

The UCI's points system and team ranking regulations have significantly influenced team strategies, race selection, and financial planning in professional road cycling. The introduction

of the UCI ProTour established a structured league of top-tier teams but led to unintended behaviors, such as regional race concentration, due to the licensing framework [2]. The allocation of points across different race classifications plays a crucial role in shaping team participation strategies, affecting both individual rider development and overall team performance, as teams adapt to maximize point accumulation [3]. Economic factors also heavily influence team competitiveness, with wealthier teams benefiting from greater investment in elite riders and advanced resources, while lower-budget teams must strategically manage race selection and rider development to remain competitive [4].

### III. DATA COLLECTION

The dataset was scraped from publicly available sources, primarily the *ProCyclingStats* (PCS) website [5], which provides comprehensive race results, rider statistics, and team performance data for professional cycling. Additionally, *team budget estimates* were collected from *ProCyclingUK* [6]. The data collection process required careful planning to ensure completeness while adhering to ethical and legal guidelines. The scraping process was conducted using a combination of *Requests*, *BeautifulSoup*, and *Selenium* to handle *HTTP requests*, *parse structured data*, and *navigate dynamic content*. Regular expressions (Regex) were employed to extract key details such as race results, rider IDs, team names, and ranking points. The process took approximately four weeks, resulting in a structured dataset containing 386,836 detailed race results, including rider profiles, team statistics, and financial estimations, forming the basis for subsequent analyses.

### IV. DATASET DESCRIPTION

This study uses a dataset compiled from *ProCyclingStats* (PCS) and *ProCyclingUK*, covering race results, UCI points allocation, rider rankings, and team budgets. The dataset integrates multiple sources to analyze how UCI regulatory changes influenced team strategies.

#### A. Race and Points Data

The primary dataset consists of race results from all UCI-sanctioned road races between 2020 and 2023. Each race entry includes rider performance metrics such as finishing position, race duration, speed, and team affiliation, alongside race characteristics such as distance, elevation gain, terrain type, and weather conditions. PCS classifies race terrains into multiple categories, distinguishing between sprint-oriented, hilly, and mountainous races, with additional categories for time trials (TT) and general classification (GC). Alongside race data, the UCI points scale provides the number of points awarded per finishing position across different race classifications. To analyze the impact of the 2023 points system modification, I computed the percentage change in total points available per race classification from year to year. Points increase trends were categorized into three levels: low ( $\leq 50\%$ ), medium (50%–200%), and high ( $> 200\%$ ). Teams were also classified by budget: low ( $< 20\text{M}$  Euro), medium (20M–50M Euro), and

high ( $> 50\text{M}$  Euro), allowing for a comparative analysis of how financial resources influenced participation and roster decisions.

#### B. Year-to-Year Rider Performance and Ranking

To evaluate long-term performance trends, I aggregated rider-level data into a yearly ranking system based on total UCI points. An additional metric, assist points, was introduced to measure a rider's contribution to their team's success. Assist points for a rider in a given race were defined as the sum of UCI points earned by all other teammates, excluding their own. This approach enables an evaluation of leadership versus support roles within teams and helps quantify how teams distribute points accumulation across their rosters.

#### C. Dataset Integration and Merging

To create a unified structure, year-to-year rider performance data was merged with the following season's race results, allowing us to analyze the influence of past performance on future participation. The UCI Points Scale was linked to each race to incorporate shifts in point allocations over time. Finally, Points Increase Shift data was assigned to each race-year combination to examine whether teams adapted their strategies based on changing incentives. The final dataset captures rider-level performance trends, race characteristics, ranking shifts, and budget classifications, providing a robust foundation for analyzing team strategies in response to UCI regulatory changes. Table I summarizes the merged dataset's key features.

