Eight Sleep
Data Analysis

Arpan Choudhury

Exploratory Data Analysis

Exploratory Data Analysis

- Cleansed Data by handling missing values
- Extracted age from the date of birth and added it to the table
- Separated the hour from start-time and end-time variables into a new column
- Visualized distributions of different features and Correlation matrices to observe relationship between features on a high level

Overview

Dataset info

Number of variables	16
Number of observations	1042308
Total Missing (%)	21.9%
Total size in memory	127.2 MiB
Average record size in memory	128.0 B

Variables types

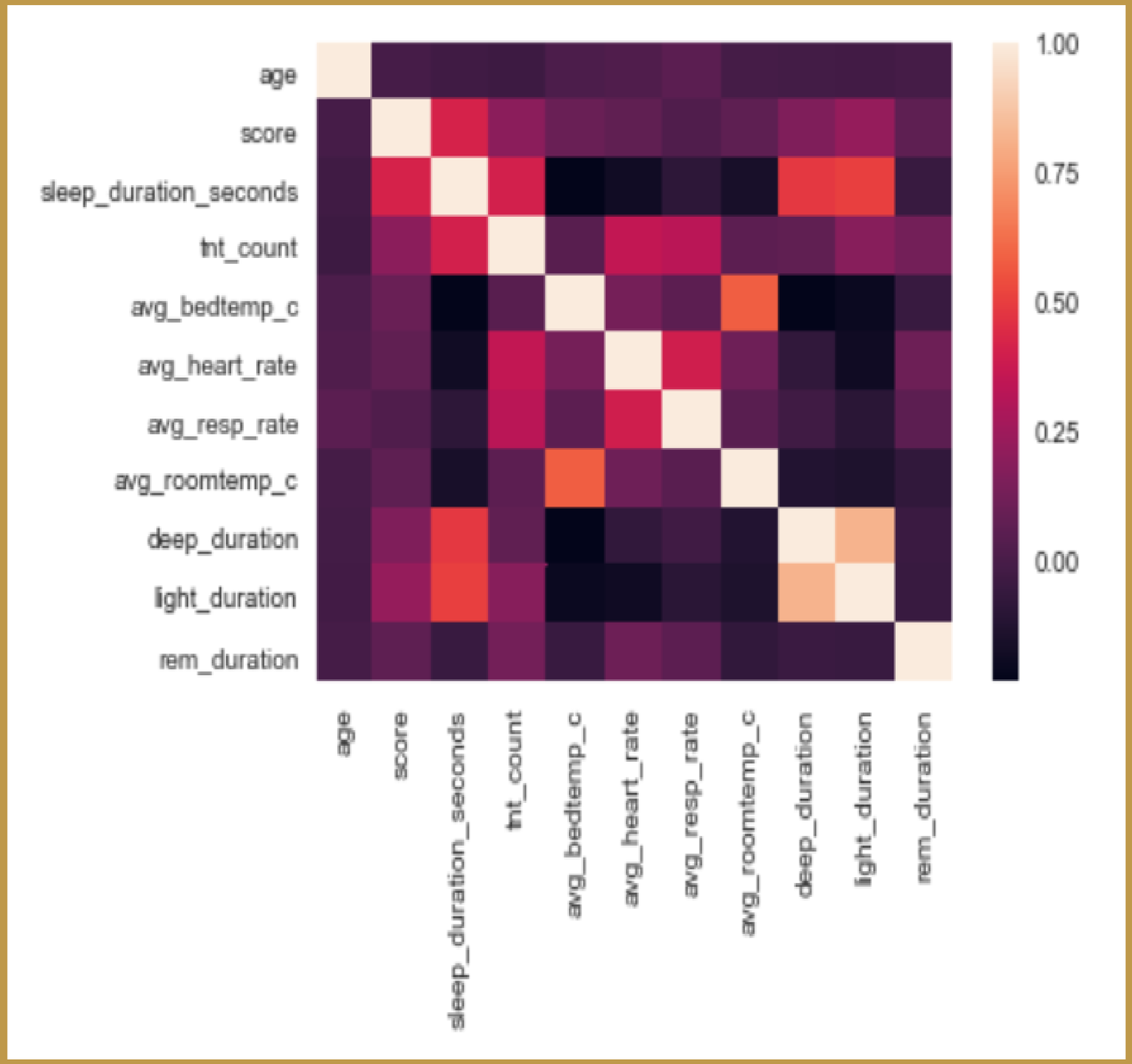
Numeric	9
Categorical	6
Boolean	0
Date	0
Text (Unique)	1
Rejected	0
Unsupported	0

Warnings

- avg_bedtemp_c has 315057 / 30.2% missing values Missing
- avg_heart_rate has 1034168 / 99.2% missing values Missing
- avg_resp_rate has 1034166 / 99.2% missing values Missing
- avg_roomtemp_c has 1034133 / 99.2% missing values Missing
- deep_duration has 137174 / 13.2% zeros Zeros
- deep_duration is highly skewed (γ1 = 120.2) Skewed
- end_ts has a high cardinality: 364775 distinct values Warning
- light_duration is highly skewed (γ1 = 84.463) Skewed
- presence_duration_seconds has 79922 / 7.7% missing values Missing
- presence_duration_seconds has a high cardinality: 3103 distinct values Warning
- rem_duration has 845004 / 81.1% zeros Zeros
- session_duration_seconds has 79922 / 7.7% missing values Missing
- session_duration_seconds has a high cardinality: 3156 distinct values Warning
- sleep_duration_seconds has 79922 / 7.7% missing values Missing
- sleep_duration_seconds has a high cardinality: 1655 distinct values Warning
- start_ts has a high cardinality: 366521 distinct values Warning
- tnt_count has 182368 / 17.5% zeros Zeros
- user_id has a high cardinality: 6513 distinct values Warning

