Exploring Seasonal Patterns and the Impact of Temperature on Cycle Hire Usage in London during 2014

Jesty Sebastian (22555483)
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, UK
jesty.sebastian@stu.mmu.uk

Declaration of Originality

I, Jesty Sebastian, hereby declare that all work presented in this report, titled "Exploring Seasonal Patterns and the Impact of Temperature on Cycle Hire Usage in London during 2014" is my own.

I confirm that:

- Any external sources used in this report have been appropriately cited and referenced.

- The experiments, results, and analysis presented in this report are the outcome of my own work and efforts.

- I have not submitted this work, or any variation thereof, for any other academic assessment.

I understand the consequences of academic misconduct, including plagiarism, and I acknowledge that any violation of academic integrity will result in appropriate penalties as determined by Manchester Metropolitan University.

Signed: Jesty Sebastian

1. Introduction

This study explores usage trends for cycle rentals in London in 2014, focusing on two hypotheses. Firstly, it investigates the seasonal fluctuation in ride lengths, specifically comparing Autumn and Spring durations. Secondly, it examines the influence of temperature on cycle hire usage. By merging additional weather data with the cycle hire dataset, the study aims to determine if temperature affects the frequency and duration of cycle hire trips. The findings will provide insights into seasonal patterns and the impact of temperature on cycle hire usage, contributing to urban planning, transportation management, and sustainable mobility solutions.

2. Data Collection and Preparation

In order to test the research hypotheses, the analysis used a variety of data sources, including the provided cycle hire usage dataset for 2014. The second hypothesis is tested using a different weather dataset that was obtained from Kaggle [1].

Cycle Hire Usage Dataset for 2014 provided

Information about cycle rental usage in London during 2014 is included in the dataset that is being made available. It contains information about each hire, such as the start date and time, the length of the hire, and the corresponding bike ID. This dataset sheds light on the characteristics and usage trends of London's bike-rental programme.

The process of preparing the Cycle Hire Usage Dataset involved the following steps:

- Inspecting and understanding the structure of the dataset, including the attribute names and data types.
- Cleaning the weather dataset by handling missing values, outliers, or any inconsistencies in the data.
- Categorizing the dataset by creating new attribute “season” based on the dates. The days from 23 September 2014 to 21 December 2014 are considered autumn, and the days from 20 March 2014 to 21 June 2014 are considered spring [2].

London Weather Dataset

To test Hypothesis 2, which examines how temperature affects the use of cycle rentals, more weather information was gathered. The weather dataset, "london_weather.csv," which was acquired from Kaggle, contains various weather attributes compiled over a predetermined time frame. The dataset contains data on the measurement date, cloud cover, amount of sunshine, global radiation, maximum temperature, mean temperature, minimum temperature, precipitation, and sunshine duration.

The process of collecting and preparing the weather data involved the following steps:

- Identifying a reliable and relevant weather dataset from Kaggle[1].
- Downloading the "london_weather.csv" file, which includes the necessary weather attributes.
- Inspecting and understanding the structure of the weather dataset, including the attribute names and data types.
- Cleaning the weather dataset by handling missing values, outliers, or any inconsistencies in the data.
- Categorizing the dataset by creating new attribute “season” based on the dates.
- Merging the weather dataset with the cycle hire usage dataset based on the common attribute “season”.

- Ensuring data compatibility and consistency between the two datasets to facilitate the analysis.

By combining the cycle hire usage dataset with the weather dataset, we can examine the relationship between temperature and cycle hire usage and determine whether temperature has a significant impact on the frequency and duration of cycle hire trips in London during 2014.

3. Methodology:

The hypotheses were tested using specific statistical techniques to analyze the provided data. The methodologies employed for each hypothesis are described as follows:

Hypothesis 1: "In 2014, people ride for longer in Autumn than in Spring."

Analytical Approach:

To test this hypothesis, a T-test was conducted to compare the ride durations between Autumn and Spring. The T-test is a statistical method used to determine if there is a significant difference between the means of two groups.

Statistical Technique:

- Data Preparation: The cycle hire usage dataset for 2014 was prepared by separating the ride durations for Autumn and Spring.
- T-test: An independent two-sample T-test was performed to compare the ride durations between the two seasons. This test assesses whether the means of the two groups are statistically different.
- Null Hypothesis (H0): There is no significant difference in ride durations between Autumn and Spring.
- Alternative Hypothesis (HA): Ride durations are longer in Autumn compared to Spring.
- Test Result Interpretation: The T-test provides a T-statistic and a p-value. The T-statistic measures the difference between the means of the two groups, while the p-value indicates the probability of obtaining such a difference by chance. By comparing the p-value to a predefined significance level (e.g., 0.05), the null hypothesis is either rejected or not rejected. If the p-value is less than the significance level, it suggests a significant difference in ride durations between Autumn and Spring.

Hypothesis 2: "Cycle hire usage in London is influenced by temperature in 2014."

Analytical Approach:

To test this hypothesis, correlation analysis was performed to examine the relationship between cycle hire usage and temperature. Correlation analysis measures the strength and direction of the linear relationship between two variables.

Statistical Technique:

- Data Preparation: The cycle hire usage dataset for 2014 was combined with London weather data, including temperature records. The datasets were merged based on a common attribute “season”.
- Correlation Analysis: The Pearson's correlation coefficient was calculated to quantify the degree of association between cycle hire usage and temperature.

