Moon Phases Impact On An Individual’s Sleep and Mood

By: Abigail Fabella and Rose Militante

**Sources of Data**

* Kaggle.com (<https://www.kaggle.com/aroojanwarkhan/fitness-data-trends>)
  + Fitness Trends Dataset (25.csv)
* Moon Phases Data scraped from <https://www.calendar-12.com/moon_calendar>

**Goal**

* Analyzing data from fitness trends in relation to moon phases to determine if the moon phases had any effect on an individual’s amount of sleep, step count, weight, and mood.

**Data Transformation**

The fitness/sleep CSV file featured recorded data of an individual's step count, mood, calories, hours of sleep, level of activity and weight(kg) on a daily basis through a time frame of 4 months (October 2017- January 2018).Weight was originally given in kilograms so a column was added in excel to convert weight in kilograms to pounds. During the data cleaning process the CSV file was uploaded onto pandas data frame. Two columns were removed: the ‘bool\_of\_active’ and ‘weight\_kg’. Once the table was cleaned it was then converted back into CSV file. A database was then created in PgAdmin called sleep\_db and the fitness\_sleep table was created and the csv file was then imported into PgAdmin. The column ‘date’ was chosen as the primary key for the table.

The moon phases data spanned through a timeframe of 4 months (October 2017 - January 2018) matching the time frame of the fitness/ sleep data. Since the moon phases were depicted on separate web pages, a separate pandas dataframe initially held the scrapped dates and moon phases for each of those months. During the data cleaning for months November - January, the first few dates scrapped were from the previous month due to the first day of that month not falling on the left-most ‘Sunday’ column. These dates needed to be removed from the list holding the dates so that the correct moon phases aligned with the right dates when joined together. Once the tables for each month containing moon phases were cleaned, all 4 tables were joined into a single dataframe which was pushed through to the database using a sqlalchemy connection.

Once both tables were uploaded onto Pgadmin, tables were inner joined by the date column where it was the primary key in the fitness\_sleep table and the foreign key in the moon\_phases table. By doing so, we were able to have the data side by side or have certain data points from each table to aggregate columns to analyze data such as frequency of moon phases, maximum and minimum hours of sleep, average mood and hours of sleep.

**Final Production Database**

We utilized a SQL (PgAdmin4) relational database titled sleep\_db containing two tables. The fitness\_sleep table holds the data from an individual’s tracked data (date, sleep hours and mood) over the span of 96 dates. The moon\_phases table holds the moon phases scrapped from October 2017 - January 2018. SQL (PgAdmin4) was chosen because it allowed the data to be inner joined at the common column (date) and be aggregated.

**Observations**

During recorded dates of having 2 hours of sleep, the least amount recorded, the moon was in its waning gibbous moon phase. 9 hours was the most amount of sleep recorded and the moon was in its last quarter phase. In reviewing the next least/greatest hours of sleep (3 and 8 hours respectively), there was not an observed single moon phase that correlated with these hours. Therefore moon phases does not have a direct correlation with an individual's sleep quantity.

When reviewing the happiest mood records during the time frame, the moon was in various phases for each day the person was at their happiest. Mood values were 100 being the lowest and 300 was the highest (happiest). Moon phases when the fitness records were at the happiest mood (300) showed no single related moon phase. The records showed a variety of moon phases were true for multiple instances of a happy mood. Therefore, there is no direct correlation between the person’s mood and moon phase.