**DSA 210 Term Project – Correlation Exploration Between My Step Counts and Weather Conditions**

**Can Kargın 32202**

**Motivation**

**Walking plays a pivotal role in my daily life. Regardless of my mood or the outside factors like weather, I prioritize walking and strive to increase my step count, as it benefits both my mental and physical well-being. For a long time, I wondered if factors like weather conditions influence my walking habits and overall step counts. Through this project, I aimed to explore any correlations between my steps and meteorological data. In the process, I gained valuable insights into my own behavior and how external factor ( in this project, weather elements) might shape my routine.**

**Data Source**

**Step Count Data : I used the Apple Health Export.xml file to extract my step count data. This file contains various health metrics such as step counts, active energy burned, flights climbed and so on. To focus specifically on my step counts, first I converted this file into csv format for easier analysis, afterwards, I filtered the converted file to include only step counts. This dataset shows how much I walked ( step counts ) during specific hours of the day and, in turn, how much I walked each day.**

**Weather Data : During most of the summer and in this fall term, I was at Sabancı University, so I focused on the weather data for its specific coordinates. The dataset was obtained from Meteostat, which is an open-source weather data API, and includes various metrics such as temperature (temp), dew point (dwpt), relative humidity (rhum), wind direction (wdir), wind speed(wspd), and so on. However, I only used wind speed and temperature. This CSV file provides hourly measurements, allowing me to calculate daily averages of these two metrics.**

**Data Analysis**

1. **Data Preprocessing and Cleaning:**

* **Datetime Parsing : I converted time columns to datetime format for easier manipulation and analysis. Also, I removed timezone information to standardize timestamps.**
* **Outlier Detection : I used the IQR method to identify and flag outliers in step count data.**
* **Feature Engineering: I extracted components such as Date, Month, Hour and from datetime columns for aggregation and seasonal analysis.**

1. **Data Visualization:**

**In this project, line plots and bar charts were used in order to visualize daily and hourly average temperature, average wind speed, step counts. So that, temporal patterns could be understood.**

**In this project heatmaps and scatter plots are used in order to visualize relationships between daily average weather variables ( wspd and temp) and step counts.**

**In this project, I used histogram(s) in order to identify the frequency of related variable(s) such as temperature and wind speed.**

1. **Correlation Analysis :**

**I computed correlation coefficients investigate / measure the strength of relationship between :**

1. **Wind Speed – Temperature**
2. **Wind Speed - Step Counts**
3. **Temperature – Step Counts**
4. **Exploratory Data Analysis ( EDA ):**

* **I used statistical descriptions such as mean, median, standard deviation, variance, minimum, and maximum values for temperature, wind speed, and step counts in order to summarize and understand the key characteristics of a dataset.**
* **I did distribution analysis by visualizing the temperature, wind speed and step counts through the use of histograms.**
* **For time series analysis, I resampled data to compute daily, weekly, and monthly averages for temperature, wind speed and totals for step counts. Furthermore, I created time series plots to observe temporal trends.**

1. **Statistical Analysis :**

* **For correlation analysis, I computed pearson correlation coefficients between weather variables and step counts to quantify relationships.**
* **For seasonal decomposition, I decomposed daily average temperature and wind speed time series into trend, seasonal and residual components in order to figure out periodic behaviors.**

1. **Integration of Multiple Datasets – Merging :**

**For integrated analysis, I combined the datasets containing weather variables ( wind speed and temperature with step count data / dataset by matching them based on their corresponding dates.**

**Findings**

**Through the analysis and thanks to this project, I discovered the following :**

* **Relationship between my step counts and temperature:**

**The correlation coefficient between the two variables was calculated to be -0.0222 which indicates there is essentially no meaningful linear relationship between daily average temperature and daily step counts. The correlation remains weak.**

* **Relationship between my step counts and wind speed:**

**The correlation coefficient between the two variables was calculated to be -0.08 which indicates there is a weak negative relationship between daily average wind speed and daily step counts.**

* **Key Insights About Myself:**

**1) My walking habit is independent from weather metrics.**

**2) A significant portion of my steps were taken during the evening ( 6 PM – 11 PM) which indicates I am really active during this time.**

**3) Some months have notably higher step counts than others. For example, in October, I had the highest activity.**

**4) I walk an average of 6159 steps daily.**

**5) On more active days, I exceed 8343 steps daily.**

**Limitations**

**Limited Variables : This project examines only temperature and wind speed, while other weather variables such as humidity and precipitation that can also influence walking habits are not included in this analysis.**

**Contextual Factors : Non-weather related factors such as holidays, work schedules, personal health are not accounted for, which potentially limit the accuracy of the conclusions.**

**Correlation vs. Causation : This analysis identifies correlations but does not establish causation between weather conditions and step counts.**

**Future Work**

* **For more advanced modeling, I plan to use machine learning algorithms ( e.g., Random Forest ) to capture non-linear relationships. I also intend to apply clustering techniques to group similar days based on both weather conditions and activity patterns. Also, with the use of these algorithms, I plan to uncover hidden patterns or relationships in the data.**
* **For behavioral insights, I plan to examine how step counts vary between weekends, weekdays, and holidays.**
* **For health insights, I plan to integrate step counts with additional metrics such as heart rate and calorie burn to examine how weather conditions affect overall well-being.**
* **Lastly, I plan to include additional weather variables—such as humidity, precipitation, and UV index—to explore a broader range of weather-related impacts.**

**Conclusion**

**Overall, this project aimed to investigate whether weather conditions influence my step counts – and, by extension, my walking habits. Although the analysis did not reveal a significant correlation, and thus did not support my alternative hypotheses about temperature and wind speed, the experience proved valuable in terms of data processing, visualization, and deepening my understanding of my personal walking patterns.**