Interval Dataset Correlation Overview

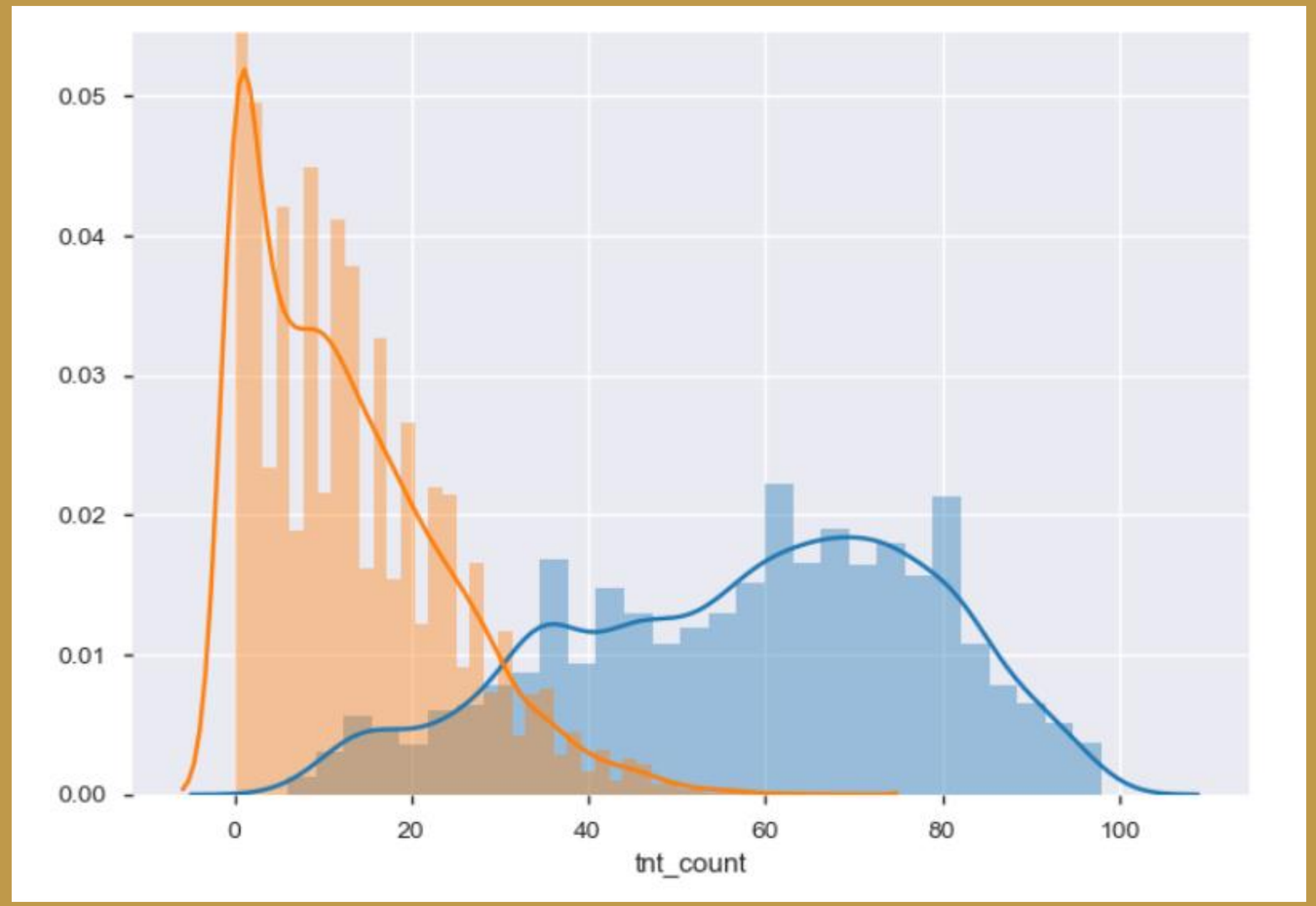
- The ‘Average Bed Temperature’ and the ‘Average Room Temperature’ (in orange) are the most highly correlated.
- ‘Average Heart Rate’ and ‘Average Respiration Rate’ are correlated with ‘Tosses and Turns Count’ .
- Durations of Deep Sleep/Light Sleep increases with increase in Sleep Duration, hence they are also highly correlated.



Variable Distributions

- This plot shows the distributions of the “Score” and “Tosses and Turns” data in orange and blue respectively.
- The Score data is a right skewed histogram
- The tosses and turns data histogram is slightly left skewed

