In today's fast-paced work environment, understanding the interplay between lifestyle factors and work performance is more crucial than ever. In this project, I wanted to explore how key variables—mood, exercise, sleep (including both hours slept and sleep quality), and screen time before bed—affect work performance.

The project is built on the premise that our daily habits and behaviors significantly influence our efficiency and cognitive function at work. For example, regular exercise may boost mood and energy levels, while quality sleep is essential for mental clarity and decision-making. Conversely, excessive screen time before bed could disrupt sleep quality and reduce overall performance.

Ultimately, I wanted to investigate whether there were lifestyle choices that impacted work outcomes that can be used as practical recommendations for individuals srtiving to achieve a positive work-life balance.

It should be kept in mind that this is a fictional data set.

This data set was downloaded from Kaggle.

sleep_cycle_producti DataFrame as df8

```
SELECT *
FROM sleep_cycle_productivity.csv
LIMIT 5;
```

	Date	P.			Sleep Start ...	Sleep En...	Total Sleep H...	Sleep ...	Exercise (mins/d...	Caffeine Intake (mg)
0	2024-04-12T00:00:00.000	1860	32	Other	23.33	4.61	5.28	3	86	87
1	2024-11-04T00:00:00.000	1769	41	Female	21.02	2.43	5.41	5	32	21
2	2024-08-31T00:00:00.000	2528	20	Male	22.1	3.45	5.35	7	17	88
3	2024-02-22T00:00:00.000	8041	37	Other	23.1	6.65	7.55	8	46	34
4	2024-02-23T00:00:00.000	4843	46	Other	21.42	4.17	6.75	10	61	269

Rows: 5

Before filtering and sorting, I want to check the overall trends in the data such as:

- What does the data consist of? Since the data is self-reported, I wanted to know what percent of the participants have multiple submissions vs single submissions.
- On average, how much sleep do people have?
- Is there a day of the week where people get more sleep?

Then, I would focus on the following:

- How do people self-report their productivity? What percent of the participants fall into top, middle, and low productivity groups?
- How much sleep did top performers(most productivity) have? How much sleep did middle and lower performers have?
- Does productivity correlate with any factors reported in this study?
- Does work stage correlate with any of these factors?

First, I wanted to make sure that there were no missing values in the productivity scores since I will use those scores to determine work performance.

sleep_cycle_producti DataFrame as df1

```
SELECT *
FROM sleep_cycle_productivity.csv
WHERE "Productivity Score" IS NULL;
```

Your query ran successfully but returned no results.

Next, I wanted to get an overall idea of what the participant input was. For this, I wanted to know if 1 participant submitted information multiple times.

sleep_cycle_producti DataFrame as df2

```
SELECT
    COUNT(*) AS total_partipants,
    COUNT(DISTINCT person_id) AS unique_ids,
    ROUND(COUNT(DISTINCT person_id)/COUNT(person_id)*100,1) AS percent_unique_submissions,
    ROUND(100-(COUNT(DISTINCT person_id)/COUNT(person_id)*100),1) AS percent_repeat_submissions
FROM sleep_cycle_productivity.csv;
```

index	total_partipants	unique_ids	percent_unique_submissions	percent_repeat_submissions
0	5000	3858	77.2	

Rows: 1

From this, we can see that of the 5000 submissions, 3858 submissions (77.2%) were unique submissions, meaning that the remaining 22.8% are multiple submissions.

sleep_cycle_producti DataFrame as df3

```
-- On average, how much sleep did participants have?
SELECT
    ROUND(AVG("Total Sleep Hours"),2) AS average_hours_slept,
    ROUND(STDDEV("Total Sleep Hours"),3) AS std_dev,
    ROUND(VARIANCE("Total Sleep Hours"),3) AS variance
FROM sleep_cycle_productivity.csv;
```

index	average_hours_slept	std_dev	variance
0	6.97	1.454	

Rows: 1

On average, particpants sleep ~7 hours a night, which is in line with the national average according to the CDC.

sleep_cycle_producti DataFrame as df4

```
-- Is there a day of the week where people get more sleep?
SELECT
    EXTRACT('DOW' FROM "Date") AS day_of_week,
    ROUND(AVG("Total Sleep Hours"),1) AS total_hours_slept
FROM sleep_cycle_productivity.csv
GROUP BY day_of_week;
--Note: 0=Sunday, 6= Saturday
```

index	day_of_week	total_hours_slept
0	0	
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	

Rows: 7

Surprisingly, there is no correlation between the day of the week and the amount of sleep that particpants slept.

