■ Data Science Project: Mental Health and Productivity

This project explores how mental health factors such as stress levels, sleep duration, and risk levels affect productivity and depression scores using Python and data science techniques.

1■■ Importing Libraries

You start by importing important tools:

- pandas and numpy for handling and analyzing data
- matplotlib and seaborn for creating graphs
- scikit-learn for building and testing machine learning models

2■■ Loading the Dataset

You load a dataset called 'mental_health_dataset.csv' that contains information about people's mental health, stress, sleep, and productivity. Then you display the first 5 rows to get a quick look at the data.

3■■ Checking Data Quality

You check if the data is clean and ready for analysis. This includes:

- Checking for missing values
- Finding duplicate rows
- Viewing basic statistics like mean, median, and standard deviation
- Checking data types and unique categories in text columns

4■■ Visualizing the Data

You create graphs to understand trends and relationships in the data:

- Distribution of Productivity Scores – shows how productivity is spread
- Correlation Heatmap – shows how features like stress or sleep relate
- Stress Level vs Productivity – scatter plot showing relationship
- Productivity by Mental Health Risk – bar chart comparison
- Sleep Hours vs Productivity – line plot to see sleep effects

5■■ Preparing Data for Machine Learning

You prepare the data so it can be used for training a model:

- Encode text data into numbers using LabelEncoder
- Set 'depression_score' as the target variable to predict
- Split the data into input features (X) and target output (y)
- Scale all numeric features to the same range
- Split data into 80% training and 20% testing

6■■ Building and Training the Model

You use a Linear Regression model to learn the relationship between input features (like stress, sleep, productivity) and the depression score.

7■■ Testing and Evaluating the Model

The model's performance is tested on new data using:

- Mean Squared Error (MSE): measures how far predictions are from real values
- R^2 Score: shows how well the model explains the data (1.0 means perfect fit)
- Scatter plot comparing actual vs predicted depression scores

8■■ Final Output and Insights

Finally, you view model accuracy and insights about how mental health factors like stress, sleep, and risk affect productivity and depression. You also visualize how well your predictions match actual scores.

■ Conclusion

This project shows how data science can be used to understand mental health and productivity. By cleaning data, visualizing relationships, preparing features, and training a regression model, you successfully built a system that predicts depression scores and provides valuable mental health insights.