Denizhan Altan
Sabanci ID - 29158

CS210- Introduction to Data Science- Course Project

Table of Contents

Project Topic: Weather Condition's Effect on my Activity Level

Project Hypothesis: Weather Conditions do affect my activity level significantly.

Denizhan Altan
Sabanci ID - 29158

Introduction – Motivation

The main motivation for undertaking this project was from my curiosity about the relations between environmental factors and human activity patterns. Specifically, my aim is to analyze in detail how different weather conditions influence the daily number of steps taken, as well as the amount of active and basal energy expended. It's recognized that even on days where the total energy consumption remains constant, the proportions of active and basal energy can vary significantly. This variance can be attributed to weather conditions, which have a profound impact on our lifestyle choices, habits, and overall physical activity. By this analysis, the project seeks to shed light on the potential behavioral adjustments made in response to the meteorological environment, thereby providing insights that could inform personal health and fitness strategies.

Denizhan Altan
Sabanci ID - 29158

Dataset Description

Data was collected from the Apple Health application on an iPhone, which consistently tracks user activity. This dataset consists of three separate CSV files, originally derived from Apple Health. The activity data spans from November 14, 2017, to October 30, 2023.

The first CSV file contains step count data. For analytical purposes, the total steps taken each day were aggregated. Additionally, three non-essential columns—device model, software version, and user information—were omitted to streamline the dataset.

The second and third CSV files pertain to active energy and basal energy, respectively. Similar to the first file, irrelevant columns in these datasets were also removed.

Moreover, weather data was acquired through API calls from https://openweathermap.org/, and this information was compiled in JSON format. This additional data provides context to the physical activity records, potentially offering insights into the influence of weather conditions on user activity.

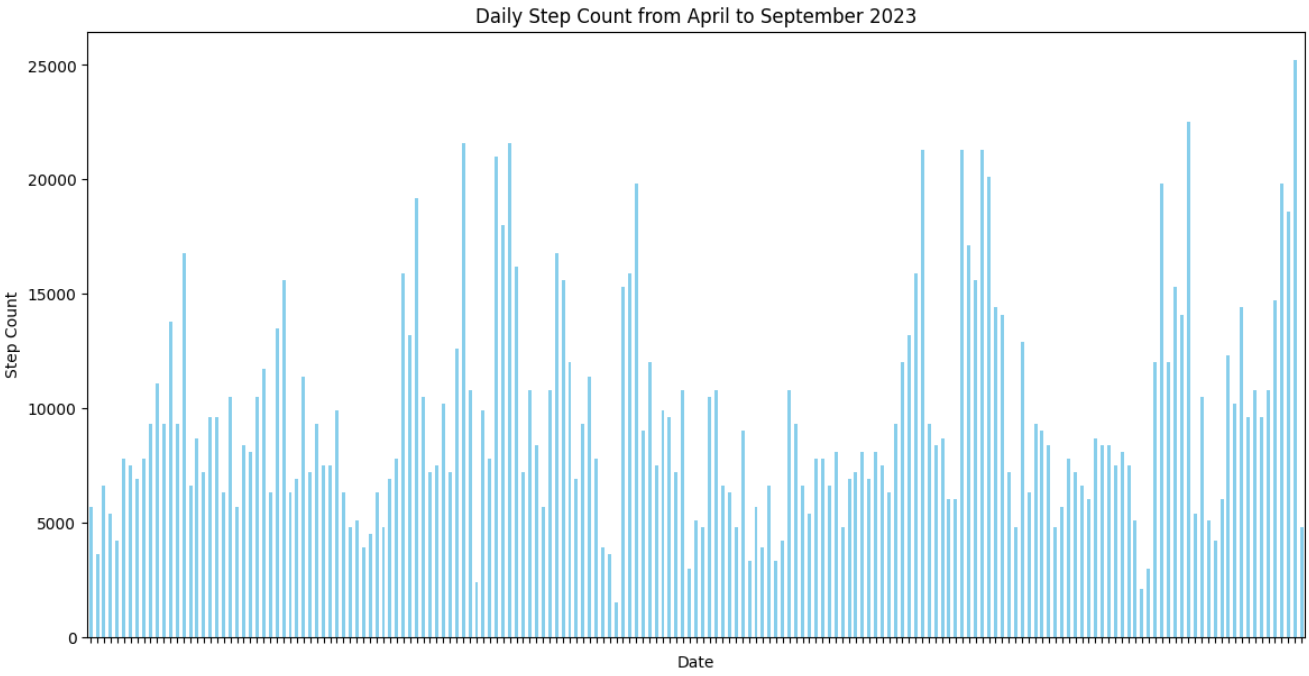
Here is the extracted data frame which includes all the relevant data:

