



DATA-DRIVEN EVALUATION OF STRATEGY AND PERFORMANCE IN FORMULA 1

*Quantifying Upgrade Effectiveness, Race
Strategy Modelling, and the Impact of the
Cost Cap and Logistical Constraints*

Bálint Csermák-Scharbert

Thesis Research Project (BENP2004)

BSc Business Engineering

Internal Advisors: Yangzhuoran Yang &
Renaud Jolivet

Internship Supervisor: Inga Stracke

June 26th, 2026

Word Count: 14,772



Abstract

Formula 1 is a sport where on track performances are defined by the hard work that goes on behind the scenes, including engineering processes, driver preparation, and transportation processes of travelling the world.

This research identifies a gap in treating these elements collectively. Prior research handles them in isolation, disregarding the notable effect of the cost cap influencing every and all decision made by Formula 1 teams. The methods used include public timing data through a shared lap level pipeline, predictive models for strategy, difference-in-differences for upgrades, and qualitative analysis for constraints.

Findings show that lap performance is highly predictable (validation MAE of 0.78s, under 1% of lap time), and strategy timing is bounded (top 3 hit rate 0.50) by safety car unpredictability. Furthermore, upgrades entail difficulties in quantification only deeming 4 out of 12 significant within the investigated upgrade packages, while development is the binding constraint in resource allocation under the cost cap.

This bounded but feasible strategy aid bodes well as a supportive model for live strategy decisions, while qualitative and partial quantitative methods clearly indicate how tackling upgrade and development decisions within the known cost and logistical constraints is critical to extract maximum performance on track.

Acknowledgements

This thesis would not have been possible without the support of several individuals who aided my research, and I am grateful to each of them.

My deepest thanks go to my internship supervisor at Pole Position Reports, Inga Stracke, for giving me the opportunity to spend this project inside the sport I love most. Her mentoring taught me how to handle formal situations and how to communicate effectively with professionals across the motorsport industry. Furthermore, the opportunities she created also allowed me to experience motorsport journalism first hand and bring my data-driven thinking into that setting.

I am equally grateful to my advisors at Maastricht University, Yangzhuoran Yang, and Renaud Jolivet. Their feedback shaped the direction of this work throughout. Yangzhuoran being my close mentor, was always open to my ideas and generous with his encouragement. The comfortable, supportive environment he provided allowed me to do my best work during this research project.

I would also like to express my appreciation to the professionals, who gave their time to speak with me and who deepened my understanding of the subject. Alex Powell from Soft Pauer guided my research by helping me judge which directions were worth pursuing and offered practical advice on building the model and building my future next to the rise of AI tools.

Catrin Mueller from DHL helped me approach my research topic from the often underrated, logistical and cost cap constraint perspective, giving direct answers to how it shapes the sport.

Steve Clark, drawing on his considerable experience in motorsport engineering, helped me successfully gauge the key objectives and perspectives that matter most over the course of a race.

Finally, I would like to thank my family for their constant support during this research. It was them who allowed me to attend to all my duties in person, and travel to different races and places during this outstanding experience.

Table of Contents

<i>Abstract.....</i>	<i>1</i>
<i>Acknowledgements.....</i>	<i>2</i>
<i>List of Figures.....</i>	<i>6</i>
<i>List of Tables.....</i>	<i>6</i>
<i>Abbreviations and Terminology.....</i>	<i>6</i>
<i>1. Introduction.....</i>	<i>8</i>
1.1 Background and Problem.....	8
1.2 Aim, Questions, and Objectives.....	8
1.3 Scope and Structure	9
<i>2. Literature Review and Theoretical Background.....</i>	<i>10</i>
2.1 Performance Modelling	10
2.2 Race Strategy and Predictive Analysis	11
2.3 Upgrade Effectiveness in Formula 1.....	11
2.4 Cost Cap, Logistics, and Operational Constraints	12
2.5 Research Gap	13
<i>3. Conceptual Framework.....</i>	<i>13</i>
3.1 Research Logic.....	13
3.2 Data-to-Model Framework	14
3.3 Upgrade-Effectiveness Evaluation Framework	15
3.4 Business Engineering Decision Context.....	16
<i>4. Methodology</i>	<i>16</i>
4.1 Research Design.....	16
4.2 Data Sources	18
4.3 Data Collection	19

4.4 Data Cleaning and Processing	19
4.5 Feature Engineering	20
4.6 Model Selection	22
4.7 Model Training and Validation	24
4.8 Evaluation Metrics	25
4.9 Upgrade-Effectiveness Methodology	26
4.10 Qualitative Methods and Constraint Analysis.....	27
5. Results.....	29
5.1 Dataset Overview	29
5.2 Model Performance.....	30
5.3 Model Error Analysis	35
5.4 Feature Importance	37
5.5 Upgrade-Effectiveness	39
5.6 Cost Cap and Logistics Findings	40
6. Discussion.....	43
6.1 Answer to Research Question	43
6.2 Interpretation of Model Results	43
6.3 Interpretation of Upgrade Effectiveness	45
6.4 Business Engineering Implications.....	46
6.5 Limitations	47
6.6 Future Research Opportunities	48
7. Conclusion.....	49
7.1 Main Findings	49
7.2 Research Contribution	49
7.3 Practical Implications.....	50
7.4 Final Reflection.....	50

<i>Future Perspective.....</i>	<i>51</i>
<i>References.....</i>	<i>52</i>
<i>Appendices.....</i>	<i>55</i>
<i>Appendix A - Tables.....</i>	<i>55</i>
<i>Appendix B - Data Dictionary.....</i>	<i>58</i>
<i>Appendix C - Model Configuration and Hyperparameters.....</i>	<i>61</i>
<i>Appendix D - Code Repository and Availability Statement.....</i>	<i>65</i>
<i>Appendix E - Generative AI Usage Declaration.....</i>	<i>66</i>

