1. Using the cleaned data that I have processed previously, I will then use query languages, Presto, to explore potentially insightful results from my current data sq2.csv, which contains the information of one's status of sleep from the experiment.

1) Query the average time_in_bed for each gender to see the gap in each gender.

```
presto>  select sex, avg(time_in_bed) from hive.zl3152_nyu_edu.sq
     ->  GROUP BY sex;
  sex   |       _col1
--------+--------------------
 Male   | 6.142857142857143
 Female | 6.800000000000002
(2 rows)

Query 20230417_025949_00062_d2d8x, FINISHED, 2 nodes
Splits: 81 total, 81 done (100.00%)
0.22 [61 rows, 8.08KB] [273 rows/s, 36.2KB/s]
```

- It can be observed that the male from the group on average spend less time in bed than the female by about 0.64 hr

2) Query the average time_in_bed for people with different ages.

```
presto> select age, avg(time_in_bed) from hive.zl3152_nyu_edu.sq
     ->  GROUP BY age
     ->  order by age asc;
 age |       _col1
-----+--------------------
 20  |  6.266666666666667
 21  | 5.7875000000000005
 22  | 6.1000000000000005
 23  |                7.0
 24  |  7.000000000000001
 25  | 6.6571428571428575
 26  | 6.6428571428571415
 27  |               5.75
 28  |                8.1
(9 rows)

Query 20230417_030623_00064_d2d8x, FINISHED, 2 nodes
Splits: 100 total, 100 done (100.00%)
0.22 [61 rows, 8.08KB] [274 rows/s, 36.4KB/s]
```

- There are no interesting findings on time in bed per age, but a sudden drop of time in bed at the age of 27 appears to be somewhat dubious, as I will have to consider the background of the subjects in this experiment to account for this data.

3) Query the sleep_duration across genders

```
presto>   select sex, avg(sleep_duration) from hive.zl3152_nyu_edu.sq
    ->  GROUP BY sex;
  sex   |       _col1
--------+--------------------
 Female | 6.222499999999999
 Male   | 5.538095238095237
(2 rows)

Query 20230417_031115_00065_d2d8x, FINISHED, 2 nodes
Splits: 81 total, 81 done (100.00%)
0.22 [61 rows, 8.08KB] [277 rows/s, 36.7KB/s]
```

- Females sleep about 0.7hrs more than males do

4) Query the sleep_duration across ages

```
presto>  select age, avg(sleep_duration) from hive.zl3152_nyu_edu.sq
    ->  GROUP BY age
    ->   order by age asc;
 age |       _col1
-----+--------------------
 20  | 5.8999999999999995
 21  |             5.2125
 22  |               5.45
 23  |  6.325000000000001
 24  |  6.483333333333332
 25  |  5.964285714285714
 26  |                6.1
 27  |                5.4
 28  |                7.7
(9 rows)

Query 20230417_031433_00067_d2d8x, FINISHED, 2 nodes
Splits: 100 total, 100 done (100.00%)
0.22 [61 rows, 8.08KB] [278 rows/s, 36.9KB/s]
```

- This resembles the pattern as we did for time_in_bed. It appears that people at age 20-23 sleeps relatively short, perhaps due to classes.

2. Notice, all code is under cmd.txt

3. There is more data to be put, but I have not completed the cleaning of them at this point, and I will reflect on the change next time.

- There are other data about performance of one's behavior, which I will be later joining the tables to the sleep quality to find the correlation between performance and sleep quality.

4. The presentation slide is the following link:
https://wepik.com/edit/33b8ddbf-c36c-40b3-83ef-ba3f1736b42a?lang=en#rs=landing-ai-slidesgo