Application Report: Analysis of Rider Performance in a Cycling Manager Game

Application for M.Sc. Data Science — Summer Semester 2026

Milad Afkhamipour Sharabyani

February 19, 2026

Contents

1   Introduction

This report examines a dataset from a cycling manager game in order to address two research questions:

1. **RQ1:** Do rider classes differ significantly in overall performance as measured by points scored?
2. **RQ2:** Does the performance difference between rider classes depend on the type of stage (flat, hills, or mountain)?

The analysis finds statistically significant performance differences across all four rider classes. More importantly, these differences are strongly shaped by terrain. Sprinters perform well on flat stages but earn almost no points on mountainous terrain, while All Rounders and Climbers reach their highest scores on mountain stages. Unclassed riders score low regardless of the stage type.

The report is organized as follows. Section 2 describes the dataset. Section 3 presents the statistical methods used, along with their mathematical definitions. Section 4 reports the results of both descriptive and inferential analyses. Section 5 discusses the main findings, their implications, the limitations of the study, and possible directions for future work.

2   Data Description

The dataset was provided by the Faculty of Statistics at TU Dortmund University. It contains 3,496 observations and 5 variables. Each row records the performance of a single rider on a single stage of the cycling manager game.

Table 1: Overview of variables in the cycling dataset.

Variable	Description	Scale Level	Values
all_riders	Rider name	Nominal	184 unique riders
rider_class	Rider category	Nominal	All Rounder, Climber, Sprinter, Unclassed
stage	Stage identifier	Ordinal	X1–X21 (X5, X13 absent)
points	Performance score	Ratio	0–304, integer
stage_class	Terrain type	Nominal	flat, hills, mount

A few features of the dataset are worth noting. There are no missing values. Out of 21 possible stages, 19 appear in the data; stages X5 and X13 are absent, likely because they correspond

to rest days. The group sizes are quite unbalanced: the Unclassed category includes 115 riders, while All Rounder has only 17. There is also substantial zero-inflation, with 61.8% of all point values being exactly zero. Points cannot be negative, and the distribution is strongly right-skewed in every rider class.

All analyses were performed using R version 4.3.3 (R Core Team, 2024) with the packages ggplot2 (Wickham, 2016), dunn.test (Dinno, 2024), kableExtra (Zhu, 2024), and car (Fox & Weisberg, 2019). The complete source code and data are available at https://github.com/milad-afkhami/cycling-performance-analysis.

3   Statistical Methods

This section describes all statistical methods applied in the analysis. Due to the heavy right skew, zero-inflation, and violations of normality and variance homogeneity (shown in Section 4), nonparametric methods are used throughout (Hollander et al., 2013).

3.1   Descriptive Statistics and Graphical Methods

The **arithmetic mean** $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ provides a measure of central tendency, while the **median** $\tilde{x}$ offers a robust alternative for skewed distributions (Fahrmeir et al., 2016). Dispersion is measured by the **standard deviation** $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$ and the **interquartile range** $\mathrm{IQR} = Q_3 - Q_1$. Given the heavy right skew and zero-inflation in this dataset, the median and IQR are preferred over the mean and standard deviation.

The **boxplot** (Tukey, 1977) provides a visual summary: the box spans Q_1 to Q_3 with a line at the median; whiskers extend to the most extreme point within $1.5 \times \mathrm{IQR}$; observations beyond are plotted as outliers. Grouped boxplots faceted by stage class allow simultaneous assessment of both research questions.

3.2   Assumption Checking

The **Shapiro-Wilk test** (Shapiro & Wilk, 1965) evaluates H_0: the data follow a normal distribution, via a test statistic W based on the correlation between observed order statistics and expected normal values. **Levene's test** (Levene, 1960) evaluates H_0: population vari-

ances are equal across groups, computing an F-statistic on absolute deviations from group medians. When both assumptions are violated, nonparametric alternatives are warranted.