TABLE I  
FINAL DATASET OVERVIEW

Category	Feature Description
<b>Race Conditions</b>	Race name, date, distance, elevation, terrain type, location
<b>Race Points</b>	Points scale, points available for race class, Yearly percentage change in total points
<b>Rider Information</b>	Name, age, team affiliation, speciality
<b>Rider-Race Performance</b>	Rank, time gap, points gained, assists, GC standing
<b>Rider Historical Performance</b>	Previous season's total points, UCI/assist rank, terrain-specific rank
<b>Team Budget</b>	Estimated budget, budget category (high, medium, low)

### V. METHODOLOGY

To evaluate team performance and race strategy, I aggregate rider-level data into a team-level dataset, capturing detailed team strength metrics. I compute a roster strength metric to quantify the competitive power of a team's lineup, assigning greater influence to higher-ranked riders while progressively reducing the impact of lower-ranked teammates. The following subsections detail the methodology used to define and calculate roster strength, outlining the specific metrics and aggregation techniques applied in this study.

### A. Team Roster Strength Computation

Each team's roster strength in a given race is computed by assigning greater influence to higher-ranked riders while progressively reducing the impact of lower-ranked teammates. The roster strength for a team  $T$  in a race  $r$  is given by:

$$S_T^r = \frac{1}{N} \sum_{i=1}^N S_i \quad (1)$$

where the individual rider score  $S_i$  is defined as:

$$S_i = \alpha \cdot \left( \frac{N - R_i + 1}{N} \right) + (1 - \alpha) \cdot \left( \frac{B - B_i + 1}{B} \right) \quad (2)$$

The term  $N$  represents the total number of riders in the team. Each rider is assigned an internal ranking  $R_i$  within the team, where the best-ranked rider has  $R_i = 1$  and the worst-ranked rider has  $R_i = N$ . This ranking is normalized so that the top-ranked rider receives the highest value, while lower-ranked teammates have progressively smaller contributions. The bin category ranking  $B_i$  is used to apply a diminishing adjustment to a rider's impact based on their percentile ranking within the team. The intention is to ensure that higher-ranked riders have a greater influence on the team's final strength score, as they contribute more significantly to overall performance. The parameter  $\alpha$  controls how much internal ranking versus bin class influences the final score. A higher  $\alpha$  places more emphasis on team hierarchy, making the internal ranking the dominant factor, while a lower  $\alpha$  allows for bin categorization to play a greater role. In this study, I set  $\alpha = 0.5$ , meaning that both internal ranking and bin class contribute equally to the final score. To capture a comprehensive representation of internal team ranking, I compute the roster strength metric across four key dimensions: leader rank across all races, leader rank specific to the race terrain, assist rank across all races, and assist rank by race terrain. These measurements provide a nuanced evaluation of team roster balance, incorporating both overall performance trends and terrain-specific strengths.

### B. One vs Multi-day races Aggregation

For one-day races, the roster strength metric  $S_T^r$  by the race terrain is computed directly for the race terrain  $p_r$  since there is only one, ensuring that the evaluation reflects the specific demands of the course:

$$S_T^{r,p_r} = S_T^r \quad (3)$$

Multi-stage races introduce additional complexity due to terrain variations across stages. To account for this, I compute stage-specific roster strengths and aggregate them using a weighted average based on the number of stages in each terrain type:

$$S_T^{\text{tour}} = \sum_p w_p \cdot S_T^p \quad (4)$$

where  $w_p$  represents the proportion of stages belonging to terrain  $p$  in the race. This ensures that terrain types occurring more frequently have a greater influence on the overall team strength evaluation. Once all race-level scores are computed,

they are aggregated at the season level to analyze trends in team strategies across different budget categories and ranking levels. Once all race-level scores are computed, they are aggregated at the season level to analyze trends in team strategies across different budget categories and ranking levels. After aggregating rider strength within teams, reducing multi-stage races to a single tour-level representation, and filtering out non-WorldTour and non-ProTeam teams, the final dataset consists of approximately 7,000 team-race records spanning the years 2020 to 2023.

