Personal Reflection – MTA Case Competition Project

Author: Mohammed Tajudeen Onigbanjo

1. Introduction

The MTA Data Series Case Competition places participants in the role of consultants for a newly formed Formula 1 constructor, MTA Racing. The objective is to design a driver hiring strategy that maximizes the team's point-to-cost ratio, requiring teams to determine which drivers to hire, how many to select, and what budget level to allocate under the competition's scoring framework.

This problem extends beyond simple performance prediction. Because driver prices are fixed and no budget constraint is imposed, overspending directly reduces efficiency, forcing teams to justify spending decisions rather than treating budget as given. As a result, effective solutions must balance expected race performance, cost efficiency, and risk while remaining consistent with the evaluation rules outlined in the case package.

My approach to this project emphasized data-driven decision-making over intuition. By evaluating drivers relative to their cost using the provided scoring system, I focused on identifying combinations that offered strong expected returns per dollar spent. More broadly, the project reflects my interest in applying quantitative analysis to strategic decision-making in applied, time-constrained environments.

---

2. Data Description and Preparation

One of the earliest challenges in the competition was preparing the raw datasets provided in the case package for meaningful analysis. The data were not immediately analysis-ready and required careful inspection to ensure consistency across seasons and drivers. In particular, I removed incomplete years, standardized lap time formats, and normalized driver naming conventions to allow for accurate comparisons across time.

A key focus of my data preparation process was addressing missing and irregular observations that could distort performance evaluation. This included explicitly handling DNFs and disqualifications, as well as cross-referencing related datasets to validate results and ensure internal consistency. Rather than treating these observations as simple noise, I made deliberate choices about how they should be represented given the competition's scoring framework.

I prioritized data cleaning steps that directly supported the core decision problem of evaluating driver performance relative to cost. This stage reinforced an important lesson from applied analytics: effective insights depend less on exhaustive data manipulation and more on aligning clean, reliable data with the strategic question being addressed.

---

3. Methodology

My first task was to translate the competition's scoring rules into a structured and reproducible point calculation system. Rather than treating total points as a black box, I decomposed driver performance into distinct components aligned with the case package, allowing each source of points and penalties to be evaluated explicitly before aggregation.Here is a brief description of the functions I created to assist me in this process:

CalculatePositionGainedPoints: Calculates the Points gained for Positions they gained in the main race. (+1 point for each position gained and -1 for each position lost)

FindingFastestLap: Find the driver information with the fastest lap in each race in our dataset.

CalculateFastestLapPoints: Calculate the additional Points gained for drivers with the Fastest lap. (10 points for the fastest lap)

CalculateStatusPoints: Calculates the Points for drivers with Did not finish (DNF) and Disqualified Status. (-20 points for each driver with these status)

CalculateQualificationPoints: Calculate the points earned for each driver in the Qualification Race based on their position.

The Code for these functions can be found in the github repository.

Together, these components were aggregated to produce a total points, expected points, and the Variance in the points for each driver, which could then be evaluated relative to fixed driver prices to assess expected point-to-cost efficiency.

To complement the score-based evaluation, I estimated fixed-effects regression models to better understand systematic differences in driver performance. Specifically, I ran the following 2 models:

Position ~ Driver Fixed effect + Team Fixed effect + Race Fixed Effect

Points ~ Driver Fixed effect + Team Fixed effect + Race Fixed Effect

These models isolate persistent driver-level performance after controlling for team strength (better cars and engineers etc) and race-specific conditions (some drivers performing better in some tracks than others). The estimated driver fixed effects provided a smoothed measure of underlying performance, helping to distinguish consistently strong drivers from those whose results were driven by favorable circumstances or isolated outcomes.

The point calculation system ensured full alignment with the competition's evaluation rules, while the regression results provided additional context by identifying persistent driver performance after controlling for team and race effects. This layered approach allowed driver selections to be justified not only by raw point totals, but also by expected value, variance, and underlying performance consistency, supporting more defensible point-to-cost trade-offs under uncertainty.

---

4. Results

Evaluating drivers through a point-to-cost framework revealed substantial differences between absolute performance and efficiency. While top-priced drivers such as Max Verstappen and Lando Norris consistently delivered the highest expected point totals, their high prices reduced overall efficiency when evaluated on a score basis. In contrast, several mid-priced drivers delivered competitive expected points at significantly lower cost, emerging as strong value candidates within the scoring framework.

Incorporating variance meaningfully altered the rankings. Drivers with the highest expected points often exhibited greater performance volatility, increasing downside risk in a single-race evaluation setting. This trade-off was particularly important given the competition's reliance on a single Grand Prix outcome. Drivers such as Oscar Piastri and Carlos Sainz stood out by combining solid expected performance with favorable pricing and manageable variance, making them attractive from a risk-adjusted perspective rather than purely on expected value.

Restricting the analysis to the most recent two seasons highlighted notable shifts in driver performance. Recent-form estimates elevated drivers such as Piastri and Charles Leclerc relative to their historical averages, while others declined. This finding reinforced the importance of weighting recent performance more heavily when forecasting near-term outcomes, particularly in a rapidly evolving competitive environment like Formula 1.

The fixed-effects regressions reinforced these results by isolating persistent driver-level performance after controlling for team strength and race-specific conditions. Drivers who ranked highly in the score-based analysis generally exhibited positive driver fixed effects, suggesting that their performance was not solely driven by favorable machinery or track characteristics. Conversely, some drivers with respectable raw point totals displayed weaker underlying effects once contextual controls were applied.

