Project Part 2

Describe (in a maximum of 2 *new* pages) the project that you proposed in part 1 (200 points)

1. **Include a short description outlining the project that you proposed in part 1.**
   This research study investigates the relationship between habits (such as caffeine intake, frequency of physical activity, and sleep hours) and university students' productivity. Using a correlational study design with a target sample of 30-50 students, anonymous surveys will collect data on variables like caffeine consumption patterns, exercise frequency, sleep duration, and daily productivity ratings (on a 0-100 scale). The original proposal included tracking over time. However, based on feedback, the study will focus on data collection "on a given day" rather than weekly measurements due to survey anonymity requirements. The analysis will primarily use multiple regression and correlation analyses to examine relationships between these variables.

2. **Describe the dataset you will have. What are the variables? Which are IVs, and which are DVs? Within IVs and DVs, which are categorical and which are continuous?**
   The dataset for this research study will consist of variables related to university students' habits and productivity levels. The focus will be on understanding how variables like hours of sleep, frequency of physical activity, and caffeine consumption influence daily productivity. Below is a detailed description of the variables included in the dataset, categorized into independent variables (IVs) and dependent variables (DVs).

   Independent Variables (IVs):

   1. **Continuous Variables:**
      a. Caffeine consumption frequency:
         Participants will track the exact number of servings they consume each day.
      b. Caffeine consumption timing:
         Participants will track caffeine timing in hours.
      c. Physical activity frequency in Hours:
         This variable measures the frequency of Activity, measured as the total hours of exercise per week.
      d. Physical activity timing:
         Like caffeine consumption timing, this variable captures physical activity timing in hours after activity for a continuous scale.
      e. Energy level Rating:
         Energy Level Variations will be measured on a continuous scale from 0 to 100.
      f. Sleep schedule consistency:
         This variable assesses how consistent participants are with their sleep schedules; it tracks the number of days per week that the sleep schedule is consistent.
      g. Hours of sleep per night:
         This variable measures the average hours of sleep participants get each night.
   2. **Categorical Variables:**

a. **None** (As we are working only with Multiple Regression Analysis and Pearson Correlation Analysis)

Dependent Variables (DVs):

1. **Continuous Variables:**
   a. Daily productivity score:
      Participants will rate their overall productivity for the day on a continuous scale from 0 to 100.
   b. Time-specific productivity levels:
      Measures productivity at different times (morning, afternoon, evening) on a continuous scale from 0 to 100.
   c. Productivity Consistency:
      Record productivity consistency as the number of days the productivity level is steady
2. **Categorical Variables:**
   a. **None** (As we are working only with Multiple Regression Analysis and Pearson Correlation Analysis)

3. **Describe the data pre-processing steps that you believe you will need to do. What data cleaning? Will you impute values from missing data? If so, how? What other pre-processing might you need to do, if any?**
Several data pre-processing steps will be necessary to ensure the dataset is clean and ready for analysis. These steps are critical to ensure the integrity and quality of the data, ultimately affecting the validity of the results.

1. **Data Cleaning**
   Data cleaning involves identifying and correcting errors or inconsistencies in the dataset. For this study, potential issues might include:
   - **Outliers:** Check for outliers in continuous variables such as hours of sleep or physical activity frequency. Outliers could result from participants entering incorrect or exaggerated values (e.g., reporting 20 hours of sleep per night or 15 days of physical activity in a week). These values should be flagged and either corrected (if possible) or removed if they are erroneous.
   - **Inconsistent Responses:** Ensure that all responses align with the continuous nature of the variables. For example, if a participant provides a categorical response for caffeine consumption (e.g., "Never"), this needs to be converted into a continuous measure (e.g., 0 cups of caffeine per day).
   - **Duplicate Entries:** Check for duplicate responses from participants, ensuring no participant has submitted multiple surveys.

2. **Handling Missing Data**
   Missing data is common in survey-based studies. It is important to decide how to handle missing values to avoid bias in the analysis. There are several strategies for dealing with missing data:
   - **Imputation:** If there are missing values for continuous variables such as hours of sleep or physical activity frequency, imputation can be used. A common approach is:
     - **Mean Imputation:** Replace missing values with the mean value of that variable across all participants. This method works well when the amount of missing data is small and when the data is normally distributed.