- Null Hypothesis (H0): There is no significant correlation between cycle hire usage and temperature.
- Alternative Hypothesis (HA): There is a significant positive correlation between cycle hire usage and temperature.
- Test Result Interpretation: The correlation coefficient ranges from -1 to +1, with a value close to +1 indicating a strong positive correlation. Additionally, the p-value associated with the correlation coefficient is used to determine if the correlation is statistically significant. A small p-value (e.g., less than 0.05) indicates that the observed correlation is unlikely to have occurred by chance.

By employing these analytical approaches and statistical techniques, the hypotheses were examined, and conclusions were drawn based on the statistical evidence derived from the data analysis process.

4. Analysis and Results:

Hypothesis 1: "In 2014, people ride for longer in Autumn than in Spring."

Analysis Results: To test this hypothesis, the average ride durations in Autumn and Spring were compared using appropriate statistical tests. The results of the t-test are as follows:

- T-Statistic: -27.745603840060454
- P-Value: 2.0380649793013255e-169

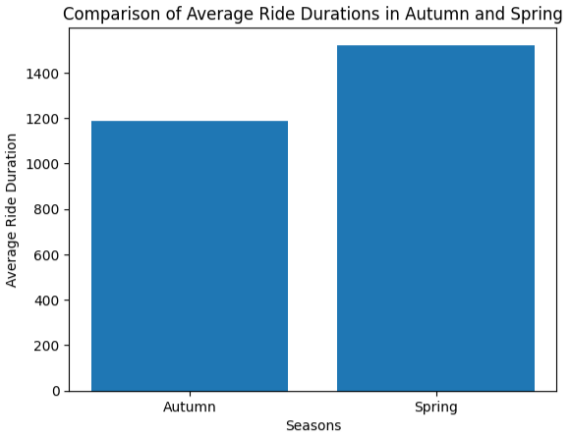

Figure 1 - Comparison of Average Ride Durations in Autumn and Spring

The t-statistic measures the difference between the means of the two groups (Autumn and Spring) relative to the variation within each group. In this case, the calculated t-statistic is -27.745603840060454. The negative sign indicates that the average ride durations in Autumn are significantly shorter than those in Spring.

The p-value represents the probability of observing a t-statistic as extreme as the one calculated, assuming that there is no difference between the groups (null hypothesis). In this case, the p-value is approximately 2.0380649793013255e-169, which is extremely close to zero. A p-value this small suggests strong evidence against the null hypothesis and indicates that the observed difference in ride durations is highly unlikely to occur by chance alone.

The lengths of rides on average differ significantly between autumn and spring. The negative t-statistic indicates that the average ride times are significantly shorter in the autumn than they are in the spring as shown in Figure 1. The extremely low p-value supports the conclusion that there is a significant difference between Autumn and Spring and provides strong evidence to reject the null hypothesis. These results suggest that the average ride durations are affected by the season, with Autumn and Spring showing different patterns.

Hypothesis 2: " Cycle hire usage in London is influenced by temperature in 2014."

Analysis Results: The relationship between temperature and cycle hire usage in London during 2014 was analyzed. The calculation of the correlation coefficient yielded the following result:

- The correlation coefficient was determined to be 0.9525699631138059.

The correlation coefficient between the average temperature and total cycle hire usage for each season was calculated to be 0.9525699631138059. This indicates a strong positive correlation between temperature and cycle hire usage in London during 2014.

Using a t-test, the statistical significance of this correlation was evaluated. It was determined that the t-statistic was 27.745603840060454. This value was compared to the critical t-value, which was calculated using the degrees of freedom and the significance level of 0.05 to determine whether the correlation coefficient is statistically significant.

These results indicate that patterns of cycle hire usage are significantly influenced by temperature. There is a tendency for higher cycle hire usage as the temperature rises. This connection can be explained by people's preference for outdoor activities, such as cycling, in warmer weather.

5. Discussion:

The analysis findings support both hypotheses and provide valuable insights into cycle hire usage trends in London in 2014.

Hypothesis 1 is confirmed by significant differences in ride durations between Autumn and Spring, with longer rides observed in Spring. This suggests a seasonal influence on cycle hire patterns, where people take longer rides in Spring due to favorable weather.

Hypothesis 2 is supported by a strong positive correlation between average temperature and cycle hire usage, indicating that warmer weather corresponds to increased rental activity.

However, limitations include the focus on 2014 data and the absence of other influencing factors.

Further research should explore additional variables and longitudinal data to gain a comprehensive understanding. Nonetheless, these findings highlight the importance of seasonality and temperature in shaping cycle hire patterns and can inform urban planning and sustainable transportation initiatives in London.

6. Conclusion:

This study investigated two hypotheses on cycle hire usage in London during 2014. Findings indicate a significant difference in average ride durations between Autumn and Spring, with longer rides in Autumn. A strong positive correlation was found between average temperature and cycle hire usage, suggesting temperature influences usage. These findings emphasize the importance of considering seasonality and temperature in cycle hire planning. Further research using longitudinal data and considering additional factors is needed. Overall, this study enhances our understanding of cycle hire patterns and informs urban transportation strategies.

7. References:

1. Emmanuel F. Werr (2022), London Weather Data [Online] [Accessed on 10th April 2023] https://www.kaggle.com/datasets/emmanuelfwerr/london-weather-data
2. DaysTo (2022) [Accessed on 10th April 2023] https://days.to/