### C. Team-Race Network Graph

To further analyze team strategies, I construct a Team-to-Race Network Graph for each season, where teams and races are represented as nodes, and participation in a race forms an edge between them. This network allows us to explore team engagement patterns and how roster strength relates to race selection. Team nodes are classified based on their budget category and include attributes such as team score and average budget. Race nodes are categorized by their points increase classification, which represents how much their UCI points allocation changed from the previous year. Additional race attributes include race classification and terrain type. The strength of a team's roster in a race is encoded in the edge weight, prioritizing leaders over support riders and emphasizing terrain-specific rankings over general rankings. The edge weight is calculated as:

$$W_{T,R} = \frac{2S_p^{\text{leader}} + S^{\text{leader}} + 2S_p^{\text{assist}} + S^{\text{assist}}}{6} \quad (5)$$

$S_p^{\text{leader}}$  and  $S_p^{\text{assist}}$  are the leader and assist scores specific to the race terrain.  $S^{\text{leader}}$  and  $S^{\text{assist}}$  are the leader and assist scores across all terrains. Edges also store additional attributes such as the team's best rider ranking in the race, the ratio of new riders in the lineup, and whether the team participated in the same race the previous season.

## VI. RESULTS ANALYSIS

In this section, I analyze the impact of the UCI points system modification by addressing the four key research questions introduced earlier.

### A. Race Participation: Shifts in Team Race Selection

To understand how teams have adjusted their race selection based on the revised UCI points allocation, I first analyze geographical patterns of participation across different budget categories. Figure 2 presents a world map illustrating regions where high-budget and low-budget teams were more active in 2023. A similar analysis is shown in Figure 3, focusing on Europe, where the highest percentage of points increases occurred. Each European region is labeled with its average points increase category: L (low), M (medium), or H (high). From the world map, it appears that both high-budget and low-budget teams have a comparable distribution of regions where they focused their participation. However, when narrowing the analysis to Europe, I observe that high-budget teams

remain more dominant, particularly in areas with medium-to-high points increases (M-H labeled). This suggests that high-budget teams prioritized races that received greater UCI point allocations, potentially maximizing their competitive advantage. Further insights into race selection strategies are provided in Figure 4, which shows the number of new and abandoned races per team budget category, compared to the percentage increase in points from 2022 to 2023. The trend lines reveal that low- and medium-budget teams dropped many races that did not receive a points increase, showing a shift in their priorities. While medium-budget teams were more careful about entering new races with low points increases, low-budget teams were more willing to try them. One possible reason for this is that with fewer medium-budget teams in these races, the competition became easier, giving low-budget teams a better chance to score points. Across all budget categories, teams increased their participation in races with high points increases, showing a common strategy of focusing on events that now offer more rewards under the new UCI system.

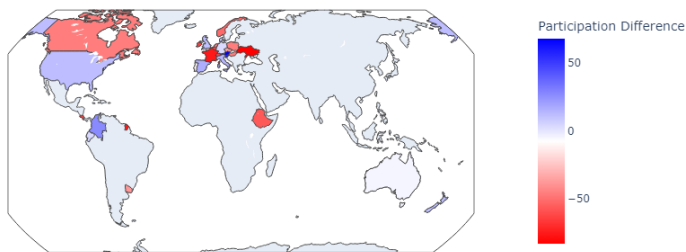


Fig. 2. Difference in average race participation between high-budget and low-budget teams across different regions (2023).

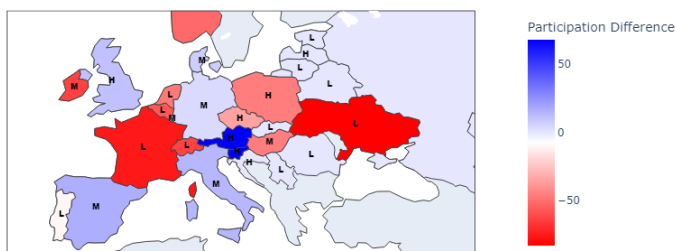


Fig. 3. Difference in race participation within Europe, labeled by race points increase category (L: low, M: medium, H: high).

### B. Roster Allocation: Strengthening Race Participation Strategies

To examine whether teams allocated stronger rosters to certain races following the 2023 UCI modification, I analyze changes in roster strength across different races.