- **Median Imputation:** If the data is skewed (e.g., sleep hours), using the median instead of the mean may be more appropriate.
- **Mode Imputation:** For categorical variables like caffeine consumption timing or physical activity timing, missing values can be replaced with the most frequent category (mode).

- **Listwise Deletion:** If a participant has too many missing responses (e.g., more than 50% of their survey is incomplete), their entire response may be removed from the dataset. This method should be used sparingly to avoid reducing the sample size too much.

3. **Bucketing and Transformation Adjustments**
   a. **Bucketing:**
      i. **Caffeine Intake:** To simplify analysis and capture usage patterns, caffeine intake can be bucketed into categories such as "Low," "Moderate," and "High," based on daily consumption frequency. These buckets will help in understanding productivity correlations for different intake levels.
      ii. **Sleep Duration:** Bucket hours of sleep into ranges, such as "Short" (0-4 hours), "Moderate" (5-7 hours), and "Optimal" (8+ hours), to analyze sleep's impact on productivity more distinctly.
      iii. **Physical Activity Frequency:** Physical activity could be categorized into "Sedentary," "Moderate," and "Active" based on weekly exercise hours. These buckets provide an easier way to examine productivity variations related to activity levels.
   b. **Transformation:**
      i. **Scaling:** Normalize continuous variables like hours of sleep, caffeine intake, and physical activity using min-max scaling. This approach brings all variables to a common scale (0-1), ensuring regression and correlation analysis consistency.
      ii. **Log Transformation:** If data for caffeine intake frequency or productivity is skewed, a log transformation can be applied to reduce skewness and stabilize variance, which enhances the model's robustness.
   c. **Handling Outliers:**
      i. **Bucketing as an Outlier Solution:** For variables with extreme values, such as high caffeine consumption, bucketing also contains outliers within a specific range.

4. **Converting Categorical Variables to Continuous**

   Since both Multiple Regression and Pearson Correlation require continuous variables, any categorical variables need to be transformed into continuous measures:

   - **Caffeine Intake:** Instead of using categories like "Never," "Rarely," or "Daily," caffeine intake should be recorded as a continuous variable based on the number of caffeinated beverages consumed per day (e.g., cups of coffee, tea, energy drinks). If participants have provided categorical responses, these can be converted into approximate daily consumption values (e.g., "Never" = 0 cups/day, "Rarely" = 0.5 cups/day, etc.).
   - **Physical Activity Frequency:** Convert physical activity frequency into total hours per week rather than days per week. This provides a more precise measure of physical activity that can be treated as continuous.

- **Sleep Consistency:** If sleep schedule consistency was originally measured on an ordinal scale (e.g., "Very Consistent" to "Very Inconsistent"), this needs to be transformed into a continuous metric. For instance, participants could report how many days per week they maintain a consistent sleep schedule (0–7 days).

5. **Normalization/Scaling**

   For continuous variables such as hours of sleep and physical activity frequency, normalizing or scaling the data may be necessary before conducting regression analysis, especially if these variables have widely different ranges.
   - **Min-Max Scaling:** This technique scales all continuous variables to a range between 0 and 1, ensuring that no variable disproportionately influences the results due to its scale.
   - **Z-score Normalization:** This method standardizes continuous variables by subtracting the mean and dividing by the standard deviation, resulting in a distribution with a mean of 0 and a standard deviation of 1.

   Normalization is particularly important when running algorithms like regression that assume all predictors are on comparable scales.

6. **Addressing Survey Response Bias**

   Since this study relies on self-reported data, there may be biases such as social desirability bias or recall bias. While these biases cannot be eliminated through pre-processing, it's important to:
   - **Check for Response Patterns:** Identify any participants who may have given identical responses across all questions (e.g., selecting "Neutral" for every item), which could indicate careless responding.
   - **Cross-check Responses:** For questions that should logically align (e.g., caffeine intake frequency and timing), cross-check responses for consistency.