3.3 Kruskal-Wallis Test

The **Kruskal-Wallis rank sum test** (Kruskal & Wallis, 1952) is a nonparametric alternative to one-way ANOVA for comparing k independent groups. All N observations are pooled and ranked. The test statistic is

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

where $N = \sum_{j=1}^{k} n_j$ is the total number of observations, n_j is the size of group j, and $R_j = \sum_{i=1}^{n_j} r_{ij}$ is the sum of ranks assigned to observations in group j. Under the null hypothesis H_0 that all k group distributions are identical, H follows approximately a chi-squared distribution with $k-1$ degrees of freedom. The alternative hypothesis H_1 states that at least two groups differ in their distributions.

3.4 Post-Hoc Testing and Multiple Comparisons

When the Kruskal-Wallis test rejects H_0, **Dunn's test** (Dunn, 1964) identifies which groups differ by performing all $m = \binom{k}{2}$ pairwise rank-sum comparisons. To control the family-wise error rate, the **Bonferroni correction** (Bonferroni, 1936) adjusts the significance level to $\alpha_{\text{adj}} = \alpha/m$. For $k = 4$ groups, $m = 6$ pairwise comparisons are conducted.

3.5 Effect Size

The **eta-squared** (η_H^2) effect size based on the Kruskal-Wallis H statistic quantifies the magnitude of group differences (Tomczak & Tomczak, 2014):

$$\eta_H^2 = \frac{H - k + 1}{N - k}$$

Values of approximately 0.01, 0.06, and 0.14 indicate small, medium, and large effects, respectively.

3.6   Analysis Strategy

All tests use a significance level of $\alpha = 0.05$. For RQ1, a single Kruskal-Wallis test with Dunn's post-hoc comparisons is applied to the full dataset. For RQ2, a **stratified analysis** is conducted: separate Kruskal-Wallis tests within each stage class (flat, hills, mount), with Bonferroni correction across the three strata ($\alpha_{\text{adj}} = 0.05/3 = 0.0167$). Significant strata are followed up with Dunn's post-hoc tests.

4   Results

4.1   Descriptive Analysis

4.1.1   Summary Statistics by Rider Class

Table 2: Summary statistics of performance points by rider class.

Rider Class	n	Riders	Mean	Median	SD	IQR	Min	Max
All Rounder	323	17	37.69	12	63.96	39.5	0	304
Climber	437	23	20.17	6	43.45	16.0	0	269
Sprinter	551	29	15.04	0	41.83	4.0	0	272
Unclassed	2185	115	6.42	0	23.28	2.0	0	260

Table 1 shows a clear ordering among rider classes. All Rounders have the highest average performance (mean = 37.69), followed by Climbers, then Sprinters, and finally Unclassed riders. In every group, there is a large gap between the mean and the median, which reflects the heavy right skew mentioned in Section 2. Both Sprinters and Unclassed riders have a median of zero, meaning that over half of their stage performances result in no points at all.

4.1.2   Mean Performance by Rider Class and Stage Class

Table 3: Mean performance points by rider class and stage class.

Rider Class	Flat	Hills	Mountain
All Rounder	15.44	35.79	67.42
Climber	5.09	21.67	35.86
Sprinter	38.98	5.20	2.04
Unclassed	5.74	9.10	2.95

