

This study analyzes the relationship between screen time (in hours) and daily step counts over a 30-day period. The analysis is performed using Python in the **Spyder IDE**. The data file is opened using a Tkinter-based file dialog to ensure flexibility during testing.

## **Data Source and Preprocessing**

- **Step Count Data:**
  - The step count data was retrieved from **Apple Health** in XML format.
  - The XML file was parsed using Python's `xml.etree.ElementTree` module.
  - Relevant fields such as `startDate`, `endDate`, and `value` were extracted, while unnecessary metadata (e.g., device details) was excluded.
  - The cleaned data was saved as a CSV file for further processing and eventually converted into an Excel file for analysis.
- **Screen Time Data:**
  - Screen time data was recorded manually and included in the same Excel file alongside step count data.

## **Libraries Used:**

1. **Pandas:** For data manipulation and analysis.
2. **Matplotlib:** For creating visualizations such as bar charts, scatter plots, and line graphs.
3. **Statsmodels:** For conducting statistical analysis, including regression modeling and hypothesis testing.
4. **Tkinter:** For file selection through a graphical interface.

## **1. Data Description**

The dataset contains daily records of screen time (in hours) and step counts. Below is a statistical summary of the data:

**Screen Time (Hours):**

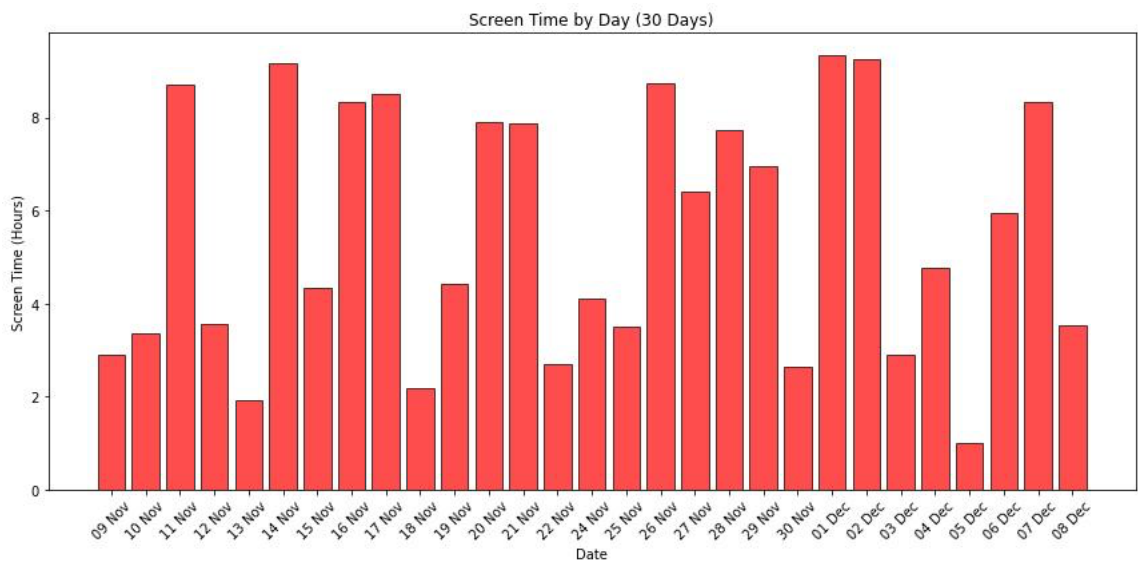
- **Mean:** 5.55 hours
- **Median:** Approximately 5.60 hours
- **Standard Deviation:** 2.69 hours
- **Minimum:** 1.2 hours
- **Maximum:** 9.5 hours

**Step Count:**

- **Mean:** 4895.62 steps
- **Median:** Approximately 4900 steps
- **Standard Deviation:** 2249.69 steps
- **Minimum:** 900 steps
- **Maximum:** 9800 steps