	creationdate	Average Daily Active Energy (kcal)	Average Daily Basal Energy (kcal)	step_count	weather	weather_encoded
0	2023-05-01	157.553	1765.439	6300	Clear	0
1	2023-05-02	263.004	1832.577	6900	Clear	0
2	2023-05-03	436.911	1709.077	11400	Clear	0
3	2023-05-04	204.112	1862.228	7200	Clear	0

Denizhan Altan
Sabanci ID - 29158

Data Analysis

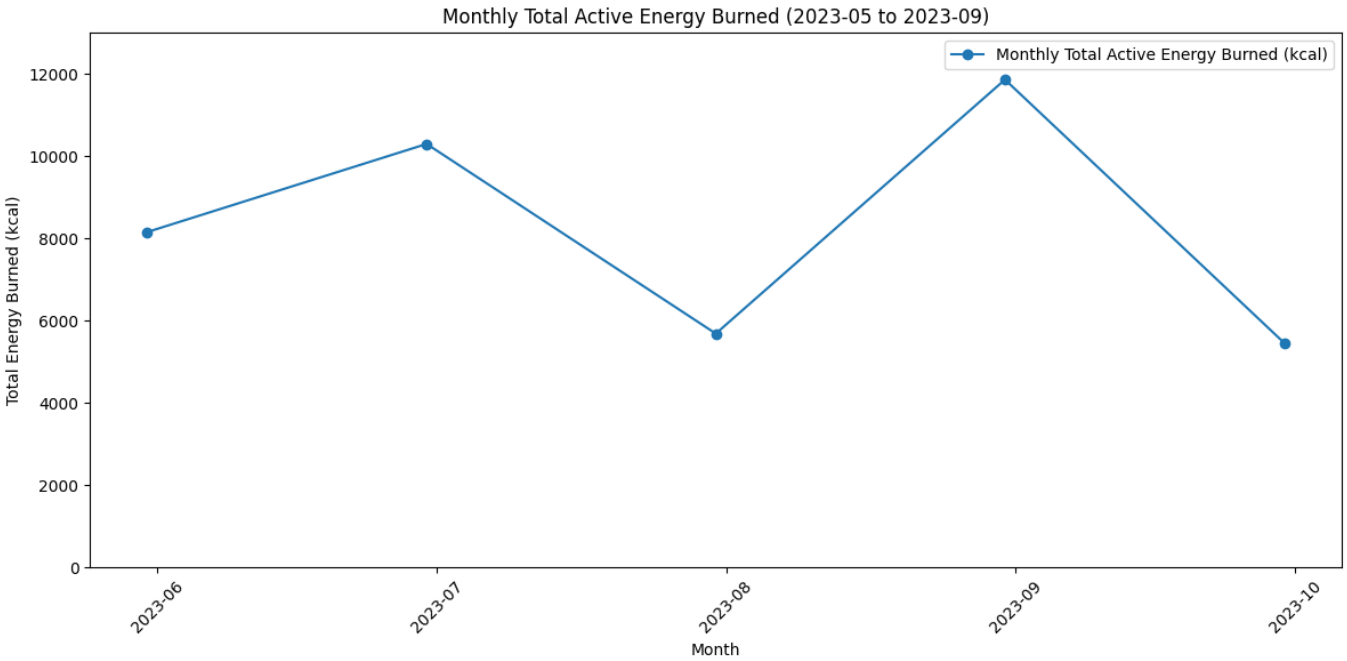
Daily Step Count From 01.04.2023 to 30.09.2023:



-The average daily step count between this period is 9529. The maximum step count for a day is 25200, whereas the minimum step count is 1500.

