**Project Part 1**

1. **A short summary (1/2 page to 1 page) of:**
   1. **Your interests**

From early in my career, I’ve been drawn to how data can uncover patterns and improve decision-making. Working across industries, I saw firsthand how data-driven solutions boost efficiency. A key experience was at Rakuten, where applying market basket analysis improved promotions, directly enhancing customer satisfaction and sales. This deepened my interest in using data for business growth and addressing societal challenges, particularly in healthcare. The combination of problem-solving and making a broader impact continues to drive my passion for data science.

* 1. **The reasons why you choose your current degree and major**

I chose to pursue a Master’s in Applied Data Science at USC to deepen my technical skills and explore emerging areas like Generative AI, Natural Language Processing, and Computer Vision. While my data science career has been rewarding, I wanted a structured academic approach to build a stronger foundation in advanced topics. USC’s hands-on projects and research opportunities align perfectly with my growth goals in this field.

* 1. **The reasons why you decided to take this class**

I enrolled in the "Research Methods and Analysis for User Studies" course to learn how to design and conduct research involving human subjects. While my work has focused on data analytics, I recognized the need to understand the ethical and methodological aspects of working with human data. This course will help me ensure that my future research is both scientifically sound and ethically responsible.

* 1. **Your personal ambitions to change the world**

I aspire to use my data science skills to make a meaningful impact on healthcare systems. One of my key ambitions is to develop AI-powered solutions that can revolutionize healthcare diagnostics and patient care, making it more personalized and accessible. I believe that through data-driven research and innovations, we can significantly improve the quality of life for individuals, especially in underserved populations where access to healthcare is limited.

* 1. **The reasons why you are interested in the topic you have chosen for your project.**

I chose this topic because I’m fascinated by how daily habits like caffeine intake, physical activity, and sleep can impact productivity and well-being. This research aligns with my interest in using data to improve everyday life. By studying these correlations, I aim to identify trends that could lead to practical recommendations for healthier, more productive lifestyles, which makes this project exciting for me.

* 1. **Show me a screenshot of your CITI certification for human subjects research.**

****

1. **Sketch out the plan for the user study that you will conduct this term, including details such as:**
   1. **What variables are you going to collect?**
      1. **Caffeine Intake:**
         1. Frequency of Consumption(Never, Rarely, Occasionally, Frequently, daily)
         2. Timing (Morning, Afternoon, Evening, Late-Night, Never)
      2. **Physical Activity:**
         1. Frequency(Scale 0-7)
         2. Timing(Morning, Afternoon, Evening, Late-Night)
         3. Variations in Energy levels(Yes, No, Somewhat, Unsure)
      3. **Sleep Quality:**
         1. Hours of sleep per night
         2. Consistency of sleep schedule (Consistent, Somewhat Consistent, Somewhat in-consistent, Very in-consistent)
      4. **Daily Productivity:**
         1. Self-reported productivity rating (scale from 1 to 5)
         2. Productivity Consistency (Consistent, Somewhat Consistent, Neutral, Somewhat in-consistent, Very in-consistent)
         3. Variation in Productivity (Morning, Afternoon, Evening, Late-Night, Constant)
   2. **What design is your study (experimental vs. correlational; if experimental, what factors are between subjects vs. within subjects)?**

**Study Design:** Correlational Study  
 **Justification:**

* + 1. 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.
    2. Our study aims to examine the relationship between daily habits (caffeine intake, physical activity, sleep quality) and productivity.
    3. 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.
  1. **Given those answers, out of those variables which are your IV(s) and DV(s)?**
     1. **Independent Variables:**
        1. Caffeine Intake:
           1. Frequency of Consumption(Never, Rarely, Occasionally, Frequently, daily)
           2. Timing (Morning, Afternoon, Evening, Late-Night, Never)
        2. Physical Activity:
           1. Frequency(Scale 0-7)
           2. Timing(Morning, Afternoon, Evening, Late-Night)
           3. Variations in Energy levels(Yes, No, Somewhat, Unsure)
        3. Sleep Quality:
           1. Hours of sleep per night
           2. Consistency of sleep schedule (Consistent, Somewhat Consistent, Somewhat in-consistent, Very in-consistent)
     2. **Dependent Variables:**
        1. Daily Productivity:
           1. Self-reported productivity rating (scale from 1 to 5)
           2. Productivity Consistency (Consistent, Somewhat Consistent, Neutral, Somewhat in-consistent, Very in-consistent)
           3. Variation in Productivity (Morning, Afternoon, Evening, Late-Night, Constant)
  2. **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. So, operational definitions for each independent variable (IV) and dependent variable (DV) will be measured. Our data will be collected through self-reported surveys.