**2. Bar Chart Insights**

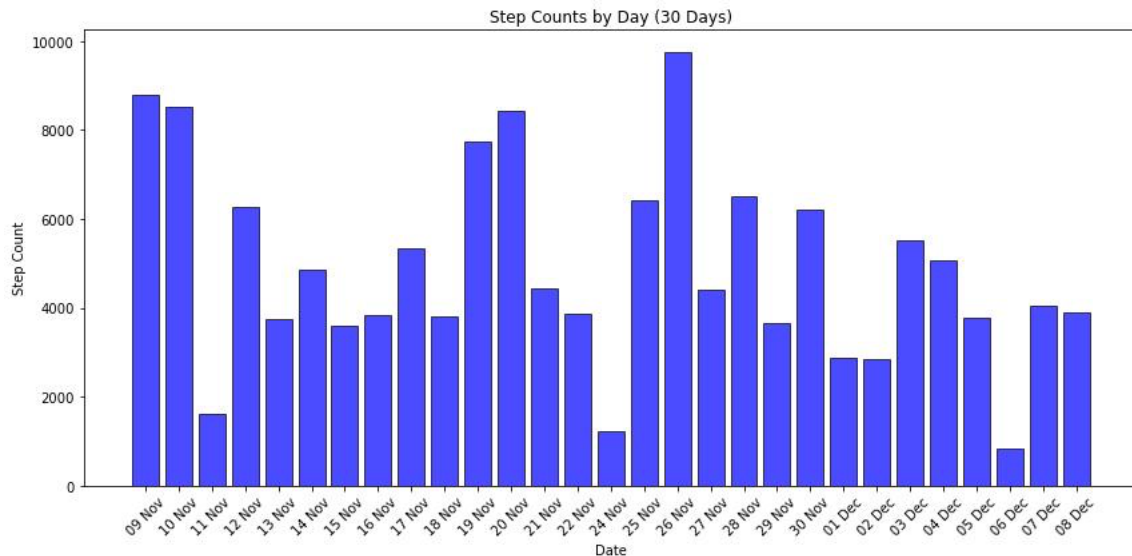
**Screen Time by Day:**



- **Observations:**

- Screen time fluctuates significantly between 1.2 hours and 9.5 hours.
- Peaks are observed on days 11, 14, 20, and 28.
- There are no consistent patterns of high or low screen time.