Conclusion

In summary, the pre-processing steps will include:

1. Cleaning data by addressing outliers, inconsistencies, and duplicates.
2. Handling missing data through imputation or listwise deletion.
3. Bucketing and Transformation Adjustments
4. Converting categorical variables into continuous measures where necessary.
5. Normalizing/scaling continuous variables if necessary.
6. Checking for response patterns to mitigate survey biases.

These steps will ensure that the dataset is clean and ready for robust statistical analysis, allowing us to accurately explore relationships between habits like caffeine intake, physical activity, sleep quality, and productivity among university students.

4. **State what kind of analysis you intend to do, and what your hypothesis is for every "test statistic" that you will generate (ie a 2 way ANOVA has a test statistic for each of the two main effects, and one for the interaction, so generate a hypothesis for each).**

We intend to do Multiple Regression Analysis and Pearson Correlation Analysis

Hypotheses for Test Statistics

With the help of t-tests, we will isolate each of the relations in the regression model, determining the strength of each of the independent variables about the dependent one, as well as employing multiple regression analysis and Pearson correlation analysis to determine how three independent variables influence the dependent variable.

Multiple Regression Analysis and T-tests Hypotheses

Multiple regression analysis shall be carried out assisted with t-tests for each of the independent variables in order to test whether or not each of them is significantly related to productivity provided that other variables are already controlled.

1. Caffeine Intake (IV1)
    - Null Hypothesis (H0): under this hypothesis, the regression coefficient for caffeine intake β1 is equal to zero which means that caffeine intake does not significantly contribute towards enhancing daily productivity in relation to physical activity and sleep quality.
        - H0:β1=0
    - Alternative Hypothesis (H1): under this hypothesis, the regression coefficient for caffeine intake β1 is not equal to zero which means that caffeine intake significantly contributes towards augmenting one's productivity levels on a daily basis in relation to physical activity and sleep quality.
        - H1:β1!=0
2. Physical Activity (IV2)
    - Null Hypothesis (H0): the regression coefficient for physical activity, β2, is equal to a zero which implies that physical activity is not viewed as an important determinant of daily productivity level after the caffeine dosage and the quality of sleep has been considered.
        - H0:β2=0
    - Alternative Hypothesis (H1): the regression coefficient for physical activity, β2, is not equal to a zero and this implies that physical activity is practiced on all days as a contributor to the daily productivity level even with the consideration of the caffeine dosage and the quality of sleep.
        - H1:β2!=0
3. Sleep Quality (IV3)

- Null Hypothesis (H0): Sleep quality has a coefficient, β3, that is equal to zero. This means that when we factor in the caffeine dosage and physical activity, sleep quality is not one of the key factors explaining the daily productivity level.
    - H0:β3=0
- Alternative Hypothesis (H1): The regression coefficient for sleep quality (β3)is not equal to zero, meaning sleep quality has a significant effect on daily productivity after accounting for caffeine intake and physical activity.
    - H1:β3!=0

Pearson Correlation Analysis Hypotheses

We will also conduct a Pearson correlation analysis to assess the relationship between each independent variable (IV) and the dependent variable (DV).

1. Correlation between Caffeine Intake and Productivity
    - Null Hypothesis (H0): There is no significant correlation between caffeine intake and productivity.
        - H0:r=0
    - Alternative Hypothesis (H1): There is a significant correlation (either positive or negative) between caffeine intake and productivity.
        - H1:r!=0
2. Correlation between Physical Activity and Productivity
    - Null Hypothesis (H0): There is no significant correlation between physical activity and productivity.
        - H0:r=0
    - Alternative Hypothesis (H1): There is a significant correlation between physical activity and productivity (either positive or negative).
        - H1:r!=0
3. Correlation between Sleep Quality and Productivity
    - Null Hypothesis (H0): There is no significant correlation between sleep quality and productivity.
        - H0:r=0
    - Alternative Hypothesis (H1): There is a significant correlation between sleep quality and productivity (either positive or negative).
        - H1:r!=0

Summary of Analysis Approach:

- We will conduct multiple regression analyses to understand the combined influence of caffeine intake, physical activity, and sleep quality on productivity.
- For each independent variable in the regression model, we will perform a t-test to determine if the regression coefficient (β) is significantly different from zero, meaning that the variable significantly impacts productivity after controlling for other variables.