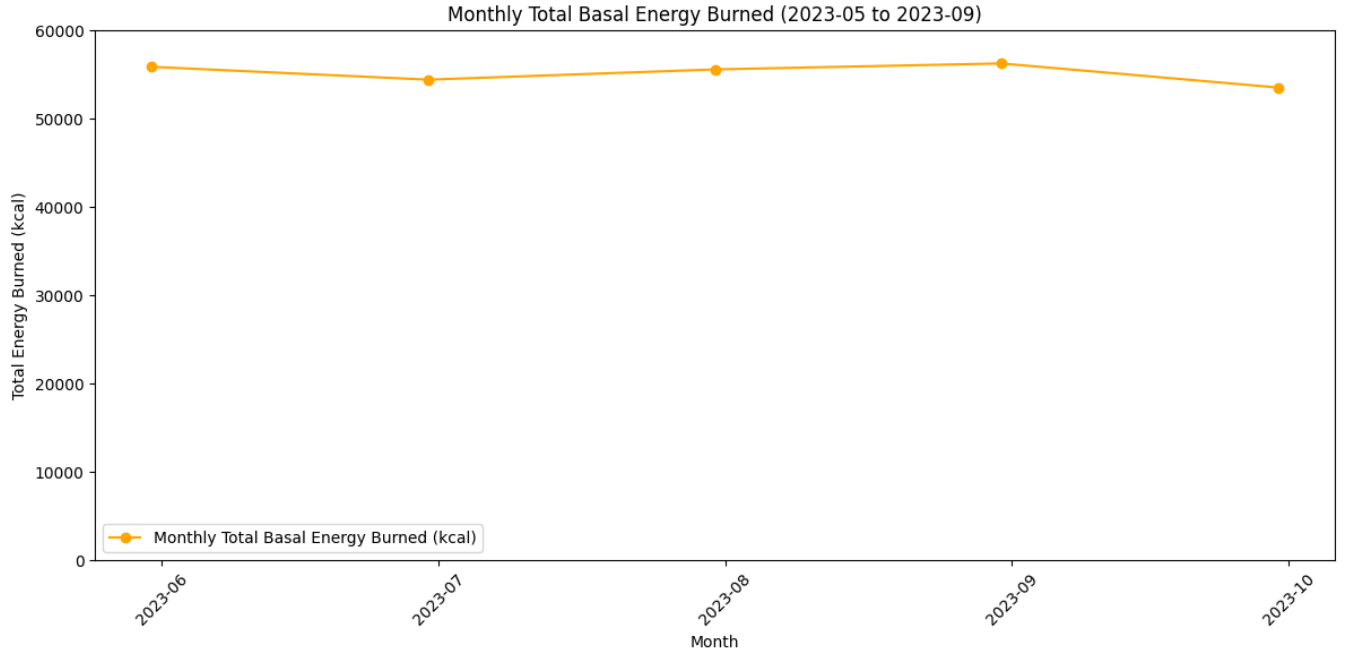
Denizhan Altan
Sabanci ID - 29158

Total amount of Active Energy burned in a Month:



-Active Energy: The energy that the user has burned due to physical activity and exercise.

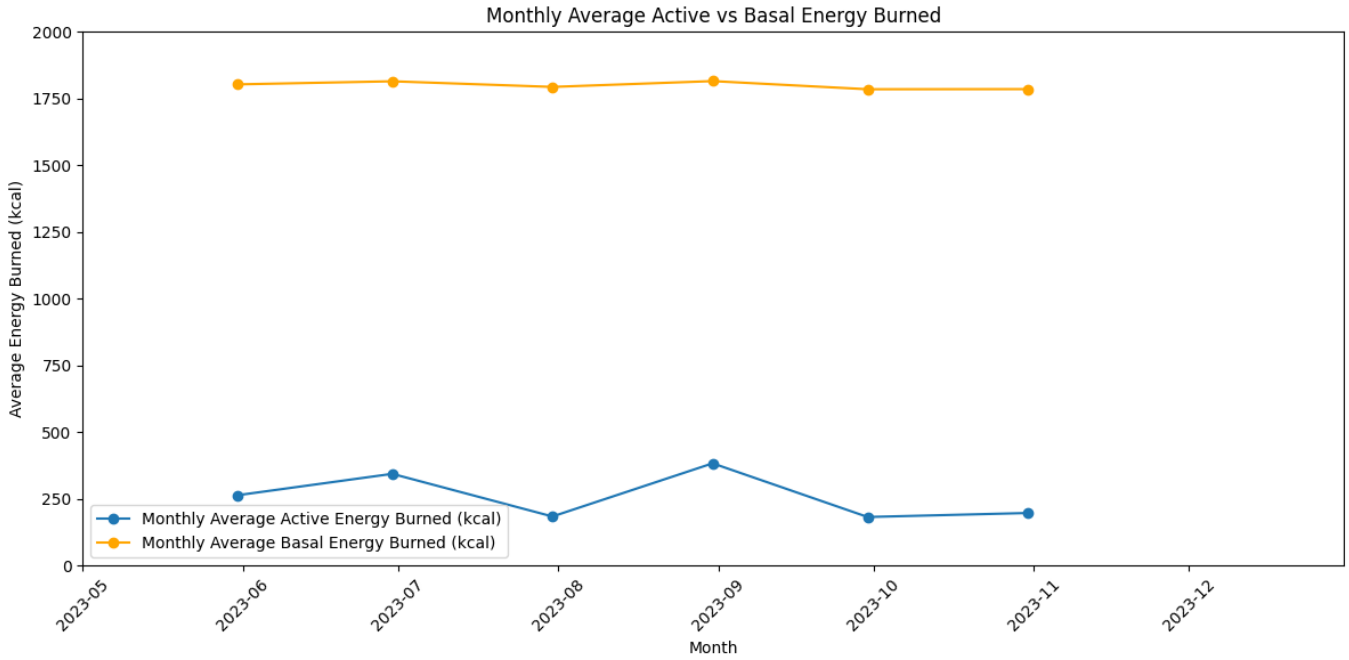
Total amount of Basal Energy burned in a Month:



-Basal Energy: The energy that the user has burned due to life sustaining functions.

Denizhan Altan
Sabanci ID - 29158

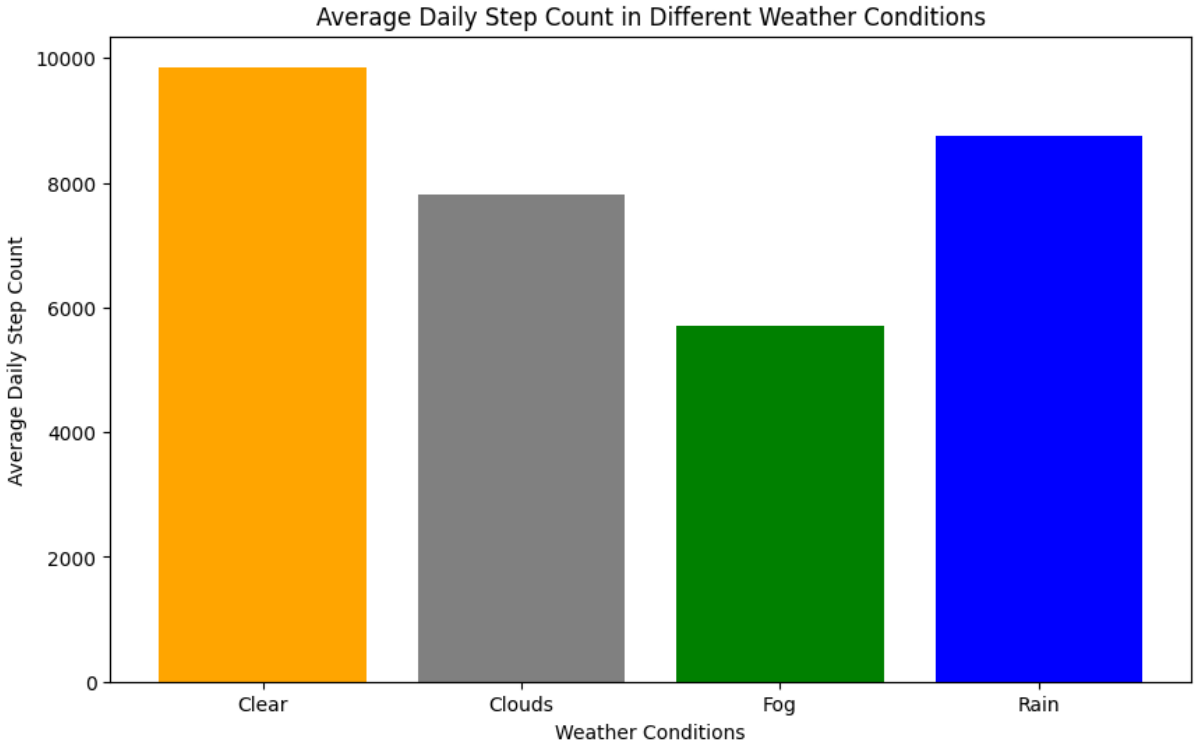
Comparison of single day average in each month for Basal and Active Energy burned:



-It can be seen from this plot that basal energy burned in a day is much higher than active energy burned in a day. This is normal due to basal energy being burned consistently through out the day whereas active energy is only burned while physical activities.