Score and Tosses/Turns data distributions

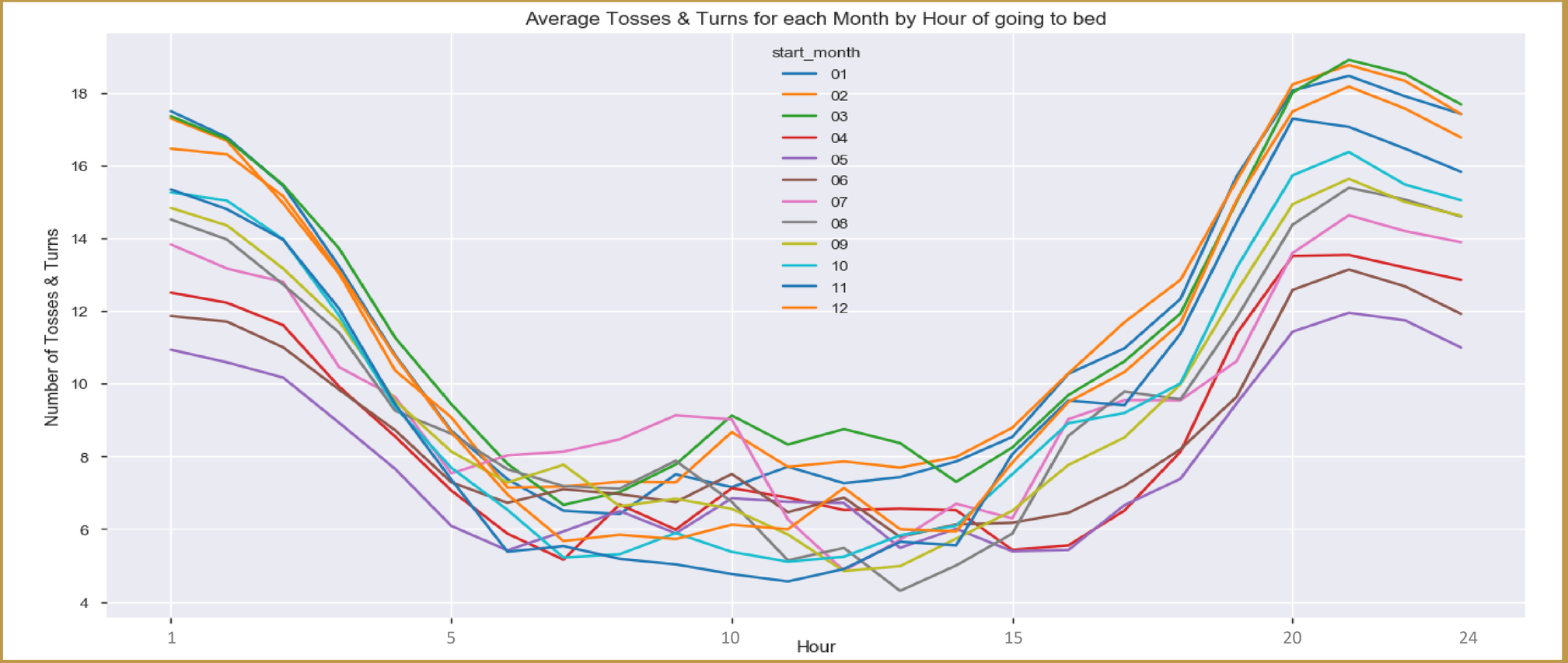


1

User Bed-time/Wake-time and Other Features Correlation

Average Tosses & Turns for each Month by Hour of going to bed

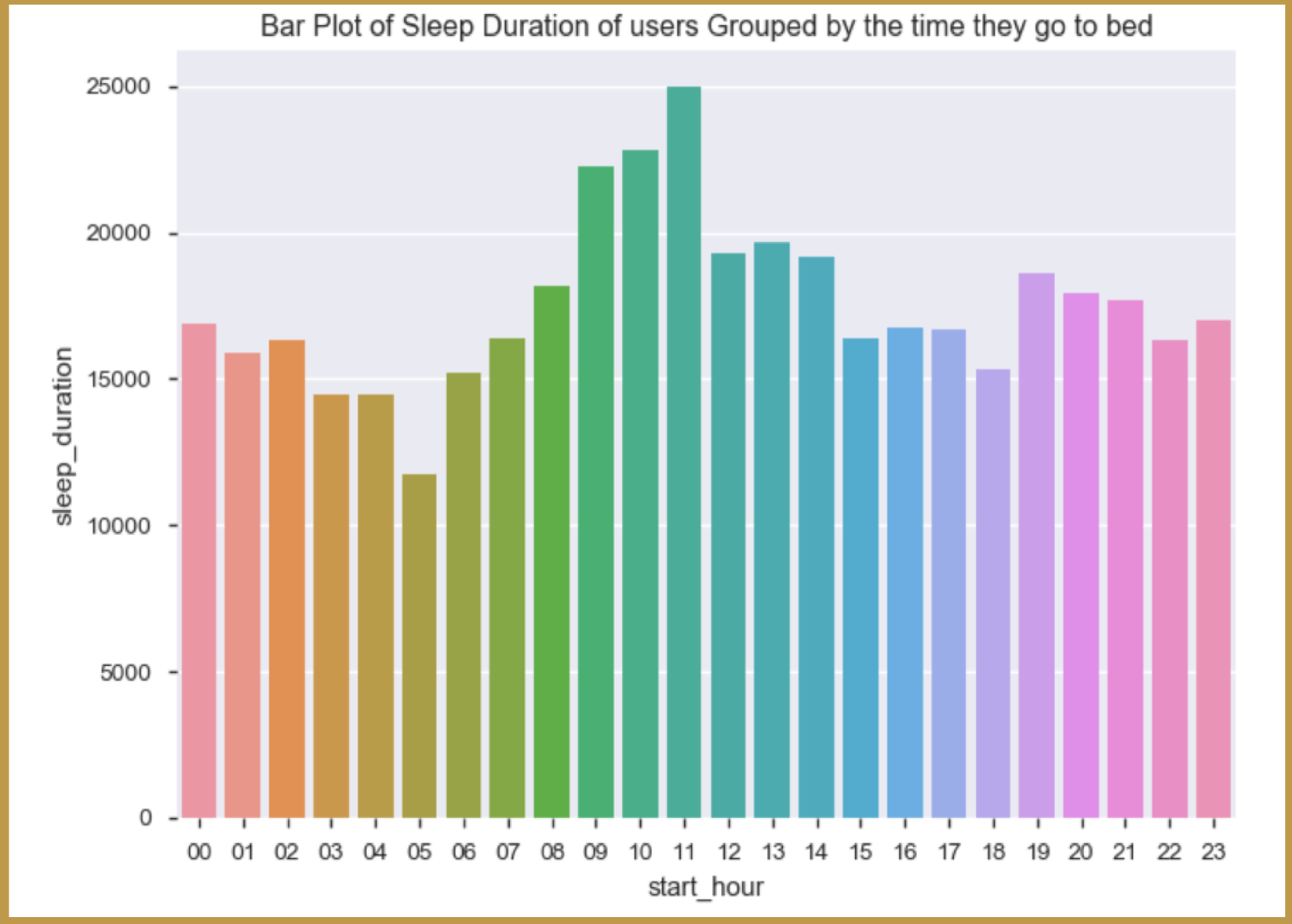
- On an average, users toss and turn the least during the Months of May and June (during Summer).
- On an average, users toss and turn the most in the months early months of Jan, Feb and March (during Winter)
- On average, tosses and turns happen least between hours 06 – 18 (i.e., during the day)
- Users' tosses in sleep during the months of March and July are the most irregular:
  - In July, tosses and turns suddenly become the highest in the hours before noon. And they become very low just after noon (Pink line)
  - In March, vice versa