- We will also conduct Pearson correlation analysis to determine the strength and direction of the relationships between each IV (caffeine intake, physical activity, sleep quality) and DV (productivity).
    - The Null hypothesis for correlation analysis (H0:r=0)implies no relationship.
    - An Alternative hypothesis for correlation analysis (H1:r!=0) implies a significant relationship.

By using multiple regression analysis and Pearson correlation, we will be able to determine each variable's unique contribution to productivity and the overall strength and direction of the relationship between these lifestyle habits and productivity levels.

5. **Include as an appendix the entire assignment from part 1, which you can update based on comments you've received and new understanding, making sure you highlight all changes in red font. This doesn't count towards your page "limit"… also if you don't have any changes you can make, then put a note in red at the top of the appendix indicating that**
    1. **Sketch out the plan for the user study that you will conduct this term, including details such as:**
        a. **What variables are you going to collect?**
            i. **Caffeine Intake:** For **Caffeine Intake**, the variables collected will include **Frequency of Consumption**, which will be measured on a continuous scale based on the **number of caffeinated beverages** consumed daily, such as cups of coffee, tea, or energy drinks. Participants will track the exact number of servings they consume each day. The **Timing of Consumption** will be recorded as the number of **hours after waking up** when caffeine is consumed. This replaces categorical time blocks like "Morning" or "Afternoon" and allows for more granular data. Participants will log the exact time (e.g., 1 hour after waking, 3 hours after waking) at which they consume caffeine.
            ii. **Physical Activity:** For **Physical Activity**, the key variables include **Frequency of Activity**, measured as the total **hours of exercise per week**, turning the data into a continuous variable. Participants will log the total time spent engaging in physical activity each week, allowing for a more accurate measurement than simply counting days of activity. The **Timing of Activity** will be recorded as the **number of hours after waking** when participants begin their physical activity session, providing precise data on when exercise is most likely to occur. Finally, **Energy Level Variations** will be measured on a continuous scale from 0 to 100, with participants rating their energy levels at different points of the day (e.g., morning, afternoon, evening), creating a continuous variable that can show fluctuations in energy throughout the day.
            iii. **Sleep Quality:** For **Sleep Quality**, the variable **Hours of Sleep per Night** will be recorded as a continuous measure of the average hours of sleep participants get each night. This data will be collected using a **sleep tracker** or manual logs to ensure precision. **Consistency of Sleep Schedule** will be measured on a continuous scale, focusing on the **regularity** of a participant's sleep schedule. Instead of categorical terms like "Consistent" or "Inconsistent," participants will track the **number of days** they maintain a consistent sleep schedule throughout a week, giving a more nuanced view of their sleep habits.
            iv. **Daily Productivity:** For **Daily Productivity**, the variable **Self-reported Productivity Rating** will be measured on a continuous scale from 0 to 100, where 0 indicates no productivity and 100

represents maximum productivity. Participants will self-assess their overall productivity at the end of each day. The **Productivity Consistency** will be recorded as the number of days a participant maintains **consistent productivity levels** throughout the week, rather than as a simple categorization of consistency. Finally, **Variation in Productivity** will be assessed by tracking productivity at different times of day (morning, afternoon, and evening) on a continuous scale from 0 to 100. Participants will rate their productivity levels throughout the day, providing detailed insight into how productivity fluctuates during various periods.

b. What design is your study (experimental vs. correlational; if experimental, what factors are between subjects vs. within subjects)?

The study will use a **correlational design** to examine the relationships between variables such as caffeine intake, physical activity, sleep quality, and daily productivity.