Denizhan Altan
Sabanci ID - 29158

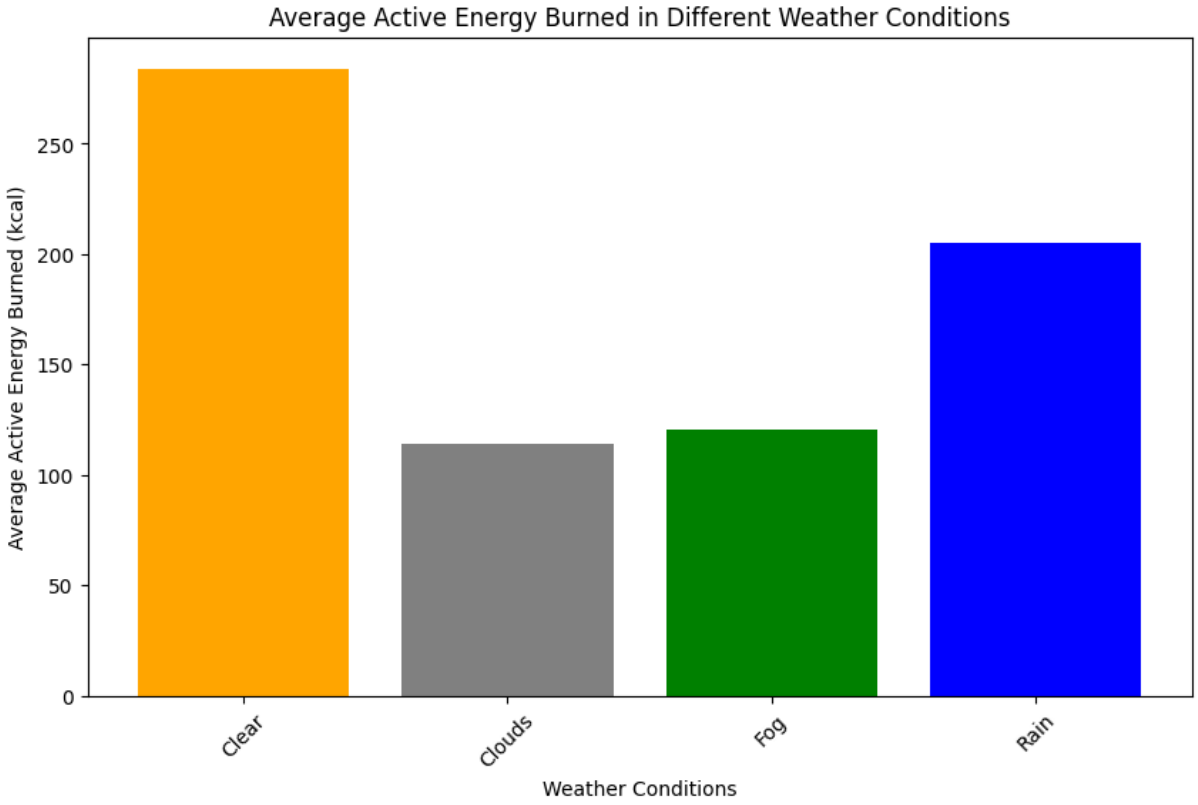
Average Daily Step Count in Different Weather Conditions:



-In this plot, we can see the average step count for a single day with the specified weather condition.