## Step Count by Day:

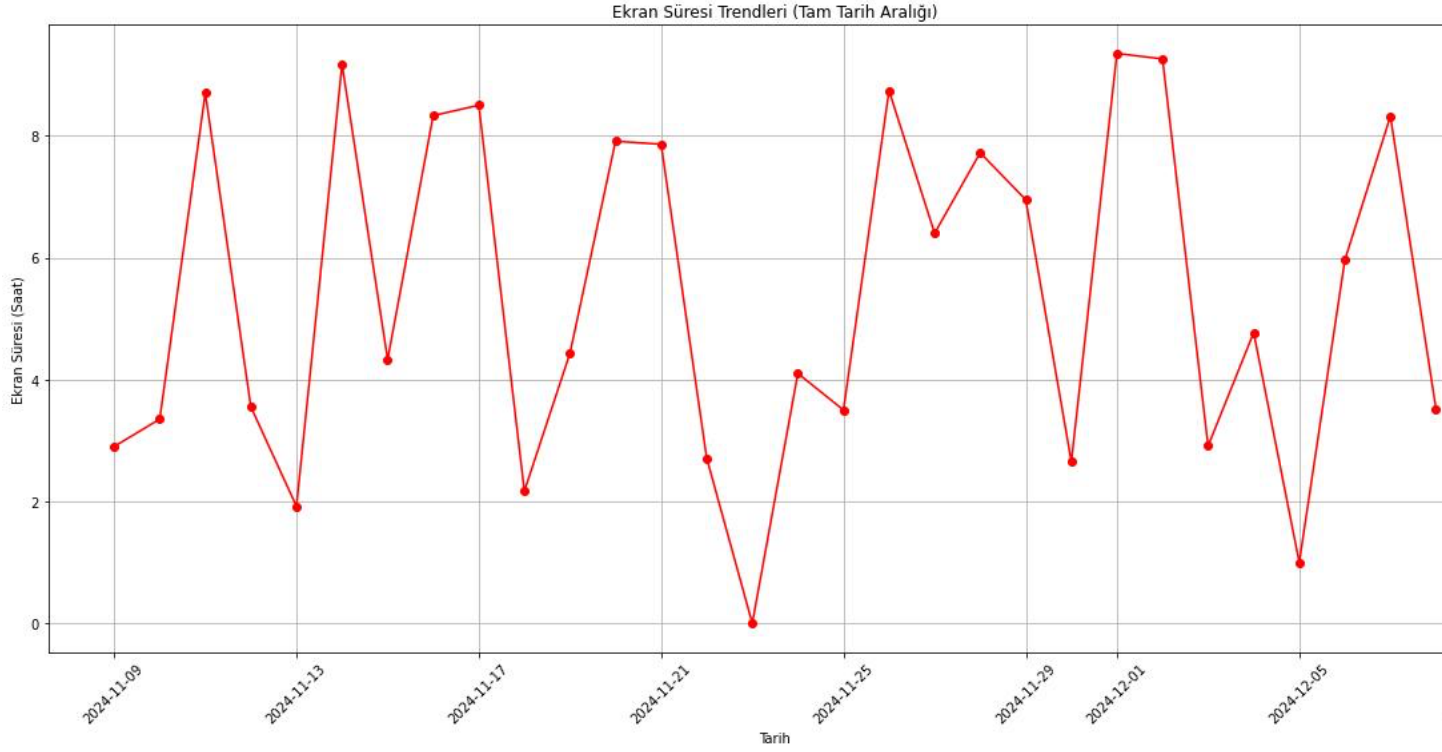


- **Observations:**

- Step counts vary between 900 and 9800 steps.
- Peaks are observed on days 9, 19, and 25.
- No observable trend connects step counts to specific days.

## 3. Line Graph Insights

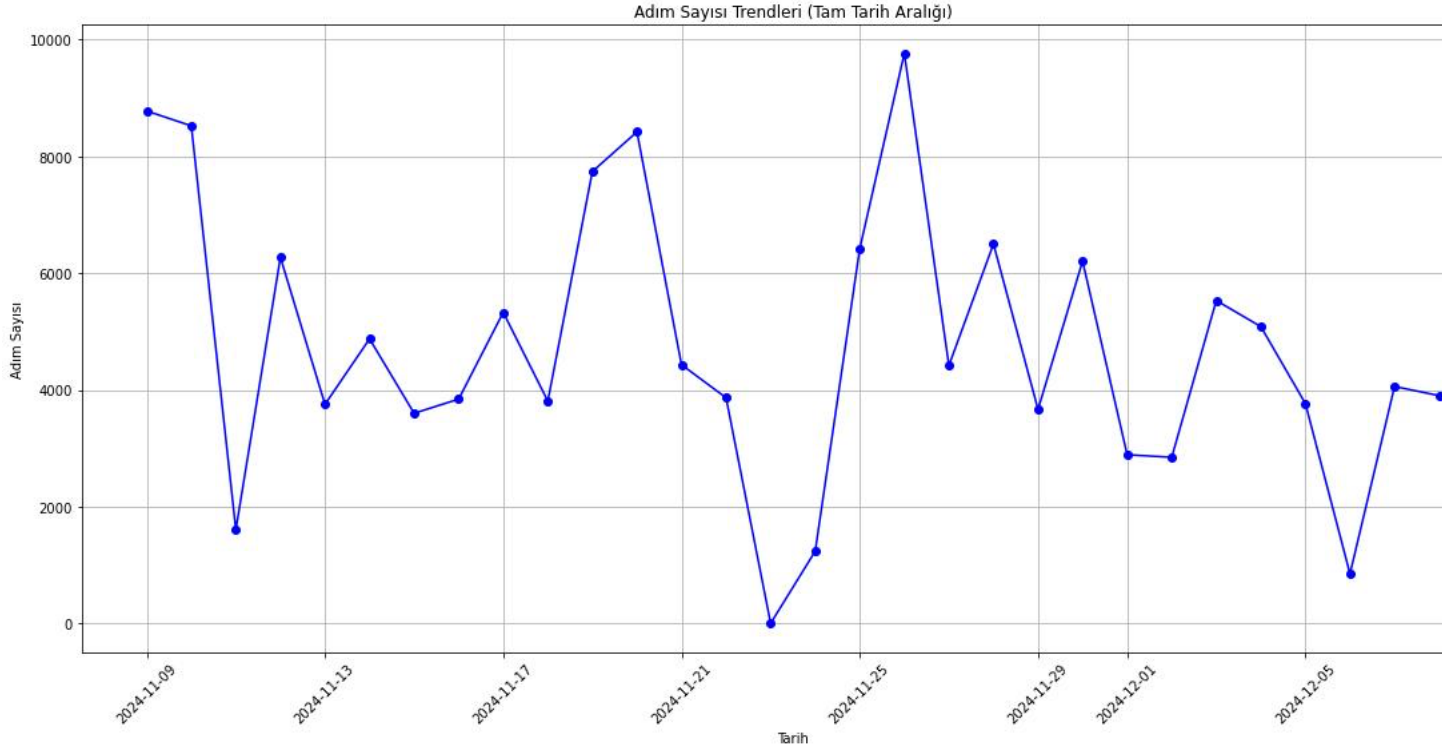
### Screen Time Trends:



- **Observations:**

- Screen time exhibits significant day-to-day fluctuations.
- The highest screen time (9.5 hours) occurs on day 28, while the lowest (1.2 hours) occurs on day 12.