Justification:

i. It's correlational as there is no active manipulation of variables, and does not involve random assignment. Also, the participants' daily routine is observed as it occurs naturally.

ii. Our study examines the relationship between daily habits (caffeine intake, physical activity, sleep quality) and productivity.

iii. We aim to identify trends and associations between these variables without manipulating them. Participants will self-report their habits and productivity over time, and we will analyze naturally occurring variations to see if they correlate.

c. Given those answers, out of those variables which are your IV(s) and DV(s)?

i. **Independent Variables:**

1. **Caffeine Intake:** For **Caffeine Intake**, the variables collected will include **Frequency of Consumption**, which will be measured on a continuous scale based on the **number of caffeinated beverages** consumed daily, such as cups of coffee, tea, or energy drinks. Participants will track the exact number of servings they consume each day. The **Timing of Consumption** will be recorded as the number of **hours after waking up** when caffeine is consumed. This replaces categorical time blocks like "Morning" or "Afternoon" and allows for more granular data. Participants will log the exact time (e.g., 1 hour after waking, 3 hours after waking) at which they consume caffeine.
2. **Physical Activity:** For **Physical Activity**, the key variables include **Frequency of Activity**, measured as the total **hours of exercise per week**, turning the data into a continuous variable. Participants will log the total time spent engaging in physical activity each week, allowing for a more accurate measurement than simply counting days of activity. The **Timing of Activity** will be recorded as the **number of hours after waking** when participants begin their physical activity session, providing precise data on when exercise is most likely to occur. Finally, **Energy Level Variations** will be measured on a continuous scale from 0 to 100, with participants rating their energy levels at different points of the day (e.g., morning, afternoon, evening), creating a continuous variable that can show fluctuations in energy throughout the day.

3. **Sleep Quality:** For **Sleep Quality**, the variable **Hours of Sleep per Night** will be recorded as a continuous measure of the average hours of sleep participants get each night. This data will be collected using a **sleep tracker** or manual logs to ensure precision. **Consistency of Sleep Schedule** will be measured on a continuous scale, focusing on the **regularity** of a participant's sleep schedule. Instead of categorical terms like "Consistent" or "Inconsistent," participants will track the **number of days** they maintain a consistent sleep schedule throughout a week, giving a more nuanced view of their sleep habits.

ii. **Dependent Variables:**

1. **Daily Productivity:** For **Daily Productivity**, the variable **Self-reported Productivity Rating** will be measured on a continuous scale from 0 to 100, where 0 indicates no productivity and 100 represents maximum productivity. Participants will self-assess their overall productivity at the end of each day. The **Productivity Consistency** will be recorded as the number of days a participant maintains **consistent productivity levels** throughout the week, rather than as a simple categorization of consistency. Finally, **Variation in Productivity** will be assessed by tracking productivity at different times of day (morning, afternoon, and evening) on a continuous scale from 0 to 100. Participants will rate their productivity levels throughout the day, providing detailed insight into how productivity fluctuates during various periods.

d. What are the operational definitions going to be for your IV(s) and DV(s)? (ie how are you going to measure or manipulate the variables)?

Our study is **correlational**, and we will use self-reported data collected through surveys to measure each variable. The operational definitions of both **independent** and **dependent variables** are designed to capture precise, continuous data, providing a detailed view of participants' lifestyle habits and their impact on daily productivity.

Independent Variables (IVs):

1. **Caffeine Consumption**:
   To assess caffeine intake, participants will report the **number of caffeinated beverages** they consume per day, such as coffee, tea, or energy drinks. This will be measured as a continuous variable (e.g., 1, 2, or 3 cups per day). This provides a direct measure of the volume of caffeine consumed. Additionally, participants will record the **timing of caffeine consumption**, measured as the **number of hours after waking** that they consume caffeine (e.g., 1 hour after waking, 3 hours after waking). This allows us to examine the relationship between the timing of caffeine intake and its potential effects on productivity.
2. **Physical Activity**:
   For physical activity, we will measure **frequency** by asking participants to report the **total hours** of exercise they engage in per week (e.g., 3.5 hours per week). This gives a continuous measure of physical activity, capturing both light and vigorous exercise. We will also record the **timing of physical activity**, which will be measured as the **number of hours after waking** that participants begin their exercise sessions (e.g., 2 hours after waking, 5 hours after

waking). This data will help us understand how the timing of physical activity relates to productivity levels throughout the day.