Table 2 shows a strong interaction between rider class and stage type. Sprinters achieve their best average score on flat stages (38.98) but drop to near-zero on mountain stages (2.04). All Rounders, on the other hand, score highest on mountain stages (67.42) and considerably less on flat terrain (15.44). Climbers follow a similar pattern favoring mountains, though the difference is less extreme. Unclassed riders remain low across all three terrain types.
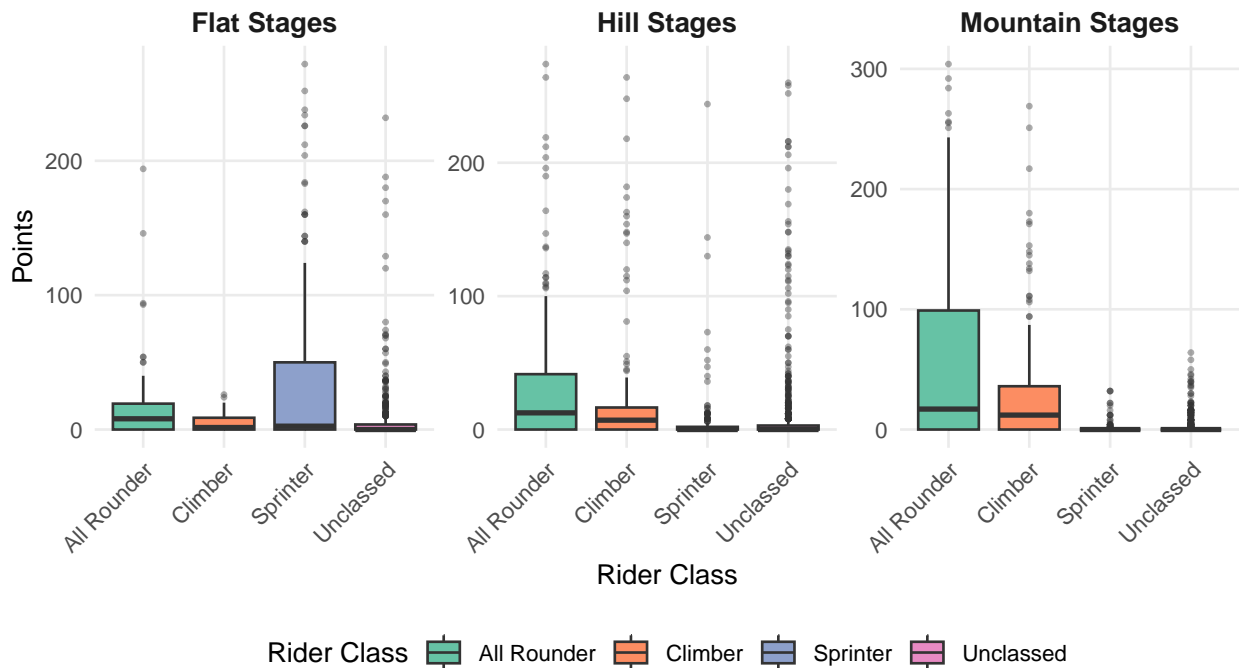
4.1.3 Graphical Analysis

Figure 1: Distribution of performance points by rider class, stratified by stage class. Boxes span the interquartile range (Q1–Q3); horizontal lines indicate the median; whiskers extend to 1.5×IQR; points beyond are outliers.

Figure 1 confirms these patterns visually. The distributions are right-skewed in all groups, with many high-scoring outliers. On flat stages, Sprinters show the widest spread and the highest median. On mountain stages, All Rounders clearly stand out. The Unclassed category shows very little variation across terrain types, which is consistent with a support-rider role where consistent but low scoring is expected.

4.2   Inferential Analysis

4.2.1   Assumption Checks

Table 4: Shapiro-Wilk normality test results by rider class.

Rider Class	W	p-value
All Rounder	0.6410	<2e-16
Climber	0.4997	<2e-16
Sprinter	0.4137	<2e-16
Unclassed	0.2898	<2e-16

The Shapiro-Wilk test rejected normality for all four rider classes ($p < 0.001$ in each case; Table 3). Levene's test also rejected homogeneity of variance ($F(3, 3492) = 83.08, p < 0.001$). Given these violations, nonparametric methods are warranted for all subsequent inferential analyses.

4.2.2   RQ1: Overall Differences Between Rider Classes

A Kruskal-Wallis test showed a statistically significant difference in performance points across the four rider classes ($H(3) = 330.2, p < 0.001$; Kruskal & Wallis (1952)). The corresponding effect size was $\eta_H^2 = 0.0937$, which indicates a medium effect (Tomczak & Tomczak, 2014).

4.2.3   Post-Hoc Pairwise Comparisons (Overall)

Table 5: Dunn's post-hoc pairwise comparisons with Bonferroni correction (overall).