Figure 5 presents the distribution of roster strength changes across races, plotted against the percentage increase in UCI points from 2022 to 2023. The results show that the number of races where teams increased their roster strength is nearly

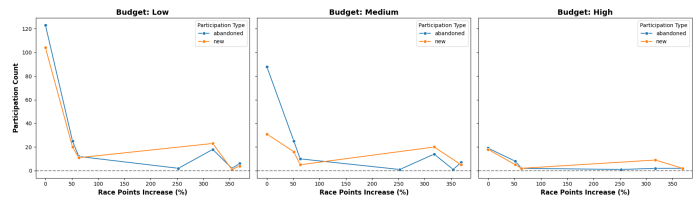


Fig. 4. New vs. abandoned race participation per budget category, plotted against race points increase percentage (2022-2023).

balanced with those where roster strength decreased. This suggests that, at an aggregate level, there is no clear trend of teams universally strengthening or weakening their lineups for particular race categories. Instead, this variability highlights that roster allocation strategies are highly individualized, with teams making trade-offs between different races. One reason for this observation is that roster optimization is influenced by team-specific strategies rather than budget constraints alone. While some teams may prioritize strengthening their lineup for specific race types, others within the same budget category might do the opposite, reallocating strength from one race to another. Furthermore, the roster strength metric used in this study is a simplified representation of team strength. Different teams may define roster power differently, adjusting the weighting of leader and assist rankings or emphasizing certain terrain-specific strengths. This flexibility in defining team strength further complicates direct budget-based clustering of roster strategies. To further illustrate the team-specific variability in roster allocation, Figure 6 visualizes the network of teams and races, where team nodes are categorized by their budget, race nodes by their UCI points increase category, and edges are colored based on the shift in roster strength from 2022 to 2023. This network highlights how different teams apply unique strategies in determining which races to prioritize. Some teams increase roster strength for certain high-value races, while others shift their focus elsewhere, possibly in response to their rivals' strategies.

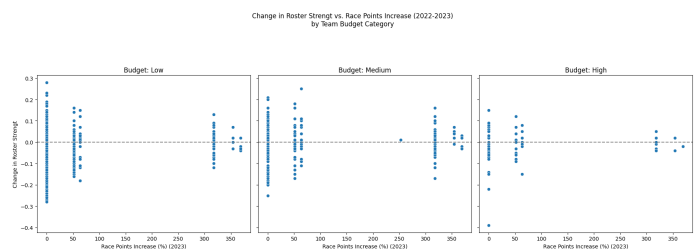


Fig. 5. Distribution of changes in roster strength across races, categorized by race points increase percentage (2022-2023).

### C. Rider Strategy: Star Riders vs. Distributed Effort

To examine whether teams have shifted from relying on a few top riders to distributing point accumulation across

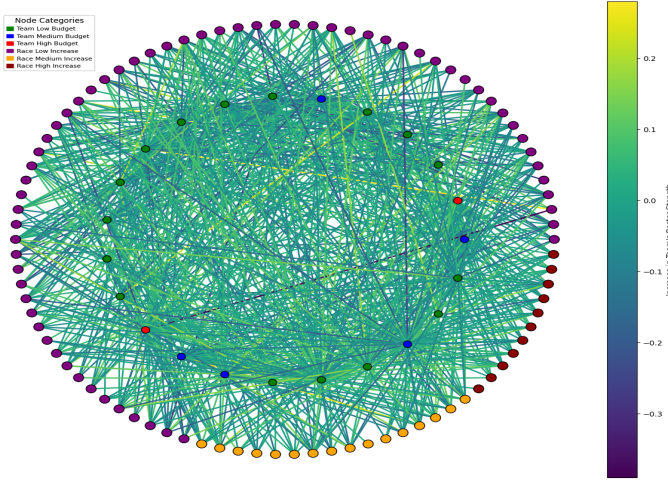


Fig. 6. Team-to-race network visualization. Teams are categorized by budget, races by points increase category, and edges are colored by roster strength shift (2022-2023).

multiple riders, I analyze participation trends by leader score and internal team ranking.