Denizhan Altan
Sabanci ID - 29158

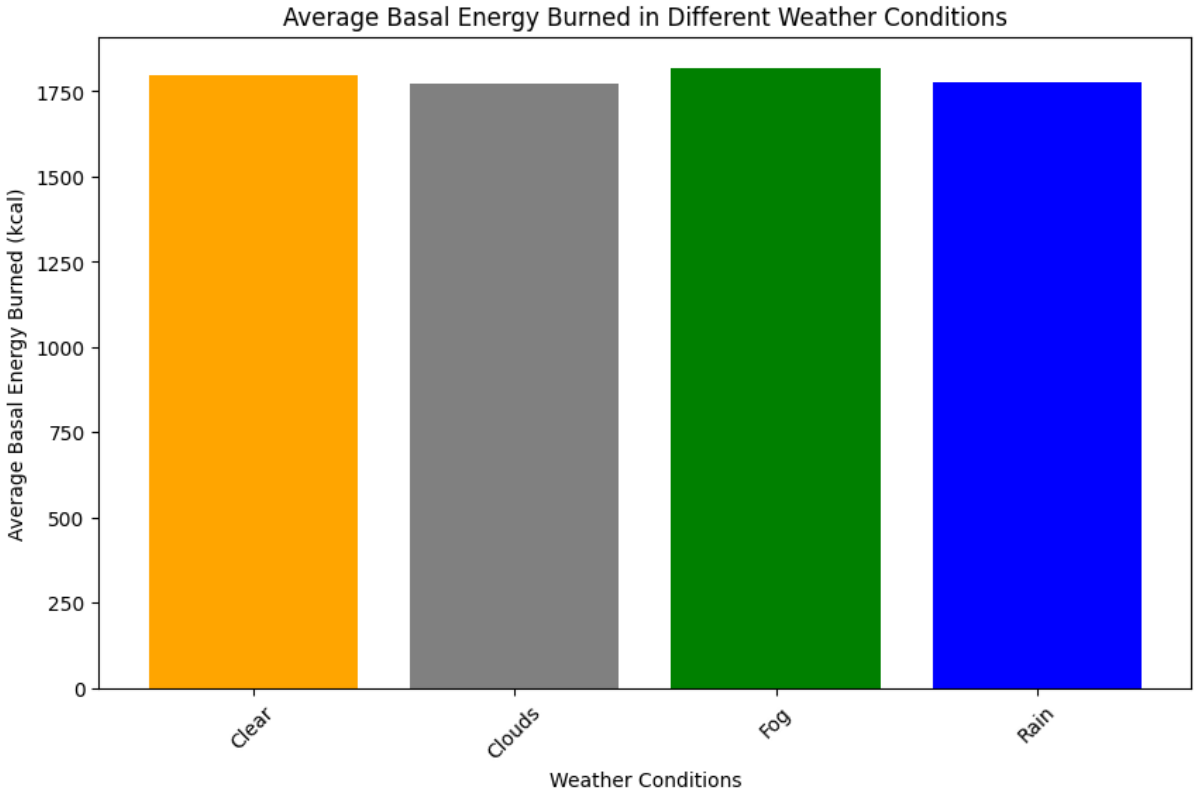
Average Active Energy Burned in Different Weather Conditions:



-In this plot, we can see the average active energy burned for a single day with the specified weather condition.

Denizhan Altan
Sabanci ID - 29158

Average Basal Energy Burned in Different Weather Conditions:



-It can be seen that in clear and rainy days, active energy burned is clear of the other two conditions, whereas basal energy burned is slightly higher for foggy days compared to other three conditions.

Denizhan Altan
Sabanci ID - 29158

Findings

From the daily step count plot, we know that average daily step count for the specified time period is 9529. To observe our hypothesis's truth, this number should be compared with the plot which shows average step count in different weather conditions. It is observed that when the weather condition is clear, the step count is significantly higher than the rest of the conditions. Step count with rain condition comes in second and is also significantly higher than the other two conditions. Just by looking at these results, it can be seen that weather conditions affected my step count and my activity level which is proportional to step count. These results can be a bit surprising, considering rainy weather can be seen as a sub-optimal condition for

Moving onto average "Average Active Energy Burned in Different Weather Conditions**"** graph, here we see a result different than expected. It is expected that if step count is higher, active energy burned value will also be higher. We can see this relation with clear and rainy weather conditions; however, it can be seen that active energy spent in foggy conditions is higher than cloudy condition. It is not possible for us to determine the specific reason for this case but foggy day number relatively being small can make this data susceptible to noise. Another reason for this can be related to human biology which can be the topic for a future project.

Moving onto the last graph named "Average Basal Energy Burned in Different Weather Conditions", it can be seen that all the values are similar and only basal energy burnt in foggy conditions is a little higher than the rest of the conditions.

In summary, step counts show that activity level significantly changes with different weather conditions. We used active and basal energy values to further analyze this statement, and it showed once more that our hypothesis, which was "My Activity Level changes significantly with different weather conditions." It is important to note that these only represent my own activity level and it can vary greatly with people who have difference preferences.

Denizhan Altan
Sabanci ID - 29158

Future Work

For future work, a larger data set must be gathered to eliminate any noise. This was only with my data, and it cannot be generalized for other people, which prevents further uses of this project. Trying this project with more people can help making this data more generalizable. One example for the usage of analyzing this data set in detail, with today's world becoming more and more digital, these generalized values can be used in forensic analysis to gather proof.

Denizhan Altan
Sabanci ID - 29158

Limitations

-Limited data set

-Accuracy of Apple Health Data