Mental Health Dataset

Feature Engineering Questions:

1. **Missing Data Imputation**
    - Task: Identify any missing values in the dataset. Discuss and implement a strategy to handle missing data (e.g., imputation with mean, median, or mode) for the columns Age, Stress_Level, and Physical_Activity_Hours.
2. **Categorical Variable Encoding**
    - Task: Encode the categorical variables Gender, Occupation, Country, and Mental_Health_Condition using suitable encoding techniques (e.g., One-Hot Encoding, Label Encoding). Discuss when to use each method.
3. **Creating Interaction Features**
    - Task: Create an interaction feature between Stress_Level and Work_Hours to study the potential combined effect on Mental_Health_Condition.
4. **Feature Scaling**
    - Task: Standardize the numerical features (Age, Stress_Level, Sleep_Hours, Work_Hours, Physical_Activity_Hours) using a suitable scaling method (e.g., Min-Max Scaling or Z-score Standardization). Discuss why scaling is important.
5. **Binning Continuous Variables**
    - Task: Bin the Age column into age groups (e.g., 18-30, 31-45, 46-60, 60+). Discuss the rationale behind binning and how it might affect the analysis.
6. **Handling Outliers**
    - Task: Identify outliers in the Work_Hours and Physical_Activity_Hours columns. Discuss and implement a method to handle outliers (e.g., capping, removal).
7. **Feature Transformation**
    - Task: Apply a log transformation to Work_Hours to reduce skewness and improve normality. Discuss when and why such transformations might be useful.
8. **Aggregation of Features**

- Task: Create a new feature called Total Activity by summing Physical_Activity_Hours and Work_Hours. Discuss how this new feature might provide additional insights.

9. **Derived Features Based on Domain Knowledge**
    - Task: Create a new feature Sleep vs Work Ratio by dividing Sleep_Hours by Work_Hours. Discuss how this ratio could be a useful predictor of mental health outcomes.
10. **Outlier Detection**
    - Task: Use the IQR method to identify outliers in Stress_Level and decide whether to remove or transform these outliers. Discuss the impact of outliers on the analysis.

Hypothesis Test Questions:

T-Test Questions:

1. **T-Test for Mean Difference in Stress Levels Across Genders**
    - Task: Perform an independent two-sample t-test to test if there is a significant difference in the average Stress_Level between Male and Female participants. State the null and alternative hypotheses, and interpret the p-value.
2. **T-Test for Comparison of Stress Levels by Consultation History**
    - Task: Use a t-test to compare the mean Stress_Level between individuals who have consulted a mental health professional (Consultation_History = 'Yes') and those who haven't (Consultation_History = 'No'). Discuss the results and provide conclusions.

ANOVA Questions:

3. **ANOVA for Stress Level Across Occupations**
    - Task: Perform a one-way ANOVA to examine if there are significant differences in Stress_Level across the different Occupation groups (e.g., Teacher, Engineer, Doctor, etc.). State the null and alternative hypotheses, and interpret the results.
4. **ANOVA for Mental Health Condition Severity by Age Groups**
    - Task: Perform a one-way ANOVA to determine if Severity (of mental health condition) differs across different Age groups (e.g., 18-30, 31-45, 46-60, 60+). Discuss the results and interpret the p-value.

Chi-Square Test Questions:

5. **Chi-Square Test for Association Between Gender and Mental Health Condition**
    - Task: Perform a chi-square test of independence to determine if there is a significant association between Gender and Mental_Health_Condition. State the null and alternative hypotheses, and discuss the findings.
6. **Chi-Square Test for Consultation History and Physical Activity**
    - Task: Use a chi-square test to assess whether there is an association between Consultation_History (Yes/No) and Physical_Activity_Hours categories (e.g., Low, Medium, High). Interpret the chi-square statistic and p-value.
7. **Chi-Square Test for Association Between Country and Occupation**
    - Task: Perform a chi-square test to evaluate if Country and Occupation are independent. Discuss the assumptions of the test and interpret the results.