* + 1. **Independent Variables (IVs):**

These variables relate to everyday habits, and we will ask participants to report on them using specific questions designed to capture the required data.

* + - 1. **Caffeine Intake:**
         1. Frequency of Consumption:

Operational Definition: Self-reported frequency of caffeine consumption.

Measurement: Participants will choose one of the following options (Never, Rarely, Occasionally, Frequently, daily)

* + - * 1. Timing of Consumption:

Operational Definition: Self-reported time of day caffeine is consumed.

Measurement: Participants will indicate the time(s) they typically consume caffeine (Morning, Afternoon, Evening, Late Night, Never)

* + - 1. **Physical Activity:**
         1. Frequency of Activity:

Operational Definition: Number of days per week participants engage in physical activity.

Measurement: Participants will report the number of days they are physically active in a typical week, rated on a scale from 0 (no activity) to 7 (daily activity).

* + - * 1. Timing of Activity:

Operational Definition: Self-reported time of day participants engage in physical activity.

Measurement: Participants will indicate when they typically exercise (Morning, Afternoon, Evening, Late Night)

* + - * 1. Variations in Energy Levels:

Operational Definition: Self-reported variation in energy levels during the day.

Measurement: Participants will choose from the following options(Yes, No, Somewhat, Unsure)

* + - 1. **Sleep Quality:**
         1. Hours of Sleep per Night:

Operational Definition: Average hours of sleep participants get per night.

Measurement: Participants will report the typical number of hours they sleep each night.

* + - * 1. Consistency of Sleep Schedule:

Operational Definition: Self-reported consistency of participants’ sleep schedules.

Measurement: Participants will rate their sleep schedule consistency(Consistent, Somewhat Consistent, Somewhat Inconsistent, Very Inconsistent)

* + 1. **Dependent Variables (DVs):**
       1. **Daily Productivity:**
          1. Self-Reported Productivity Rating:

Operational Definition: Participants’ subjective rating of their daily productivity.

Measurement: Rated on a scale from 1 (very unproductive) to 5 (very productive).

* + - * 1. Productivity Consistency:

Operational Definition: How consistent participants feel their productivity is over time.

Measurement: Rated with the following options Consistent, Somewhat Consistent, Neutral, Somewhat Inconsistent, Very Inconsistent)

* + - * 1. Variation in Productivity:

Operational Definition: Self-reported variations in productivity throughout the day.

Measurement: Participants will indicate when they feel most productive(Morning, Afternoon, Evening, Late-Night, Constant)

* 1. **What is your population? How are you going to get participants from that population? How many are you planning to recruit for the study?**
     1. 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.
     2. 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.
     3. 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.

1. **Sketch out your plan for analysis:**
   1. **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.**
      1. **Research question:** How do caffeine intake, physical activity, and sleep quality influence self-reported productivity levels among university students, both on a daily and weekly basis?
      2. We will answer this question by analyzing the correlations between our independent variables 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. By comparing the relationships between different habits and productivity measures, we can determine which factors have the strongest association with productivity outcomes.
      3. **Data Collection Mapping:**  
         **Primary DVs from Survey:   
         -** Daily productivity rating (1-5 scale)  
         - Productivity consistency rating  
         - Time-of-day productivity variations  
         **Primary IVs from Survey:  
         -** Caffeine consumption frequency and timing  
         - Physical activity frequency (0-7 scale) and timing  
         - Sleep hours and consistency rating
   2. **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:

* + 1. **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. **Hypotheses:**
          1. Null Hypothesis (H0): The combination of caffeine intake, physical activity, and sleep quality does not significantly predict productivity ratings (R² = 0)
          2. Alternative Hypothesis (H1): At least one daily habit significantly predicts productivity ratings (R² > 0)
       3. **Significance Level:** α = 0.05 (as per general trend)
          1. **Independent Variables (IVs):**

Caffeine Intake: Frequency of Caffeinated Beverage Consumption

Physical Activity: Frequency of Physical Exercise per Week

Sleep Quality: Average Sleep per Night and Consistency of Sleep Schedule

* + - * 1. **Dependent Variable (DV):**

Overall Productivity Rating

* + 1. **Pearson Correlation Analysis** - Detect Continuous DV Effects over Productivity (positive, negative, or neural)
       1. **Purpose:** To determine the strength and direction of relationships between continuous variables such as xyz
       2. **Main Effect of Caffeine Intake:**
          1. **H0:** There is no correlation between caffeine intake frequency and productivity ratings (r = 0)
          2. **H1:** There is a significant correlation between caffeine intake frequency and productivity ratings (r ≠ 0)
          3. **IV:** Frequency of Caffeinated Beverage Consumption
          4. **DV:** Overall Productivity Rating
       3. **Main Effect of Physical Activity:**
          1. **H0:** There is no correlation between physical activity frequency and productivity ratings (r = 0)
          2. **H1:** There is a significant correlation between physical activity frequency and productivity ratings (r ≠ 0)
          3. **IV:** Frequency of Physical Exercise per Week
          4. **DV:** Overall Productivity Rating
       4. **Main Effect of Sleep Quality:**
          1. **H0:** There is no correlation between hours of sleep and productivity ratings (r = 0)
          2. **H1:** There is a significant correlation between hours of sleep and productivity ratings (r ≠ 0)
          3. **IV:** Average Sleep per Night
          4. **DV:** Overall Productivity Rating
    2. **Chi-Square Tests of Independence:**
       1. **Purpose**: To analyze relationships between categorical variables
       2. **Variables to be analyzed:**
          1. Timing of habits (Morning/Afternoon/Evening/Late-Night) vs. Productivity variations

IVs: Timing of Caffeinated Beverage Consumption, Time of Day for Physical Activity

DV: Productivity Consistency (e.g., Consistency in Productivity Throughout the Week)

* + - * 1. Consistency measures vs. Productivity consistency

IVs: Consistency of Sleep Schedule, Energy Level Patterns

DV: Consistency in Productivity Throughout the Week

* + - 1. **Hypotheses**:
         1. H0: No association exists between timing/consistency variables and productivity patterns
         2. H1: Significant associations exist between timing/consistency variables and productivity patterns
    1. **Weekly Analysis:**
       1. **Purpose:** To examine weekly patterns and aggregate effects or to analyze changes across weeks
       2. **Repeated Measures ANOVA:**
          1. **Variables:**

Weekly average productivity ratings  
DV: Overall Productivity Rating (assessed weekly)

Weekly patterns of caffeine consumption  
IV: Frequency of Caffeinated Beverage Consumption (assessed weekly)

Weekly physical activity frequency  
IV: Frequency of Physical Exercise per Week (assessed weekly)

Weekly sleep consistency  
IVs: Average Sleep per Night, Consistency of Sleep Schedule (assessed weekly)

* + - * 1. **Hypotheses:**

H0: No significant differences in productivity patterns across weeks

H1: Significant differences exist in productivity patterns across weeks

The analysis focuses on three key components:

1. Overall predictive relationships through multiple regression
2. Individual correlations between continuous variables
3. Associations between categorical timing and consistency measures

This comprehensive analysis approach will allow us to:

* Identify which daily habits have the strongest relationships with productivity
* Understand how 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 multiple statistical approaches, we can provide robust insights into how caffeine intake, physical activity, and sleep quality correlate with productivity among university students.