3. **Sleep Quality**:
   For sleep quality, participants will provide the **average number of hours** they sleep per night, measured as a continuous variable (e.g., 6.5 hours per night). This allows us to analyze sleep duration about daily productivity. In addition, we will assess **sleep schedule consistency** by asking participants to report how **regular** their sleep patterns are. Rather than using categories like "consistent" or "inconsistent," we will measure consistency on a scale (e.g., the number of days per week they maintain a consistent sleep schedule). This will provide a more accurate view of sleep patterns and their potential influence on productivity.

Dependent Variables (DVs):

1. **Daily Productivity**:
   For daily productivity, participants will rate their **overall productivity** each day on a continuous scale from 0 to 100, where 0 indicates no productivity and 100 represents maximum productivity. This allows for more precise measurement than traditional Likert scales. We will also measure **productivity consistency** by asking participants how consistent their productivity is daily. This will be tracked as a continuous variable based on the number of days they feel their productivity remains consistent throughout the week (e.g., 0 to 7 days). Finally, to capture variations in productivity, participants will report their productivity at different times of the day (morning, afternoon, and evening), using a continuous scale from 0 to 100. This will allow us to see how productivity fluctuates during the day and how it relates to other lifestyle factors like caffeine consumption and physical activity.

e. What is your population? How are you going to get participants from that population? How many are you planning to recruit for the study?

i. The population for this study consists of university students who are likely to consume caffeine daily, engage in physical activity, and experience varying sleep habits. This demographic is ideal because they often manage their productivity in academic or work environments.

ii. Recruitment will target students from the University of Southern California (USC), where the study is being conducted. We will share the survey through Email, Student organization platforms, and word of mouth.

iii. A sample of 30-50 participants is a reasonable size for detecting patterns in correlational research while maintaining manageable data collection and analysis efforts for our study.

2. Sketch out your plan for analysis:

a. State your research question(s), and discuss how it could be answered by analyzing the data that you listed in the previous question. That is, affirm for me that your research question is answerable using the data you will collect.

i. **Research question:** How do caffeine intake, physical activity, and sleep quality influence self-reported productivity levels among university students on a daily basis?

ii. We will answer this question by analyzing the correlations between our independent and dependent variables (productivity). Our collected survey data will measure specific aspects of each habit and productivity metric, allowing us to examine relationships between these variables. We can determine which factors are strongly associated with productivity outcomes by comparing the relationships between different habits and productivity measures.

iii. For **Data Collection Mapping**, the primary **dependent variables (DVs)** in the survey will include participants' **daily productivity rating**, which will be measured on a continuous scale from 0 to 100, where 0 indicates no productivity and 100 represents maximum productivity. This will provide a precise measure of productivity throughout the day. Additionally, **productivity consistency** will be assessed by asking participants how consistent their productivity is over a week, measured on a scale from 0 to 7 days, where 0 indicates no consistency and 7 indicates perfect consistency. Finally, **time-of-day productivity variations** will be captured by asking participants to rate their productivity at different times of the day (morning, afternoon, and evening) on a scale from 0 to 100. This will provide detailed insights into how productivity fluctuates throughout the day and whether it is influenced by other factors.

b. Describe in your own words what kinds of analysis could be done with the data to answer each question. Be specific about what analysis -within null hypothesis significance testing- you would use and why.