Correlation of Sleep Duration & Bed Time of the Users

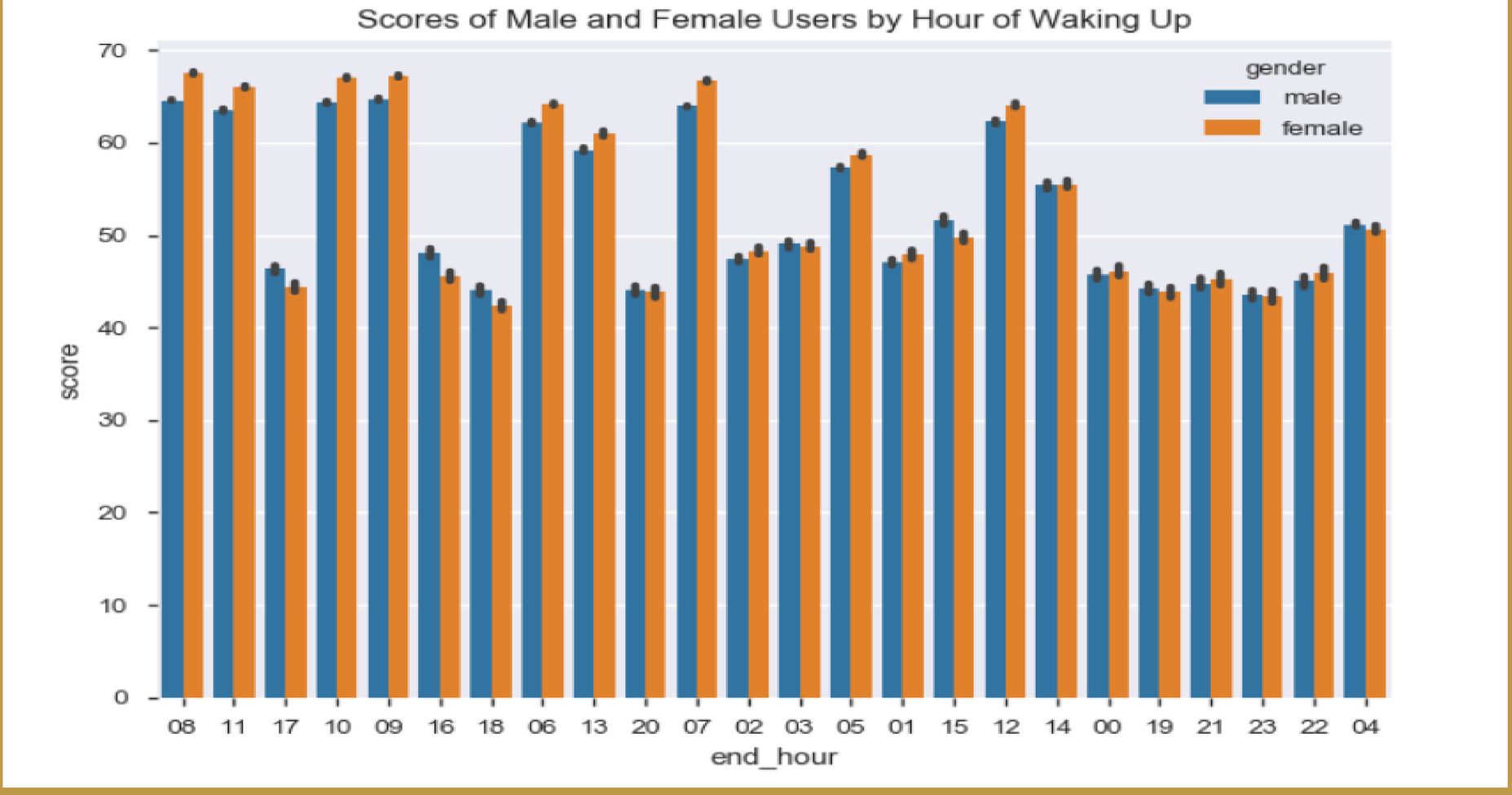
- On average, users going to bed during the 12th hour (11am – 12pm) of the day sleep the longest. In general, users going to bed during the 10th, 11th, and 12th hours (i.e. just in the hours before noon) sleep for much longer than the average sleep duration.

- Also, users going to bed in the 6th hour (i.e. after 5am) sleep for the shortest duration



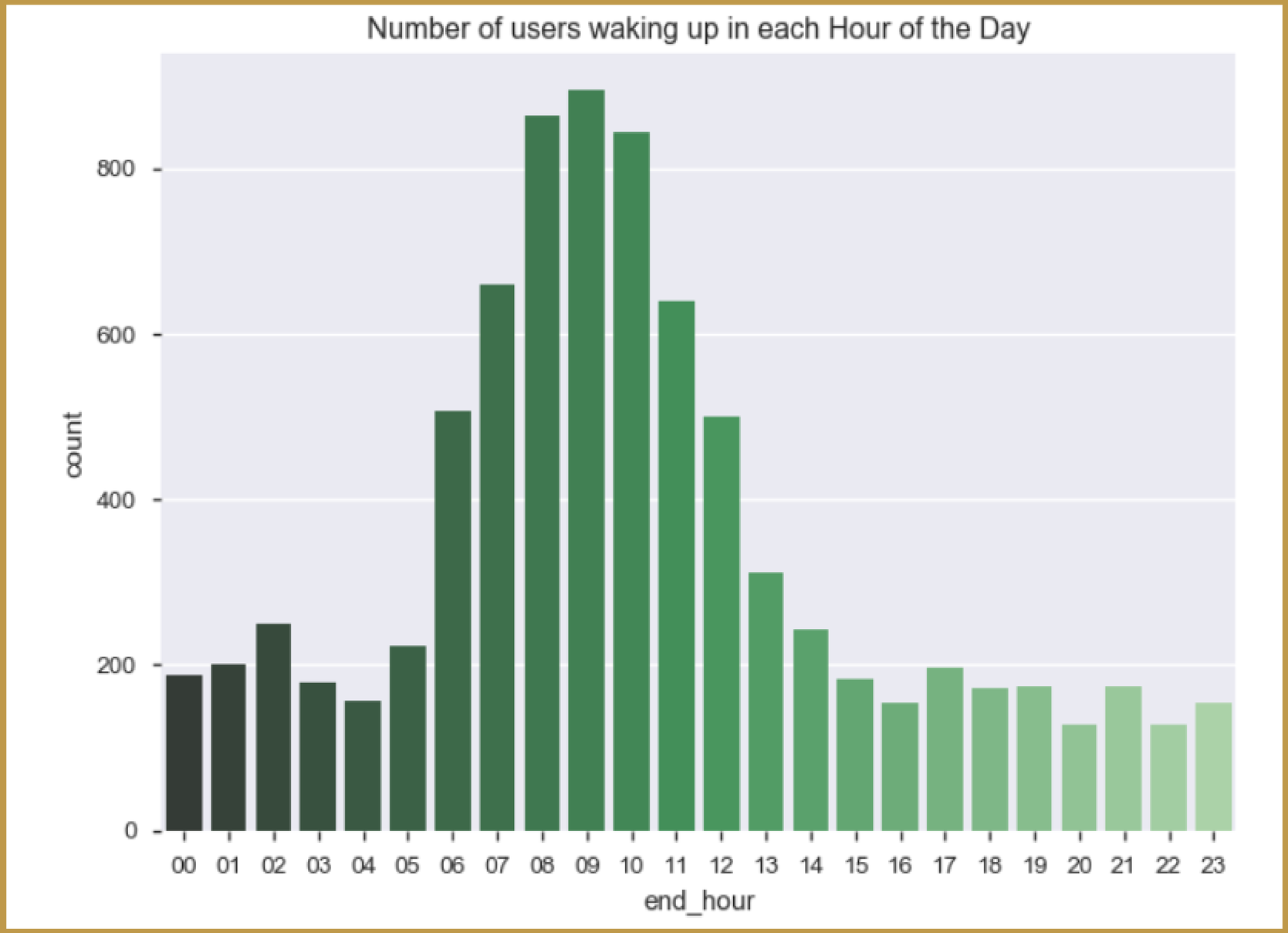
Scores of Male and Female Users by Hour of Waking Up