Comparison	Z	Adjusted p-value
All Rounder - Climber	2.545	0.0328
All Rounder - Sprinter	10.383	<2e-16
Climber - Sprinter	8.444	<2e-16
All Rounder - Unclassed	14.479	<2e-16
Climber - Unclassed	12.908	<2e-16
Sprinter - Unclassed	2.843	0.0134

The results of Dunn's post-hoc tests with Bonferroni correction are shown in Table 4 (Bonferroni, 1936; Dunn, 1964). All six pairwise comparisons reached statistical significance at the corrected level. This means that every rider class differs from every other class in its performance distribution.

4.2.4 RQ2: Stratified Analysis by Stage Class

Table 6: Kruskal-Wallis tests within each stage class (Bonferroni-corrected $\alpha = 0.0167$).

	Stage Class	H	df	p-value	η_H^2	Significant
Kruskal-Wallis chi-squared...1	flat	82.10	3	<2e-16	0.0719	Yes
Kruskal-Wallis chi-squared...2	hills	156.40	3	<2e-16	0.1045	Yes
Kruskal-Wallis chi-squared...3	mount	183.16	3	<2e-16	0.1967	Yes

As shown in Table 5, the Kruskal-Wallis test is significant within each stage class at the Bonferroni-corrected level of $\alpha = 0.0167$. This confirms that rider class differences are present across all terrain types, not just in the overall data. The effect sizes do vary by terrain, with the largest effects appearing on mountain and flat stages where the specialization is most relevant.

4.2.5 Post-Hoc Comparisons Within Stage Classes

Dunn's post-hoc tests with Bonferroni correction were also carried out within each stage class separately. On **flat stages**, 3 of the 6 pairwise comparisons were significant. All of these involved the Unclassed group, which differed from All Rounders, Climbers, and Sprinters. The three classified rider types, however, did not differ significantly from one another on flat terrain. On **hill stages**, 4 of 6 comparisons were significant, with All Rounders and Climbers forming one performance group and Sprinters and Unclassed riders forming another. On **mountain stages**, the pattern was very similar: 4 of 6 comparisons were significant, and the same two-cluster structure appeared. These results show that the relative performance ranking among rider classes changes considerably depending on terrain.

5 Summary of Findings

This report analyzed rider performance in a cycling manager game dataset containing 3,496 observations across 184 riders and 19 stages.

RQ1: The Kruskal-Wallis test found significant differences across all four rider classes ($p < 0.001$), and all six Dunn's post-hoc comparisons remained significant after Bonferroni correction. Average performance follows the order All Rounder > Climber > Sprinter > Unclassed, with a medium effect ($\eta_H^2 = 0.0937$).

RQ2: The stratified analysis confirmed significant differences within all three stage classes. The key finding is that the pattern of differences depends on terrain. Sprinters perform best on flat stages (mean 38.98) but earn almost no points on mountains (mean 2.04), while All Rounders reach their peak on mountain stages (mean 67.42). This terrain-dependent shift in performance is the most notable result of the analysis.

Overall, the rider classifications in this game appear to reflect meaningful performance specializations that align with real-world cycling. Sprinters excel on flat terrain, All Rounders maintain solid performance across stage types with particular strength in the mountains, and Unclassed riders score consistently low — a pattern consistent with a domestique or support-rider role. From a practical standpoint, these findings suggest that team managers should select riders based on the expected terrain profile of a given race.

References

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche e Commerciali Di Firenze*, *8*, 3–62.

Dinno, A. (2024). *Dunn.test: Dunn's test of multiple comparisons using rank sums.* https://CRAN.R-project.org/package=dunn.test

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252.

Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & Tutz, G. (2016). *Statistik: Der weg zur datenanalyse* (8th ed.). Springer.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://www.john-fox.ca/Companion/

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods* (3rd ed.). John Wiley & Sons.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to probability and statistics* (pp. 278–292). Stanford University Press.

R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited.

Trends in Sport Sciences, *21*(1), 19–25.

Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org

Zhu, H. (2024). *kableExtra: Construct complex table with 'kable' and pipe syntax.* https://CRAN.R-project.org/package=kableExtra