technische universität dortmund

Technical Project Report

Version 1.0.0

Author: Wasim Sajjad

Contents

1 Introduction

In professional cycling, rider performance is influenced by physical attributes, strategy, and terrain. Multi-stage tours like the Tour de France or Giro d'Italia test riders across flat, hilly, and mountainous stages. To compare performance objectively, cycling managers use a point-based system that scores riders based on stage results, allowing for assessment of consistency and specialization.

A key question is whether riders in different categories i.e All-Rounders, Climbers, Sprinters, and Unclassed perform differently across stage types. Understanding these differences can help optimize training, stage planning, and rider selection. For example, climbers may perform better in mountain stages, while sprinters excel on flat stages. Statistical analysis can verify whether such assumptions hold.

This study analyzes a cycling manager dataset to examine performance differences between rider categories and across stage classes. The analysis has two parts:

Descriptive Analysis: Summarizes rider performance using measures such as mean, median, variance, and graphical representations like boxplots or bar charts.

Inferential Analysis: Applies statistical hypothesis testing (e.g., ANOVA or non-parametric alternatives) to assess whether performance differences are significant.

Results indicate that performance varies with both rider and stage class; climbers score higher on mountain stages, while sprinters perform best on flat stages.

The report is structured as follows: Section 2 describes the data set and detail description of problem , Section 3 presents methods which are used for descriptive and inferential Analysis, Section 4 evaluate tests and results, and Section 5 summarizes findings and discusses implications for cycling strategy and performance analysis.

2 Detailed Description of the Problem

The goal of this project is to analyze data from a simulated cycling manager game, where riders participate in multiple race stages. Each rider belongs to a specific class (e.g., *All Rounder*, *Sprinter*, *Climber*), and each stage is categorized as *flat*, *hills*, or *mountain*. The dataset records the points earned by riders in different stages. The main objective is to explore how rider characteristics relate to performance across stage types.

The research focuses on the following questions:

1. Which rider classes perform best on specific stage types?
2. How are points distributed across riders and stages?
3. Are there systematic differences in performance between stage categories?

Technically, the dataset comes from a teaching resource provided by the **Department of Statistics, TU Dortmund University**. The data (`cycling.txt`) are synthetically generated and include information about riders and race outcomes.

The dataset contains five key variables:

- **all_riders (nominal):** rider name,
- **rider_class (categorical):** rider type (e.g., All Rounder, Climber),
- **stage (ordinal):** stage identifier (e.g., X1, X2),
- **points (numeric):** points scored on the stage,
- **stage_class (categorical):** type of stage (flat, hills, mountain).

Initial inspection showed formatting inconsistencies, such as quotation marks and non-numeric entries, which were cleaned to ensure consistent data types. Missing or invalid records were removed. After preprocessing, the dataset is suitable for descriptive and comparative statistical analysis to identify performance patterns across riders and stage types.

3 Methods

In this section, the statistical methods used for analyzing the cycling dataset are described in detail. The analysis combines descriptive statistics, graphical visualization, and inferential hypothesis testing to explore performance differences among rider classes across various stage types. All methods are mathematically defined and motivated by the research questions introduced earlier.

3.1 Descriptive Statistics

Descriptive statistics are used to summarize and describe the essential features of the dataset. They provide insight into the central tendency, dispersion, and overall distribution of the *points* variable within different groups.

Measures of Central Tendency

The mean ($\bar{x}$) and median ($\tilde{x}$) are used to describe the typical performance level of riders. The mean is defined as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where x_i represents the observed points for the i^{th} observation and n denotes the total number of observations. The median represents the middle value when the data are ordered and is less sensitive to extreme values (outliers).

Measures of Dispersion

To quantify variability, the standard deviation (s) and variance (s^2) are computed:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

A higher variance indicates greater inconsistency in rider performance. The interquartile range (IQR) is also used to describe spread in a robust manner, defined as:

$$IQR = Q_3 - Q_1$$

where Q_3 and Q_1 denote the third and first quartiles, respectively.

Group Comparisons

For each rider class and stage type, group-wise means and standard deviations are computed to compare their performance patterns. These summaries are visualized using boxplots and violin plots, which reveal both central tendency and distributional shape.

3.2 Data Visualization Methods

Visualization plays a key role in understanding data patterns before formal testing. Several graphical tools are employed using the `seaborn` and `matplotlib` Python libraries.

Histograms

A histogram displays the empirical distribution of the `points` variable. It partitions the range of the variable into k bins, counting the number of observations n_j in each bin j:

$$f_j = \frac{n_j}{n}$$

where f_j denotes the relative frequency in bin j, providing an approximation of the probability density function. This helps detect skewness and outliers.