- This bar plot shows that mostly when scores of both males and females are at least 50 in a month, Females have a higher score than males.
- And mostly when scores are below 50 for both genders, males have a score greater than or equal to that of their female counterparts.



Number of users waking up in each Hour of the Day

- ✓ From this bar plot, we can see that the number of users waking up after 9am (i.e., between 9 am and 10 am) are the maximum.
- ✓ Also, only during hours 08, 09 and 10, there are at least 800 users waking up.
- ✓ On an average, the number of users waking from the early evening to all night (2pm -5 am), are same i.e. 200



2

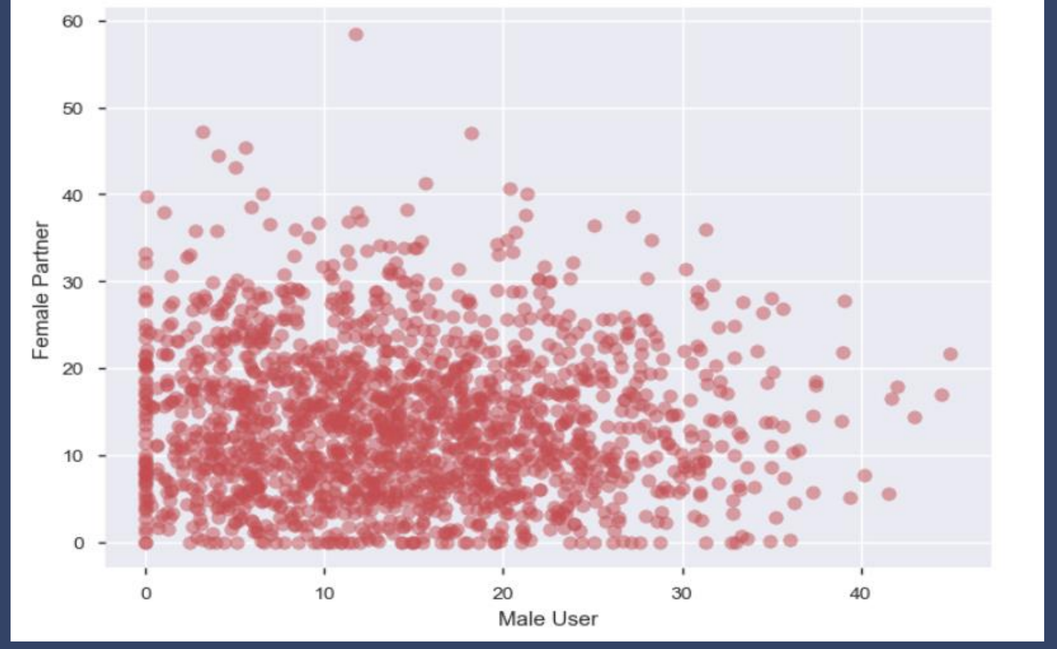
User – Partner Correlation

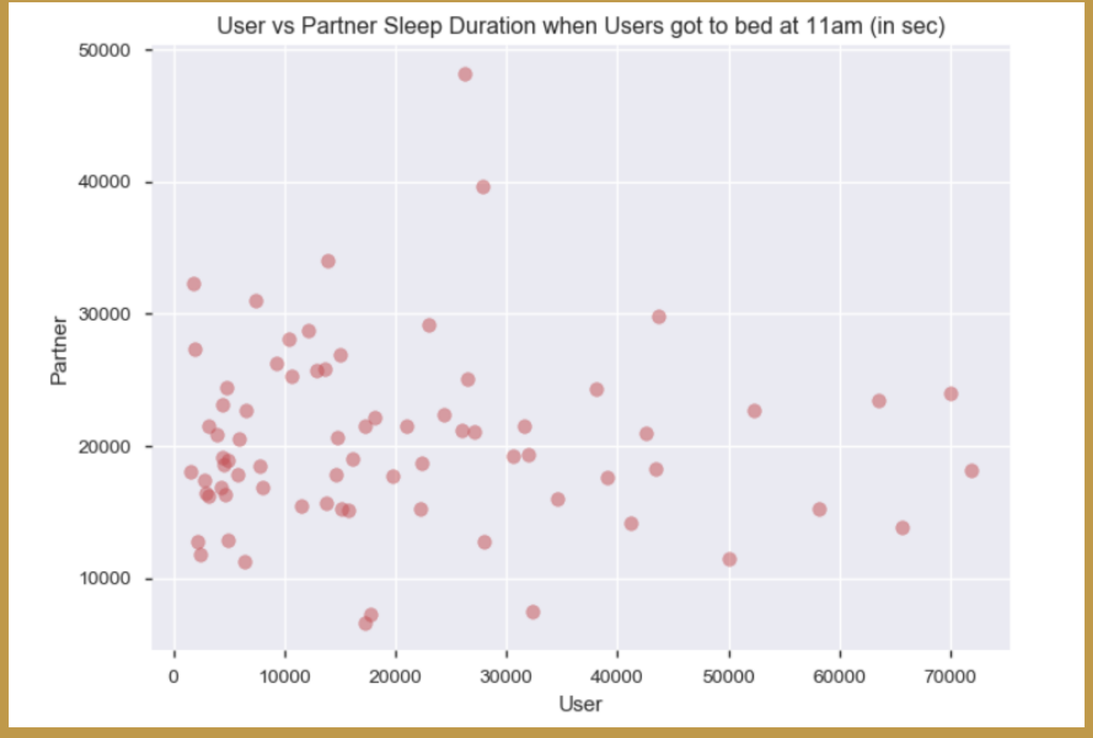
User & Partner Analysis

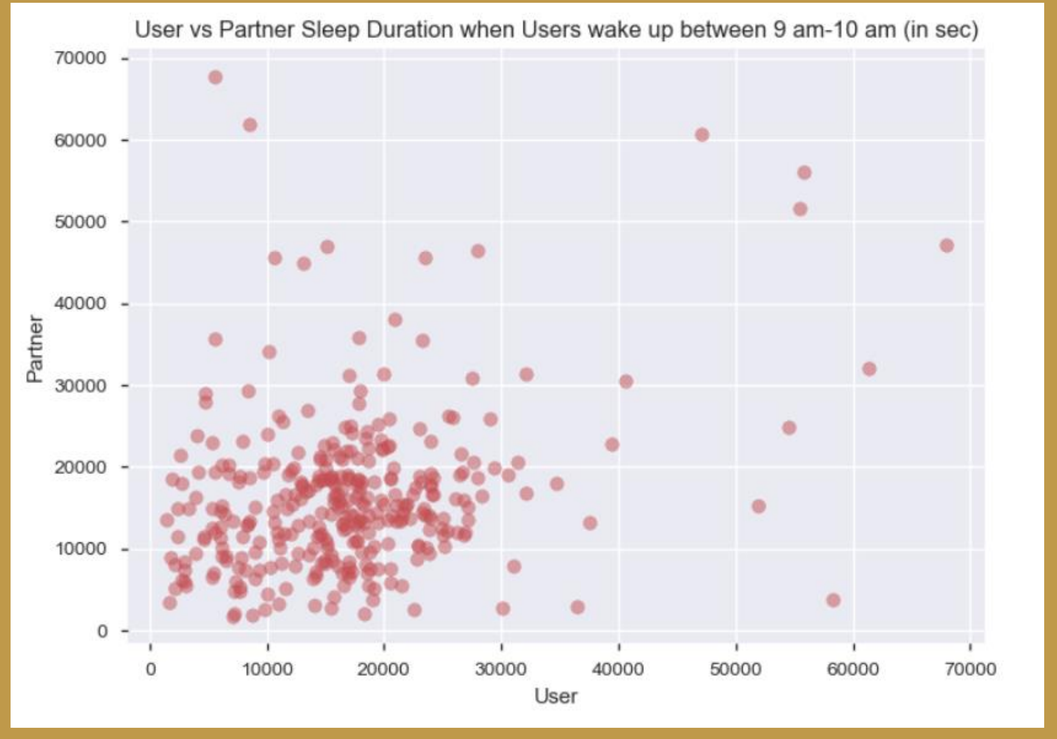
1. The plot on the right shows that for majority cases, male users and their female partners both turn 30 times or less during sleep.

2. The scatter plot below (left) shows that for users going to bed at 11 am in the morning, users sleep much longer than their partners

3. Male users and Female partners waking up between 9am and 10 am (which is when the largest number of users wake up) sleep fairly evenly. Mostly, both users and partners sleep at most 30,000 seconds.