## Step Count Trends:



- **Observations:**

- Step counts show notable variability across the 30 days.
- The highest step count (9800 steps) occurs on day 25, while the lowest (900 steps) occurs on day 12.
- There is no alignment between peaks and troughs in step count and screen time.

#### 4. Hypothesis Formulation

- **Null Hypothesis ( $H_0$ ):** There is no correlation between screen time and step count.
- **Alternative Hypothesis ( $H_1$ ):** There is a correlation between screen time and step count.

---

#### 5. Statistical Overview

- **Average Screen Time (Hours):** 5.55 hours
- **Standard Deviation (Screen Time):** 2.69 hours
- **Average Step Count:** 4895.62 steps
- **Standard Deviation (Step Count):** 2249.69 steps
- **Correlation Coefficient (Screen Time vs. Step Count):** -0.09
  - The weak negative correlation suggests that as screen time increases, step count slightly decreases, but the relationship is extremely weak.

## 6. Hypothesis Testing

- **P-value:** 0.6405
  - Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This indicates that there is no statistically significant correlation between screen time and step count.

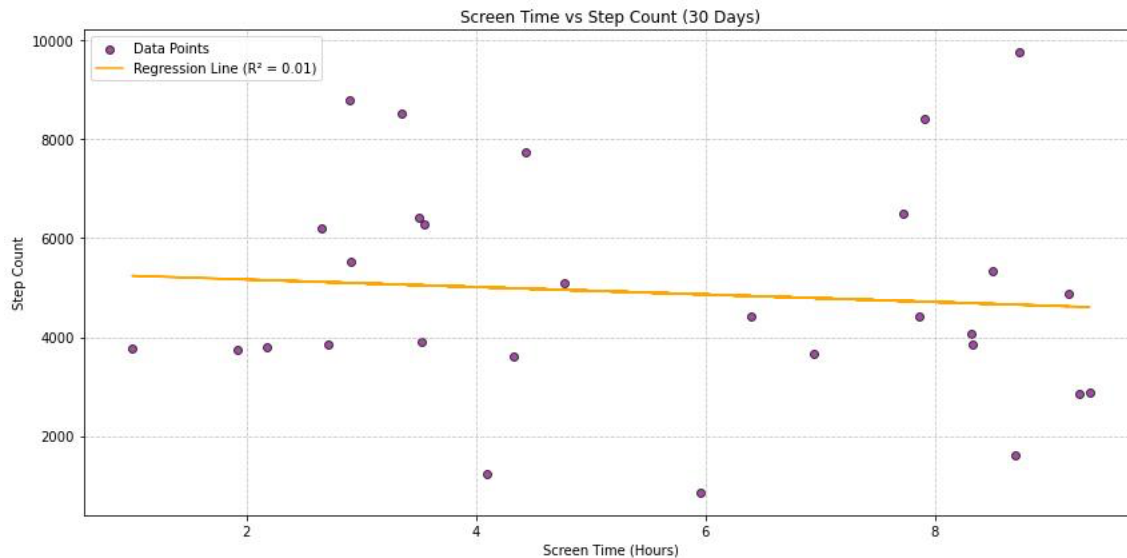
## 7. Linear Regression Analysis

### Regression Equation:

Steps = 5316.08 - 75.74 × Screen Time (Hours)

- **Intercept (Constant):** 5316.08 steps (predicted step count when screen time is 0).
- **Slope:** For every additional hour of screen time, step count decreases by approximately 75.74 steps.
- **R<sup>2</sup> Value:** 0.01
  - Only 1% of the variability in step count is explained by screen time, indicating a very poor fit.

## 8. Scatter Plot Insights



- **Observations:**

- The scatter plot shows data points with a regression line.
- The  $R^2$  value of 0.01 confirms the weak explanatory power of screen time in predicting step count.
- The data points are widely scattered, supporting the lack of a meaningful relationship.

## 9. Conclusion

- **Correlation Analysis:** The weak negative correlation (-0.09) and high p-value (0.6405) indicate no significant linear relationship between screen time and step count.
- **Regression Model:** The  $R^2$  value (0.01) demonstrates that screen time explains only a tiny fraction of the variation in step count.
- **Practical Implications:** Screen time and step count appear to be largely independent in this dataset. Other factors, such as exercise habits, weather, or personal schedules, may have a stronger influence on step count.