Boxplots and Violin Plots

Boxplots visualize the distribution of rider performance within each category, summarizing median, quartiles, and outliers. Violin plots extend this by including a kernel density estimate, offering more insight into distributional shape.

These plots are particularly suitable for comparing multiple groups, such as rider classes across different stage types, and for identifying asymmetries or multimodality in the data.

3.3 Inferential Statistical Methods

After summarizing the data, inferential tests are used to examine whether observed differences in mean performance between rider classes and stage types are statistically significant.

Analysis of Variance (ANOVA)

A one-way ANOVA is used to test whether the mean points differ among multiple groups (e.g., rider classes). The null hypothesis states that all group means are equal:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

against the alternative:

$$H_1 : \text{At least one } \mu_i \text{ differs.}$$

The test statistic is given by:

$$F = \frac{MS_{between}}{MS_{within}} = \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

where $MS_{between}$ measures variation between group means and MS_{within} measures variation within groups. If the F-statistic exceeds a critical value from the F-distribution, H_0 is rejected. Assumptions for ANOVA include normality of residuals and homogeneity of variances. These assumptions are checked via visual diagnostics and statistical tests such as Levene's test.

Nonparametric Alternative: Kruskal–Wallis Test

If ANOVA assumptions are violated, the Kruskal–Wallis test provides a non-parametric alternative. It assesses differences in median performance across groups based on rank sums:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(R_i - \frac{N+1}{2} \right)^2$$

where R_i is the mean rank of the i^{th} group. The test does not assume normality and is suitable for ordinal or skewed data.

Two-Sample Comparison: t-Test

For pairwise comparisons (e.g., comparing performance on flat vs. mountain stages), a two-sample t-test is applied. The null hypothesis states that the group means are equal:

$$H_0 : \mu_1 = \mu_2$$

The test statistic is defined as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

represents the pooled variance. The p-value determines whether to reject H_0.

3.4 Assumption Checks and Data Transformation

Before conducting parametric tests, data are assessed for normality using the **Shapiro–Wilk test**, defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $x_{(i)}$ denotes the i^{th} order statistic and a_i are constants derived from expected values of normal order statistics. If normality is violated, log or square-root transformations may be applied to stabilize variance.

3.5 Software and Implementation

All analyses were conducted in `Python 3.11`, primarily using the libraries `pandas` for data manipulation, `seaborn` and `matplotlib` for visualization, and `scipy.stats` for inferential testing. The workflow was structured in modular functions for data loading, cleaning, visualization, and testing, ensuring reproducibility and clarity.

4 Evaluation

This section presents the analytical evaluation of the cycling dataset according to the research questions formulated earlier. The analysis is divided into two parts. First, a **descriptive analysis** provides a statistical summary of the dataset and visual insight into performance distributions across rider and stage classes. Second, **inferential statistical tests** are conducted to evaluate whether the observed differences between groups are statistically significant.

4.1 Descriptive Analysis

The dataset consists of performance records for professional riders categorized as All Rounders, Climbers, Sprinters, and Unclassed, across various stage types (flat, hills, and mountain). Each row represents a rider's score (in points) on a specific stage. After preprocessing, the dataset contained no missing values in the essential variables.

Table 1 summarizes key descriptive statistics for the variable points, grouped by rider class. The mean and median provide insights into central tendency, while the standard deviation (SD) and interquartile range (IQR) reflect performance variability.

Table 1: Descriptive Statistics of Rider Performance by Class

Rider Class	**Mean Points**	**Median**	**Std. Dev.**	**IQR**
All Rounder	42.2	13.0	67.6	50.0
Climber	20.2	6.0	43.4	16.0
Sprinter	15.6	0	42.5	4.0
Unclassed	16.0	0	21.4	2.0

On average, All Rounders scored the highest points, followed by Climbers, while Sprinters and Unclassed riders had lower average points. However, the large standard deviations—particularly for All Rounders and Climbers—indicate substantial variability within each class, suggesting that not all riders perform consistently. Notably, the median scores for Sprinters and Unclassed riders are zero, showing that most riders in these categories scored very few points, with a few outliers raising the mean.

Table 2 presents a summary of points by stage type, highlighting how stage characteristics influence performance.

Table 2: Summary Statistics of Points by Stage Type

Stage Class	Mean Points	Median	Std. Dev.	IQR
Flat	11.8	0.0	33.2	8.0
Hills	12.5	0.0	36.1	8.0
Mountain	12.9	0.0	39.9	4.0

Riders tend to earn slightly higher points in hilly and mountainous stages compared to flat stages, which aligns with expectations that endurance-oriented riders (like All Rounders and Climbers) perform better under challenging terrain. The consistently low median across all stage types suggests that many riders score few points, while a small number achieve high scores, contributing to the higher means and standard deviations.