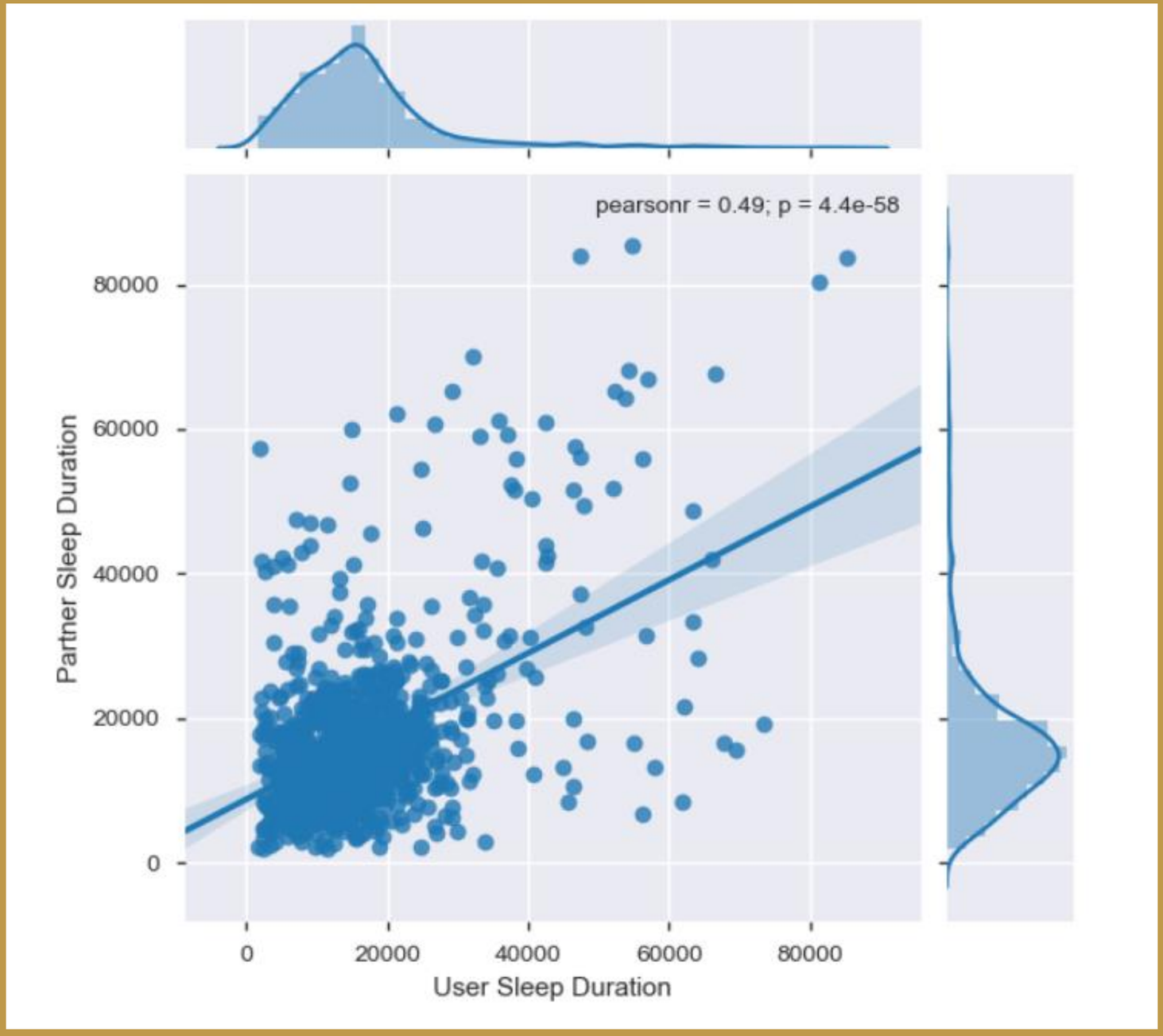






User & Partner Sleep Duration

- This regression cum scatter plot fits a straight line to the User and Partner Sleep Duration behaviors.
- Generally, as User sleep duration increases, the duration of partners also increases.



