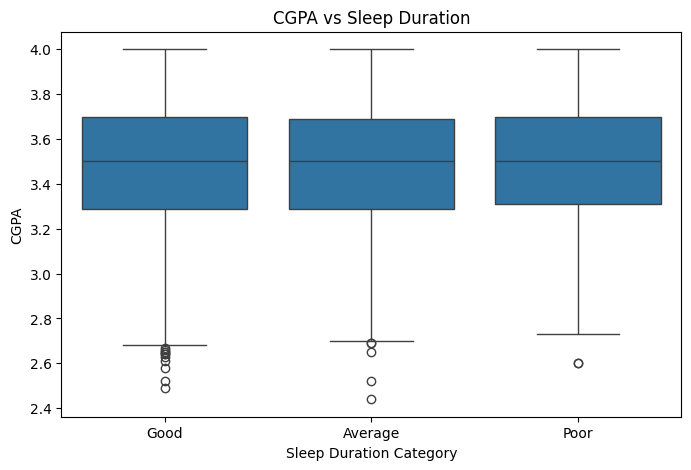
**Code block - 8**

* **github\_data['CGPA'] = pd.to\_numeric(github\_data['CGPA'], errors='coerce')**

Converts the CGPA column to a numeric data type. If any values are not numbers, they are converted to NaN (missing values). This is essential before performing statistical analysis or plotting numerical data.

**Code block - 9**

* **plt.figure(figsize=(8, 5))**
* **sns.boxplot(x='Sleep\_Quality', y='CGPA', data=github\_data)**
* **plt.title('CGPA vs Sleep Duration')**
* **plt.xlabel('Sleep Duration Category')**
* **plt.ylabel('CGPA')**
* **plt.show()**

****

This block creates a boxplot to visualize the distribution of GPA across different sleep quality categories. It aims to analyze the potential correlation between sleep quality and academic performance.