Figure 7 presents the distribution of rider participation by leader score across teams of different budget categories, comparing the years 2021 to 2023. The results indicate a clear shift among low- and medium-budget teams, where the distribution in 2023 becomes wider and shorter compared to previous years. This suggests that these teams have diversified their race participation strategies, allowing more riders to compete rather than depending solely on a few high-performing individuals to secure points. In contrast, the distribution for high-budget teams remains largely unchanged between 2022 and 2023, indicating that their reliance on top riders has remained stable. A similar trend is observed in Figure 8, which visualizes rider participation based on internal team ranking. The 2023 distribution for low- and medium-budget teams appears wider, with participation spread across a larger range of team members rather than concentrated among the highest-ranked riders observed in the high-budget teams. These findings highlight a strategic divergence between budget groups. While lower-budget teams appear to be distributing race opportunities more evenly across their roster, high-budget teams continue to focus on their established leaders.

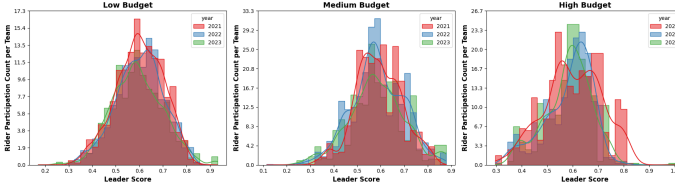


Fig. 7. Rider participation count by leader score, comparing distributions across 2021-2023 for different budget categories.

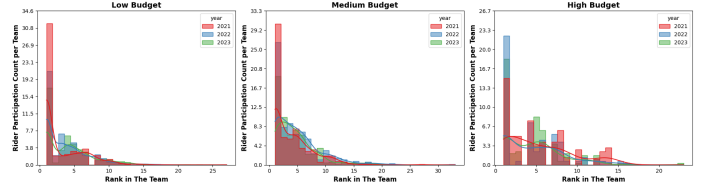


Fig. 8. Rider participation count by internal ranking within teams, comparing distributions across 2021-2023.

#### D. Team Rankings and Year-End Points Accumulation

To evaluate the impact of the new UCI points distribution on team rankings and budget-related competitiveness, I analyze changes in total points and UCI rankings across seasons. Figure 9 shows the yearly change in total points per budget category, while Figure 10 illustrates the corresponding ranking shifts. The effects of external disruptions, such as COVID-19, are evident in these trends. In 2020-2021, low-budget teams improved their rankings due to reduced participation from higher-budget teams, despite a decline in overall points gained. As normal race schedules resumed in 2022, high- and medium-budget teams regained dominance, accumulating significantly more points and stabilizing their rankings. By 2023, total points accumulation remained relatively stable, with rankings showing little movement across budget groups. However, within the lower-budget category, rankings became more volatile—some teams significantly improved, while others fell further behind. A key distinction emerged between low-budget WorldTour and ProTeams: WorldTour teams dropped by an average of  $-4.8$  positions, while ProTeams improved by  $+1.76$  ranks. This suggests that WorldTour teams, constrained by mandatory high-level race participation, struggled to secure points, whereas ProTeams benefited from greater race selection flexibility, focusing on lower-tier events where they could accumulate points more effectively. While the overall team hierarchy remained largely unchanged, these findings indicate that the UCI modification intensified competition within lower-budget teams, favoring those with the ability to optimize their race calendars.

### VII. CONCLUSION AND FUTURE DIRECTIONS

This study examined the impact of the 2023 UCI points system modification on team strategies, including race participation, roster allocation, reliance on star riders, and overall rankings. The findings suggest that while high-budget teams remained largely unaffected, lower-budget teams experienced greater variability. ProTeams benefited from race selection flexibility, improving their rankings, while low-budget WorldTour teams, constrained by mandatory high-level race participation, struggled to accumulate points effectively. The modifications reinforced the importance of strategic race selection and roster planning, particularly for teams with fewer financial resources. Future research could explore how individual rider characteristics influence performance under the new system.