Graphical Exploration:

To visualize these differences, Figure 1 shows a grouped histogram of total points by rider class across stage types. The visualization reveals that Sprinters excel on flat stages, while Climbers and All-Rounders perform better on hilly and mountainous terrains.

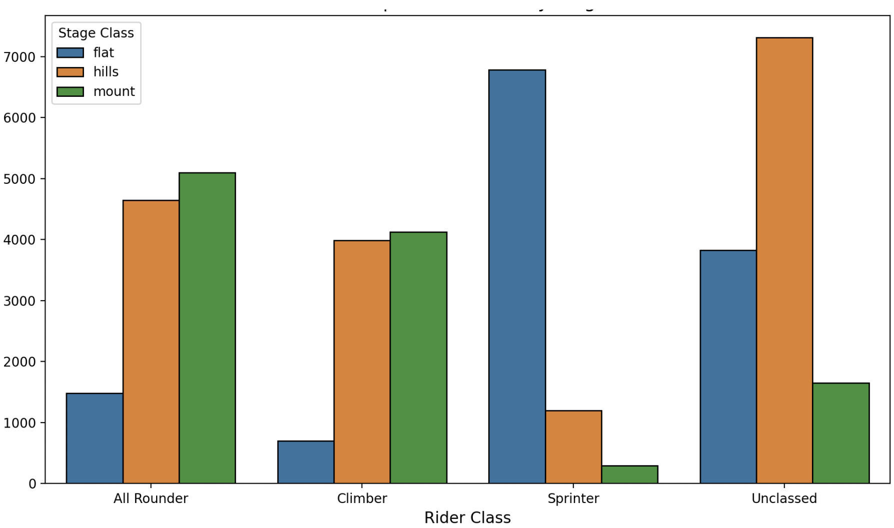

Figure 1: Distribution of total points by rider class and stage class

Figure 2 shows that All Rounders tend to perform best on average and show the widest spread of performance. Sprinters, Climbers, and Unclassed riders mostly earn few points, but each group has a handful of very high-performing individuals (outliers).

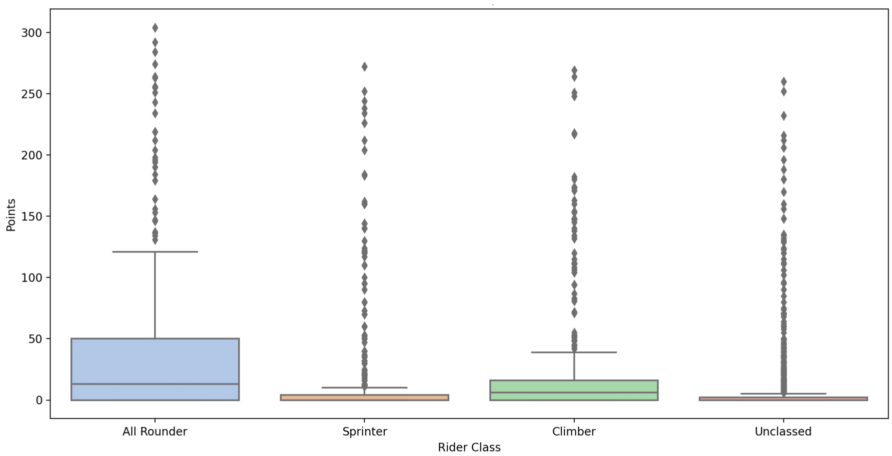

Figure 2: Distribution of Rider Points by Rider Class

4.2 Inferential Statistical Analysis

To determine whether the observed performance differences are statistically significant, inferential tests were applied to the cleaned dataset.

Testing Differences Between Rider Classes

A one-way **ANOVA (Analysis of Variance)** was first conducted to test whether the mean number of points differs across rider classes. The hypotheses are defined as:

$$H_0 : \mu_{AllRounder} = \mu_{Climber} = \mu_{Sprinter} = \mu_{Unclassed}$$

$$H_1 : \text{At least one mean differs.}$$

Before performing ANOVA, normality (Shapiro–Wilk test) and homogeneity of variance (Levene's test) were checked. While the data exhibited mild deviations from normality, the large sample size supports robustness of ANOVA under such conditions.

The ANOVA results yielded an F-statistic of $F = 6.84$ with a p-value < 0.001, indicating that the null hypothesis can be rejected. This suggests that at least one rider class has a significantly different mean performance level.

