# **Project Report:** Structured Approach to Data Analytics

## **Introduction**

This project aims to help me gain practical experience with the structured approach to data analytics. By applying this approach to a personal dataset, the goal is to gain a deeper understanding of the process and its core principles. Data analytics is a problem-solving methodology that leverages data to uncover insights and drive data-driven solutions. In this specific project, I focused on analyzing and potentially improving my daily step count. By exploring personal step data through the six phases of data analysis (Ask, Prepare, Process, Analyze, Share, and Act), the project will provide valuable insights into my activity levels and help develop strategies for increasing my daily steps.

## **1. Ask**

There is growing concern about the negative health effects associated with a sedentary lifestyle. Research suggests that sitting for extended periods can contribute to a higher risk of chronic diseases, including heart disease, type 2 diabetes, and some cancers [1]. The commonly cited goal for daily steps is 10,000, often seen as the "standard" number. However, it's not a one-size-fits-all recommendation [2].

In my current role as a virtual tutor, I spend most of my workday (9am-5pm) sitting at a desk. This sedentary lifestyle has me questioning my daily step count and its potential impact on my health. This project will utilize a structured data analytics approach to analyze my personal step data.

A successful outcome for this project will be:

* **Understanding current activity level:** Gaining a clear picture of my average daily step count.
* **Identifying influential factors:** Determining which daily activities contribute to either higher or lower step counts.
* **Developing actionable insights:** Using the data to identify opportunities for increasing my daily step intake and potentially incorporating more movement into my routine.

#### Key Questions

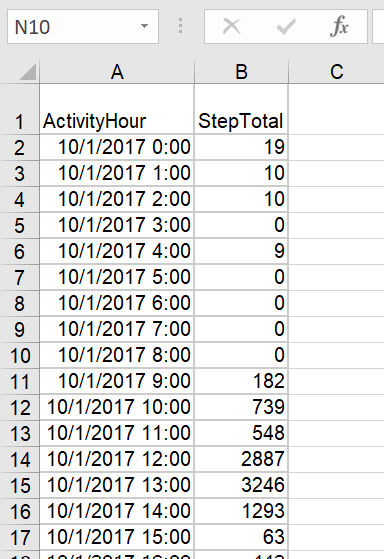
* What activities throughout my day contribute to fewer steps? (e.g., work hours, screen time)
* What activities throughout my day contribute to more steps? (e.g., errands, exercise routines)
* How does my current daily step count compare to recommendations for healthy activity levels?
* Do I have any existing planned walks or exercise breaks throughout the day?

By answering these questions, I can develop strategies to increase my daily step count and improve my overall health.

## **2. Prepare**

In this phase, I focused on utilizing readily available data to expedite the project timeline. The provided dataset contained two relevant columns:

* **ActivityHour:** This column captures the date and time for each data point, represented as a DateTime data type.
* **StepTotal:** This column holds the numerical values for the total number of steps taken during each recorded hour.

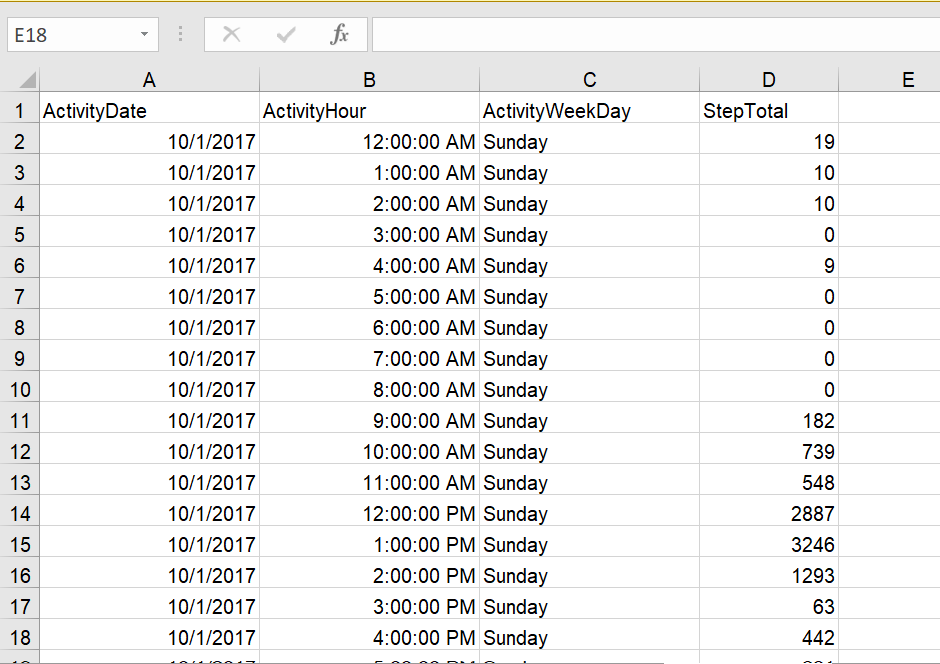


Since the data was pre-curated, the "Prepare" phase primarily involved familiarizing myself with the data structure and ensuring its suitability for the analysis.