Next, I split the particpants into 4 groups based on their productivity scores: high productivity (9-10), good productivity (7-8), mediocre productivity (4-6), and poor productivity (1-3) and further investigated trends.

sleep_cycle_producti DataFrame as df5

```sql
-- How many people consider themselves to have high, good, mediocre, and poor productivity
SELECT
    CASE WHEN "Productivity Score" BETWEEN 9 AND 10 THEN 'high'
         WHEN "Productivity Score" BETWEEN 7 AND 8 THEN 'good'
         WHEN "Productivity Score" BETWEEN 4 AND 6 THEN 'mediocre'
         ELSE 'poor' END AS performance,
    ROUND((COUNT(*)/5000)*100,2) AS percent_total --Number of participants: 5000
FROM sleep_cycle_productivity.csv
GROUP BY performance;
```

index	performance	percent_total
0	poor	
1	good	
2	high	
3	medicore	

Rows: 4

From this we can see that the majority of participants believe themselves to have poor (~28.6%) and mediocre productivity (~29.1%) consisting of in 57.7% of the total participants. While the remaining 42.3% of the participants believe themselves to have high (~22.0%) and good (~20.3%) productivity.

sleep_cycle_producti DataFrame as df6

```sql
SELECT
    CASE WHEN "Productivity Score" BETWEEN 9 AND 10 THEN 'high'
         WHEN "Productivity Score" BETWEEN 7 AND 8 THEN 'good'
         WHEN "Productivity Score" BETWEEN 4 AND 6 THEN 'mediocre'
         ELSE 'poor' END AS performance,
    ROUND(AVG("Total Sleep Hours"),2) AS avg_hours_slept,
    ROUND(AVG("Sleep Quality"),2) AS avg_sleep_quality,
    ROUND(AVG("Exercise (mins/day)"),2) AS avg_exercise_mins_day,
    ROUND(AVG("Caffeine Intake (mg)"),2) AS avg_caffine_intake,
    ROUND(AVG("Screen Time Before Bed (mins)"),2) AS avg_screentime_before_bed,
    ROUND(AVG("Work Hours (hrs/day)"),2) AS average_work_hrs,
    ROUND(AVG("Mood Score"),2) AS average_mood_score,
    ROUND(AVG("Stress Level"),2) AS average_stress_level
FROM sleep_cycle_productivity.csv
GROUP BY performance;
```

	per...	avg_h...	avg_slee...	avg_exercise_mins_...	avg_caffine...	avg_screentime_bef...	average_w...	average_mo...	average_stre...
0	poor	6.98	5.59	44.91	146.05	89.65	7.97	5.41	5.55
1	mediocre	6.89	5.44	43.78	145.98	94.45	8.08	5.36	5.58
2	good	7.06	5.5	43.33	147.29	88.28	7.93	5.4	5.48
3	high	7	5.57	43.56	148.01	92.62	7.95	5.31	5.57

Rows: 4

Surprisingly, there looks like there aren't any overarching trends between productivity and the data provided- mood, sleep, sleep quality, and exercise.

Instead of looking at productivity, let's look to see if career stage effects outcomes.

sleep_cycle_producti DataFrame as df7

```
SELECT
    CASE WHEN age BETWEEN 18 AND 24 THEN 'Young Adult'
        WHEN age BETWEEN 25 AND 34 THEN 'Early Career'
        WHEN age BETWEEN 35 AND 44 THEN 'Mid-Career'
        WHEN age BETWEEN 45 AND 54 THEN 'Late Adult'
    ELSE 'Pre-retirement' END as career_stage,
    ROUND(AVG("Total Sleep Hours"),2) AS avg_hours_slept,
    ROUND(AVG("Sleep Quality"),2) AS avg_sleep_quality,
    ROUND(AVG("Exercise (mins/day)"),2) AS avg_exercise_mins_day,
    ROUND(AVG("Caffeine Intake (mg)"),2) AS avg_caffine_intake,
    ROUND(AVG("Screen Time Before Bed (mins)"),2) AS avg_screentime_before_bed,
    ROUND(AVG("Work Hours (hrs/day)"),2) AS average_work_hrs,
    ROUND(AVG("Mood Score"),2) AS average_mood_score,
    ROUND(AVG("Stress Level"),2) AS average_stress_level
FROM sleep_cycle_productivity.csv
GROUP BY career_stage;
```

	career_s...	avg_ho...	avg_slee...	avg_exercise_mins_...	avg_caffine...	avg_screentime_bef...	average...	average_...	average_stress_lev...
0	Mid-Career	6.91	5.61	43.82	146.33	91.84	8.01	5.39	
1	Young Adult	6.96	5.49	44.2	147.23	91.24	7.94	5.43	
2	Late Adult	6.99	5.4	44.48	149.67	89.87	7.97	5.37	
3	Pre-retirement	6.93	5.56	44.58	142.72	91.61	8.04	5.41	
4	Early Career	7.05	5.55	43.06	145.9	92.61	7.99	5.29	

Rows: 5

According to the data, there are no identifiable trends between career stage and the data provided.

These results could be due to the fact that this is a synthetic data set which may not be representative of real-world trends. Additonally, if these were actual participants, these values were modeled to be "self-reported", which may skew the data on productivity scoring.