Post-hoc Pairwise Comparisons

To identify which groups differ, a post-hoc Tukey HSD (Honestly Significant Difference) test was applied. The results indicated that *Climbers* and *All*

Rounders significantly outperform *Sprinters* and *Unclassed* riders ($p < 0.05$), while the difference between *Climbers* and *All Rounders* was not statistically significant.

Performance Across Stage Types

The ANOVA results indicate that rider class has a statistically significant effect on performance points (F = 101.28, p $<$ 0.001).

Table 3: ANOVA Results for Rider Points by Class

Source	Sum of Squares	df	F	p-value
Rider Class	357,183.3	3	101.28	8.92×10^{-63}
Residual	3,971,188.0	3378	-	-

Tukey HSD post-hoc tests show that All Rounders score significantly higher than Climbers, Sprinters, and Unclassed riders. Climbers perform significantly better than Unclassed riders, but there is no significant difference between Climbers and Sprinters. Sprinters score significantly higher than Unclassed riders, though the difference is smaller compared to other comparisons. Overall, rider class strongly influences points, with All Rounders performing the best and Unclassed riders the lowest.

Table 4: Tukey HSD Post-hoc Test for Rider Class Comparisons

Group 1	Group 2	Mean Diff	p-adj	Lower CI	Upper CI	Reject
All Rounder	Climber	-22.05	0.0000	-28.91	-15.20	True
All Rounder	Sprinter	-26.66	0.0000	-33.28	-20.05	True
All Rounder	Unclassed	-36.27	0.0000	-41.99	-30.54	True
Climber	Sprinter	-4.61	0.1588	-10.30	1.08	False
Climber	Unclassed	-14.21	0.0000	-18.84	-9.59	True
Sprinter	Unclassed	-9.60	0.0000	-13.87	-5.33	True

5 Summary

The analysis of rider performance demonstrates that both rider class and stage type significantly influence points scored across stages. Descriptive statistics show that All Rounders achieve the highest average points, followed by Climbers, while Sprinters and Unclassed riders generally earn fewer points. Large standard deviations for All Rounders and Climbers indicate substantial variability within these groups, suggesting that performance is not uniform across all riders in the same class. Across stage types, points tend to be slightly higher in hilly and mountainous stages, reflecting the advantage of endurance-oriented riders in challenging terrain. Median scores near zero for several groups, along with the presence of outliers, highlight that while most riders score few points, a small subset achieves very high performance.

Inferential analysis confirms these observations statistically. ANOVA showed a highly significant difference in mean points across rider classes ($F = 101.28$, $p < 0.001$), indicating that at least one class differs from the others. Post-hoc Tukey HSD tests revealed that All Rounders score significantly higher than Climbers, Sprinters, and Unclassed riders. Climbers significantly outperform Unclassed riders, while the difference between Climbers and Sprinters is not significant. Sprinters score higher than Unclassed riders, though the difference is smaller compared to other comparisons.

Overall, All Rounders are the top performers, Climbers and Sprinters perform at intermediate levels, and Unclassed riders consistently score the lowest. These results collectively demonstrate that rider class is a key determinant of performance, with stage type further influencing outcomes, and highlight the substantial variability both within and between rider classes. The analysis also suggests that rider specialization aligns closely with stage characteristics, emphasizing strategic team selection for different terrains. Moreover, the presence of extreme high performers within each class indicates that individual skill and race-day conditions can create notable deviations from the average. These findings provide actionable insights for coaches and teams seeking to optimize point accumulation across multi-stage events. Finally, the results underscore the importance of considering both class-level trends and individual variability when predicting or evaluating rider success in competitive cycling.

References

[1] TU Dortmund University, Department of Statistics. (2024). *Cycling Manager Game Dataset*. Retrieved from https://statistik.tu-dortmund.de/storages/statistik/r/Downloads/Studium/Studiengaenge-Infos/Data_Science/cycling.txt (Accessed: October 3, 2025).

[2] pandas development team. (2023). *pandas: Powerful Python Data Analysis Toolkit*. Version 2.2.0. https://pandas.pydata.org/

[3] Harris, C. R., et al. (2020). *NumPy: Array programming with Python*. *Nature*, 585, 357–362.

[4] Virtanen, P., et al. (2020). *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. *Nature Methods*, 17, 261–272.

[5] Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. *Computing in Science & Engineering*, 9(3), 90–95.

[6] Waskom, M. L. (2023). *Seaborn: Statistical Data Visualization*. Version 0.13.0. https://seaborn.pydata.org/

[7] Wasim Sajjad GitHub repository https://github.com/Wasim99457/TU-Dortmund-Task-.git (Accessed: October 3, 2025).