## **3. Process**

In the "Process" phase, I focused on manipulating the existing data to create new features and prepare it for analysis. Here's a breakdown of the steps I took:

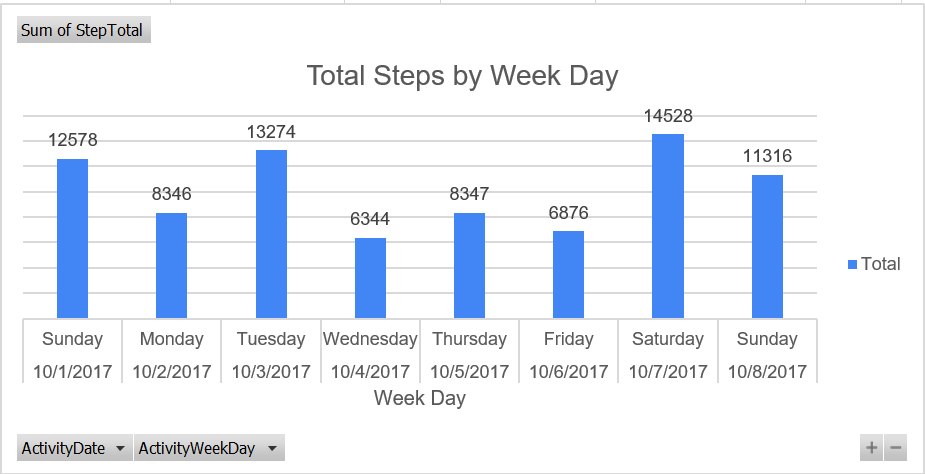
* **Data Familiarization:** I started by getting acquainted with the overall structure of the data. This involved checking the number of rows and columns (192 rows and 2 columns initially), along with data types for each column (DateTime for "ActivityHour" and numerical for "StepTotal").
* **Feature Engineering:** I identified opportunities to extract additional features from the existing "ActivityHour" column. This resulted in creating three new columns:
  + **ActiveDate:** This column extracts the date portion from the "ActivityHour" column.
  + **ActiveTime:** I renamed the second extracted column "ActiveTime" to avoid confusion with the original "ActivityHour" column. It extracts the time portion from "ActivityHour".
  + **ActiveWeekDay:** This column extracts the day of the week from "ActivityHour" using a combination of the TEXT and WEEKDAY functions in Excel.
* **Data Cleaning:** I was happy to find that the initial data inspection revealed no missing values or duplicates. This is a positive sign, indicating the data is ready for further analysis without the need for additional cleaning steps.
* **Summary Statistics:** I extracted key metrics like the total number of steps taken (81609) across the 8 days in the data. This provides a high-level overview of my activity levels.



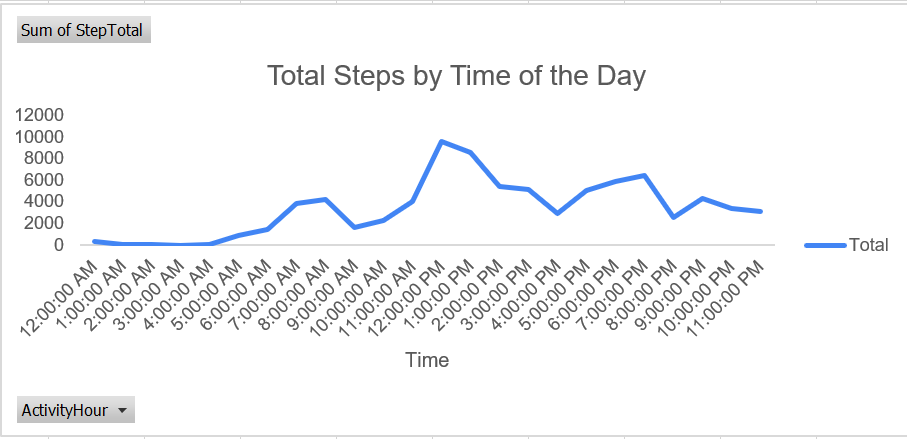
## **4. Analyze**

In the "Analyze" phase, I leveraged pivot tables and pivot charts to gain insights from the processed data. Here's a breakdown of the analyses performed:

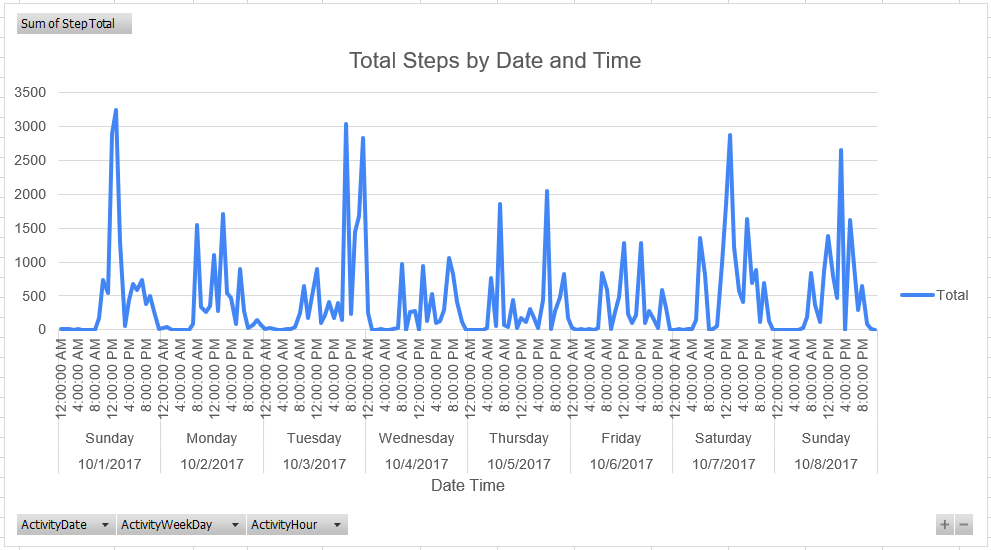
* **Total Steps by Day**
  + I created a pivot table to aggregate (sum) the total steps for each "ActivityDate".
  + To visualize this data, I used a pivot chart with a bar graph format.
  + This likely involved setting the "StepTotal" on the y-axis and the "ActivityDate" on the x-axis. This bar graph helps identify potential trends or variations in daily step counts.



* **Total Steps by Hour**
  + I created a separate pivot table to analyze the total steps taken per hour of the day.
  + Using the pivot table functionality, I then generated a pivot chart likely in the format of a line chart.
  + Here, the "StepTotal" would be placed on the y-axis and the "ActivityHour" on the x-axis. This line chart helps visualize the typical hourly distribution of steps throughout the day, potentially revealing patterns related to work hours or leisure activities.



* **Detailed Breakdown by Day and Hour**
  + I created a more granular pivot table to explore the data by both "ActivityDate" and "ActivityHour". Additionally, "ActivityDay" might have been included to provide a more comprehensive breakdown.
  + A corresponding pivot chart, likely a line chart, was created to visualize this detailed breakdown.
  + This chart allows for a deeper understanding of how step patterns vary across days and times of the day. For example, it might reveal if there are specific days or times associated with higher or lower activity levels.



By employing these pivot tables and pivot charts, I was able to extract valuable insights from the data, which will be further explored in the "Share" and "Act" phases.

## **5. Share**

The "Share" phase focuses on communicating the findings from the analysis. Here's a breakdown of the insights I obtained from the pivot tables and pivot charts:

**Daily Step Count:** The bar graph revealed a trend of higher step counts on weekends (Saturday and Sunday) and Tuesdays because these are the days I do not go to work. This suggests that I engage in activities that involve more walking on weekends, potentially due to morning walks or errands. On weekdays, particularly during my work hours, sitting seems to be a dominant activity, leading to lower step counts.

**Step Distribution by Hour:** The "Total Steps by Hour" plot indicated that most of my steps occur during the day, with a peak at 12 pm (9599 steps). This aligns with my typical schedule, where I'm mostly inactive between 12 am and 5 am (likely sleeping, except for occasional nighttime snack breaks). My overall activity seems to start around 6 am, likely due to being a morning person, and continues until 11 pm.

**Detailed Breakdown:** Visualizing the data by date, day, and hour using the line chart revealed a potentially seasonal pattern or recurring trends. Weekends and Tuesday consistently showed higher step counts compared to other weekdays. This might be due to specific activities undertaken on those days.

#### Conclusions and Findings

The analysis of my daily step data provided valuable insights into my activity patterns. Weekends and Tuesdays seem to be associated with increased activity levels, possibly due to leisure activities or errands. Conversely, weekdays, especially during work hours, involve more sitting, resulting in lower step counts. The hourly distribution suggests most steps occur during waking hours, with a peak around lunchtime. Additionally, a potential seasonal or recurring pattern emerged, with weekends and Tuesdays showing higher step counts compared to other weekdays. By understanding these patterns, I can now move forward to the "Act" phase and develop strategies to increase my overall daily step count.

## **5. Act**

The "Act" phase is where you leverage the insights from your analysis to develop actionable plans. Here are the strategies I identified to increase my daily steps:

1. Prioritize Stairs over Elevators: This I think is a great first step, incorporating more movement into my daily routine at work.
2. Short Walking Breaks: Implementing short walking breaks every hour during my workday is an excellent strategy to interrupt prolonged sitting and increase my step count. I could use this time for stretching, walking around the office, or grabbing a coffee further away.
3. Explore my Surroundings: Take advantage of my lunch break or free time to explore my surroundings on foot. I could take a new route during my walk, or discover nearby parks or walking trails.

By implementing these strategies, I aim to increase my daily step count and improve my overall health and well-being. This concludes the structured approach to data analytics applied to my step data project. By following these phases, I gained valuable insights into my activity patterns and developed actionable plans to reach my goals.

# **Reference**

1. Sitting risks: How harmful is too much sitting? - Mayo Clinic: https://www.topresume.com/career-advice/the-implications-of-sitting-for-long-periods-of-time
2. Mayo Clinic:https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/10000-steps/art-20317391