Since this is a correlational study with multiple variables, we will employ several statistical analyses within null hypothesis significance testing:

i. Multiple Regression Analysis - Combined IV Test

1. **Purpose:** To examine the combined influence of caffeine intake, physical activity, and sleep quality on productivity ratings.
2. **Hypothesis:**
    1. **Caffeine Intake:**
        a. Null Hypothesis (H0): There is no unique effect of caffeine intake (β1) controlling for Physical activity & Sleep quality.
        b. Alternative Hypothesis (H1): There is a unique effect of caffeine intake (β1) controlling for Physical activity & Sleep quality.
    2. **Physical Activity:**
        a. Null Hypothesis (H0): There is no unique effect of Physical Activity (β2) controlling for Caffeine Intake & Sleep quality.
        b. Alternative Hypothesis (H1): There is a unique effect of Physical Activity (β2) controlling Caffeine Intake & Sleep quality.
    3. **Sleep Quality:**
        a. Null Hypothesis (H0): There is no unique effect of Sleep Quality (β3) controlling for Caffeine Intake & Physical activity.

b. Alternative Hypothesis (H1): Sleep Quality (β3) has a unique effect on controlling for Caffeine Intake & Physical activity.

3. **Significance Level:** α = 0.05 (as per general trend)
   a. The **independent variables (IVs)** in this study will include **Caffeine Intake**, measured by the frequency of caffeinated beverage consumption, which reflects how often participants consume caffeine throughout the day. **Physical Activity** will be assessed by the frequency of physical exercise per week(0-7 scale), indicating how many days per week participants engage in exercise. Finally, **Sleep Quality** will be measured by two factors: the average number of hours of sleep participants get per night and the consistency of their sleep schedule, reflecting how regular and stable their sleep patterns are. These independent variables are expected to influence productivity ratings, the primary focus of the study.
   b. This study's **dependent variable (DV)** will be the **Overall Productivity Rating**, which will be assessed through participant self-reports. They will rate their daily productivity on a scale, indicating how productive they feel throughout the day. This variable will help determine how factors such as caffeine intake, physical activity, and sleep quality are related to perceived productivity.

ii. **Pearson Correlation Analysis** - Detect Continuous DV Effects over Productivity (positive, negative, or neutral) for a given day.

1. **Purpose:** To determine the strength and direction of relationships between continuous variables such as XYZ
2. **Correlation between Caffeine Intake and Productivity:-**
   The hypotheses for this study are as follows: The **null hypothesis (H_0)** suggests no correlation between **caffeine intake frequency** and **productivity ratings** ($r = 0$), meaning caffeine consumption does not affect productivity. The **alternative hypothesis (H_1)** proposes a significant correlation ($r \neq 0$), indicating that the frequency of caffeine intake influences productivity ratings. In this analysis, the **independent variable (IV)** is **caffeine intake frequency**, and the **dependent variable (DV)** is the **overall productivity rating**.
3. **Correlation between Physical Activity and Productivity:**
   The hypotheses for the relationship between physical activity and productivity are as follows: The **null hypothesis (H_0)** states that there is no correlation between **physical activity frequency** and **productivity ratings** ($r = 0$), meaning exercise frequency does not affect productivity. The **alternative hypothesis (H_1)** suggests a significant correlation ($r \neq 0$), indicating that the frequency of physical activity influences productivity ratings. In this context, the **independent variable (IV)** is the **frequency of physical exercise per week**, and the **dependent variable (DV)** is the **overall productivity rating**.
4. **Correlation between Sleep Quality and Productivity:**
   The hypotheses for the relationship between sleep and productivity are as follows: The **null hypothesis (H_0)** posits no correlation between **hours of sleep** and **productivity ratings** ($r = 0$), meaning the amount of sleep a person gets does not impact their productivity. The

alternative hypothesis (H_1) suggests a significant correlation ($r \neq 0$), indicating that the number of hours of sleep significantly affects productivity ratings. In this analysis, the **independent variable (IV)** is the **average sleep per night**, and the **dependent variable (DV)** is the **overall productivity rating**.

The analysis focuses on two key components:

1. Overall predictive relationships through multiple regression
2. Individual correlations between continuous variables

This comprehensive analysis approach will allow us to:

- Identify which habits have the strongest relationships with productivity
- Understand how the timing of different activities relates to productivity patterns
- Determine the role of consistency in habits and its relationship with consistent productivity
- Account for both continuous and categorical variables in our dataset

By examining these relationships through the above-mentioned statistical approaches, we can provide robust insights into how caffeine intake, physical activity, and sleep quality correlate with productivity among university students.