Taken together, these insights informed our final driver selection of **Max Verstappen, Oscar Piastri, and Carlos Sainz**, which balanced elite point potential, cost efficiency, and risk. Overall, the analysis demonstrated that optimal driver selection depends on managing trade-offs between expected points, variance, and price, rather than maximizing any single metric in isolation.

---

5. Visualization and Power BI Dashboard

To support interpretation and decision-making, I designed an interactive Power BI dashboard that translates the analytical results into an accessible, exploratory interface. The dashboard was structured to mirror the progression of the case competition itself,moving from high-level evaluation to targeted comparison and, finally, to individual driver analysis.

The landing page presents a consolidated view of key performance metrics across all drivers, including expected points relative to cost, points and position residuals, and the overall score. These visuals allow users to quickly identify trade-offs between absolute performance, efficiency, and risk, reinforcing the central insight that high expected points do not necessarily imply strong point-to-cost value. By summarizing these dimensions in a single view, the dashboard provides a clear starting point for narrowing the driver pool.

From this overview, users can interactively select and compare multiple drivers of interest. This comparison view enables side-by-side evaluation across expected points, cost, and score, allowing users to directly observe how different drivers perform under the same evaluation framework. This functionality was designed to reflect the analytical process used in the competition, where candidate driver sets were iteratively refined based on relative performance rather than raw rankings.

Finally, the dashboard supports drill-through functionality to individual driver pages, which provide detailed breakdowns of performance over time. These pages display driver-specific summaries, including total points by round and year, component-level point contributions (such as qualifying points, position gains, and status penalties), and measures of variability. This level of detail allows users to understand not only how much a driver scores, but *how* those points are generated and how consistent that performance has been.

Overall, the Power BI dashboard complements the case competition results by serving as a decision-support tool rather than a standalone analytical artifact. While the final driver selection was informed by quantitative analysis and regression results, the dashboard enabled clear communication of trade-offs and facilitated intuitive comparison under uncertainty. This integration of analysis and visualization reinforced the final recommendations and ensured that insights remained interpretable and aligned with the competition's objectives.

---

6. Discussion and Strategic Implications

The race outcome provided a useful real-world test of the analytical framework and final driver selection. Prior to Oscar Piastri's disqualification, the selected drivers performed strongly, finishing **1st (Max Verstappen), 4th (Oscar Piastri), and 7th (Carlos Sainz)**. These results aligned closely with the model's expectations, as the selected drivers combined high expected points with competitive efficiency relative to cost.

Piastri's disqualification revealed a limitation in the analytical framework rather than a failure of the selection strategy itself. While the model incorporated performance variance and historical reliability through point volatility, it did not explicitly model the probability or impact of disqualifications as a separate risk channel. As a result, the framework understated the downside risk associated with rare but high-impact events, which became evident in a single-race evaluation setting.

More broadly, the race results demonstrate that the analytical approach was effective in identifying drivers with strong underlying performance potential. Verstappen's first-place finish validated the model's identification of elite point-generating capacity, while Sainz's solid finish further supported his classification as a high-efficiency, lower-cost contributor. Even with the disqualification, the overall performance of the selected drivers suggests that the strategy was robust and well-aligned with the competition's point-to-cost objective.

---

7. Personal Reflection

This was my first case competition, and I approached it with an open mind and a strong willingness to learn. Beyond the technical aspects of the project, I gained valuable insight from the sessions organized by the competition hosts as well as from observing the approaches taken by other contestants. Seeing a range of analytical strategies applied to the same problem helped broaden my perspective on how data-driven decisions can be formulated.

One key takeaway was the importance of model selection in predictive settings. While my analysis relied on structured scoring rules and linear fixed-effects models to prioritize transparency and interpretability, several teams employed more flexible machine learning techniques, such as gradient-boosted models. Exploring these approaches could have improved predictive accuracy and revealed non-linear relationships not captured by simpler specifications.

Another area for improvement was deeper contextualization. Although race fixed effects were included to account for event-specific conditions, the model did not fully exploit differences in track characteristics. Explicitly analyzing driver performance across different track types could have provided additional insight into how contextual factors shape outcomes and may have influenced the final driver selection.

Finally, the treatment of disqualifications highlighted a limitation in the framework. While DNFs and disqualifications were penalized within the scoring system, the model did not explicitly estimate the likelihood of such events. As a result, drivers with higher disqualification risk may have appeared more attractive than they should have in a single-race evaluation. Incorporating explicit reliability or disqualification risk measures would strengthen future analyses.

Overall, this experience reinforced the importance of aligning analytical methods with the decision environment. The project strengthened my ability to balance interpretability, predictive power, and risk awareness, and it has shaped how I would approach similar case competitions and applied decision-making problems in the future.

---

8. Conclusion

This project applied a data-driven framework to the driver selection problem in the MTA Data Series Case Competition, evaluating performance through a point-to-cost and risk-aware lens rather than raw outcomes alone. By combining a transparent scoring system with fixed-effects regression analysis, the approach supported defensible decision-making under uncertainty.

The analysis informed the final selection of Max Verstappen, Oscar Piastri, and Carlos Sainz, and the race outcome broadly aligned with the model's expectations despite the inherent unpredictability of single-race events. Overall, the project reinforced the importance of balancing expected performance, risk, and cost, while highlighting how clear analytical frameworks and effective visualization can support strategic decisions in applied settings.