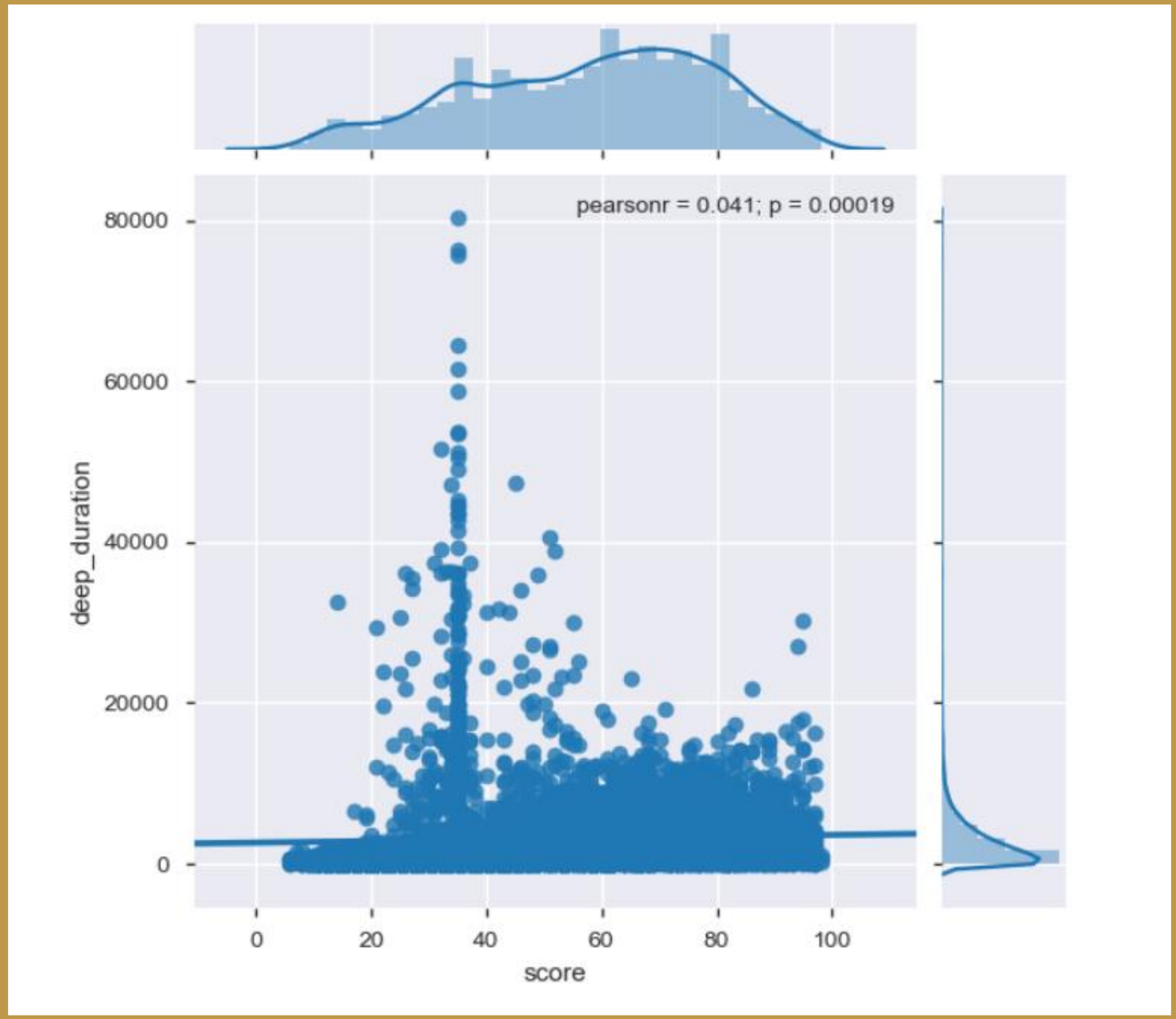
3

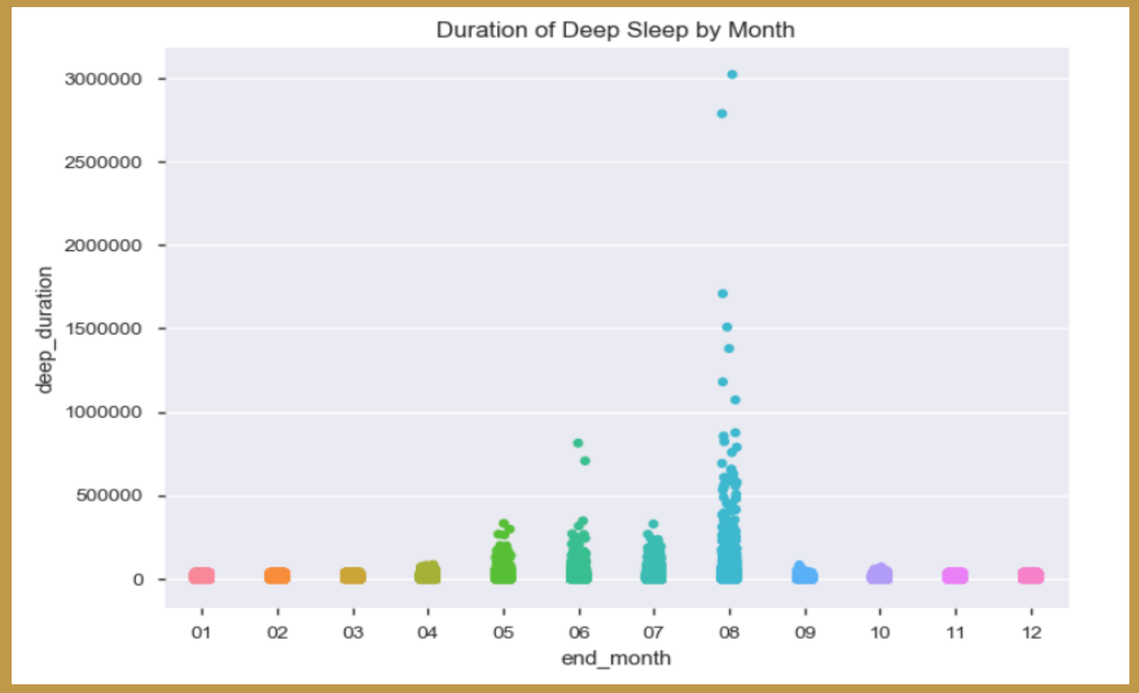
Correlation Between Other Features

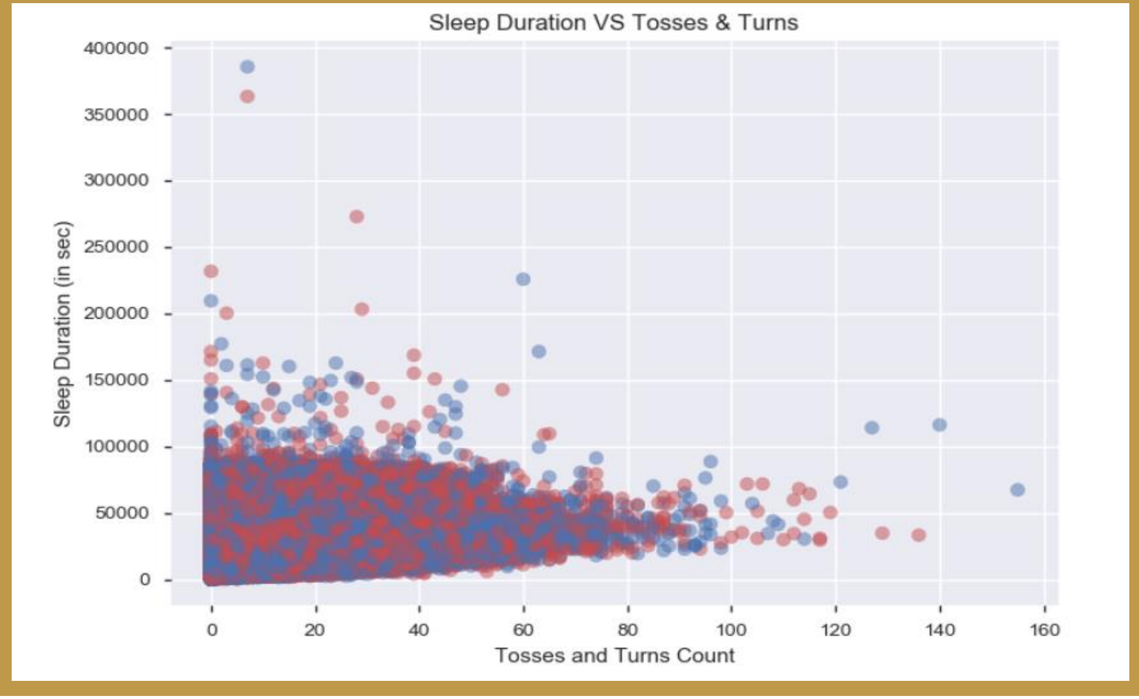
Deep Sleep Duration VS Score

- How users' deep sleep duration affects their score is captured in this scatter plot.

- Although it mostly follows a linear trend, there seems to be a lot of users with a score of 35 that have much longer durations of deep sleep.







Duration of Deep Sleep Analysis

1. The first plot shows the duration of deep sleep users get every month of the year
    - What is strikingly clear is that in the month of August (08), users typically get a very longer period of deep sleep.
    - Also, deep sleep is usually longer during the summer months of May, June and July in comparison to rest of the months (except August)
2. The second scatter plot gives a sense of how the tosses and turns of users varies with their sleep duration.
    - It can be said that almost all users sleep for at most 100,000 seconds and majority of the users toss/turn at most 60 times in their sleep.

Tools Used:

- Python
- Pandas and Numpy packages (Python)
- Matplotlib package (Python) for visualization
- Seaborn package (Python) for visualization

Thank You!