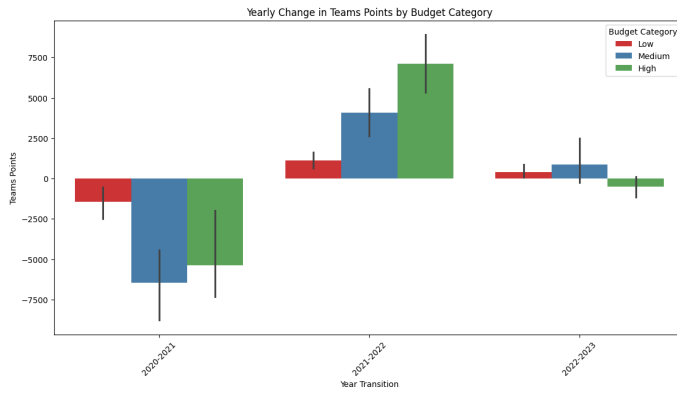


Fig. 9. Yearly change in total team points by budget category (2020-2023).

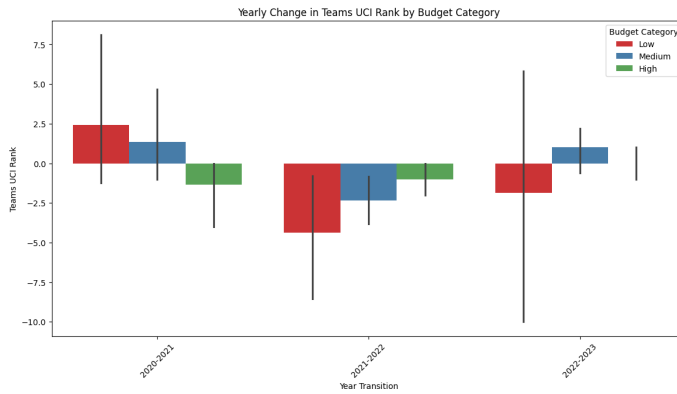


Fig. 10. Yearly change in UCI team rankings by budget category (2020-2023).

and assess broader impacts beyond point accumulation, such as sponsorship stability and financial sustainability.

#### ACKNOWLEDGMENT

##### A. Resources Used

This research relies on multiple publicly available data sources and computational tools. The primary dataset was extracted from *ProCyclingStats* [5], providing detailed race results, rider statistics, and UCI points allocation. Team budget estimates were obtained from *ProCyclingUK* [6]. Data processing and analysis were conducted using Python, leveraging libraries such as Pandas, NumPy, and NetworkX for data manipulation and graph-based analysis. Visualization and statistical modeling were performed using Matplotlib and Seaborn. Additionally, ChatGPT was used to assist in various aspects of this project, including code optimization, debugging, and structuring and refining the academic writing of this paper by improving clarity and coherence.

##### B. Code and Data Availability

The full dataset and analysis code used in this study are available in a public repository on GitHub: <https://github.com/denisrize/BigDataMininngProject>.

However, the code for the web crawler used to extract the data from ProCyclingStats is not included in this repository. Since this crawler is part of an ongoing research project for my thesis, my advisor has advised against making it publicly available until the project is completed.

#### REFERENCES

- [1] J.-F. Mignot, "The history of professional road cycling," pp. 7–31, 2015.
- [2] J. Delozier and T. Carter, "Uci protour and its impact on team behavior and race participation," *European Journal of Sports Management*, vol. 5, no. 2, pp. 112–128, 2006.
- [3] L. Lenten and R. Smith, "The effects of uci points allocation on competitive balance in professional cycling," *Journal of Sports Economics*, vol. 20, no. 3, pp. 225–245, 2019.
- [4] J. Garcia and L. Perez, "Economic disparities in professional cycling and their effect on competitive performance," *International Review of Sport and Economics*, vol. 18, no. 4, pp. 333–349, 2020.
- [5] ProCyclingStats, "ProCyclingstats - results and statistics," 2024, accessed: February 2024. [Online]. Available: <https://www.procyclingstats.com>
- [6] ProCyclingUK, "Pro cycling team budgets 2023," 2024, accessed: February 2024. [Online]. Available: <https://www.procyclinguk.com>