List of Figures

[Figure 2.3. Example of an FIA mandated upgrade document](#)

[Figure 3.1. Data-to-model framework](#)

[Figure 5.3. Per-race top 3 hit rate, 2025 season](#)

[Figure 5.5. Upgrade effectiveness: DiD point estimates with 95% confidence intervals](#)

List of Tables

[Table 4.1. Overview of models and methods across Parts A, B, and C](#)

[Table 5.1a. Lap-level data attrition](#)

[Table 5.2a. Lap-time prediction model performance](#)

[Table 5.2b. Tyre degradation model performance](#)

[Table 5.2c. Pit classifier performance](#)

[Table 5.2d. Value regressor performance](#)

[Table 5.2e. Strategy policy performance](#)

[Table 5.3a. Per-race top 3 hit rate, 2025 season](#)

[Table 5.3b. Top 3 hit rate summary by race type](#)

[Table 5.4a. Feature groups by category](#)

[Table 5.4b. Key findings on feature importance](#)

[Table 5.4c. Degradation feature detail](#)

[Table 5.5a. Summary of upgrade-effectiveness results](#)

[Table 5.6a. Cost cap inclusion/exclusion matrix](#)

[Table A.1. Missing values by column](#)

[Table A.2. Upgrade effectiveness: difference-in-differences estimates](#)

Abbreviations and Terminology

Term	Meaning
Abbreviations and acronyms	
AI	Artificial Intelligence
AP	Average Precision
API	Application Programming Interface
AUC	Area Under the Curve (see ROC-AUC)
CFD	Computational Fluid Dynamics
CI	Confidence Interval
CSV	Comma-Separated Values
DiD	Difference-in-Differences
DRS	Drag Reduction System
FIA	Fédération Internationale de l'Automobile (Formula 1's governing body)

Term	Meaning
GBM	Gradient Boosting Machine
GPS	Global Positioning System
HistGBM	Histogram-based Gradient Boosting Machine
JSON	JavaScript Object Notation
MAE	Mean Absolute Error
MRR	Mean Reciprocal Rank
NaN	Not a Number
R²	Coefficient of Determination
RMSE	Root Mean Squared Error
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve
SC	Safety Car
VSC	Virtual Safety Car
Formula 1 and domain terms	
Aero-balance	The front-to-rear distribution of aerodynamic load on the car
B-spec	A major mid-season car upgrade amounting to a substantial redesign
Cost cap	The FIA-imposed annual limit on performance-related team spending, introduced in 2021
FastF1	Python library providing Formula 1 timing and telemetry data; the primary data source
Green-flag laps	Uninterrupted racing laps, with no Safety Car, Virtual Safety Car, or yellow flag
Ground-effect era	The 2022-2025 regulation set, which relies heavily on underbody aerodynamics
OpenF1	Open API for Formula 1 timing and telemetry data; identified as an optional future extension
Pit window	The range of laps over which a pit stop is strategically optimal
Stint	A continuous run of laps on one set of tyres between pit stops
Tyre compounds	Soft, medium, and hard slicks plus intermediate and wet tyres, varying in grip and durability
Undercut / Overcut	Pitting earlier (undercut) or later (overcut) than a rival to gain track position
Triple Header	Three races held on three consecutive weekends
Modelling and statistical terms	
Counterfactual target	An outcome observed for only one branch of a decision; the central difficulty for the value regressor
deg_penalty_vs_fresh	Derived feature: predicted pace loss of the current tyres versus a fresh set of the same compound
DiD estimator	Measures a team's performance change against the field's change over the same window, netting out general development
Gradient boosting (GBM)	An ensemble model built from decision trees added sequentially
L2 regularisation	A penalty that shrinks model coefficients to improve cross-season generalisation
Logistic regression	A linear classifier for binary outcomes (here, pit versus no-pit)
Monte Carlo simulation	Repeated random sampling used to model probabilistic race outcomes
One-hot encoding	Representation of a categorical variable as a set of binary indicator columns
Polynomial regression	Regression fitting a polynomial curve; applied per circuit-compound for tyre degradation
Top 3 hit rate	Share of driver sessions in which the actual pit lap falls within the model's three highest-ranked laps

1. Introduction

1.1 Background and Problem

Formula 1 is an ever-changing sport, requiring constant engineering and operational innovation, while adhering to cyclic regulation changes that overhaul established methods. It is an engineering sport, meaning that a driver in a better-engineered car can outperform a more talented driver who lacks a high-performing machine. This norm is broken either by driver excellence or team strategy decisions, as getting strategy precisely right can earn substantial reductions in race time. These attributes of the sport are exactly why data analysis and evaluation are so vital, and why approximately 50 engineers work on each team each race weekend (Mercedes-AMG F1 Team, n.d.).

Depending on variables such as practice run plans, race time, and other factors, a modern Formula 1 car normally generates 1 terabyte or more of data over a full race weekend (Mitchell, 2022). These data sets are incredibly vast, containing key information that can be analysed, understood, and applied for future improvement. It gives an understanding of position compared to rivals in telemetry graphs, or where a team's race car package is stronger or weaker against the competition.

As teams identify where most time can be gained, development starts as soon as the next hour, with engineers back at team headquarters beginning to plan, design, execute, and implement. Since 2021, this process has been limited. The introduction of the 'cost cap' means teams are limited in their budgets for upgrades or other development that affects performance. Alongside budget constraints, development allowance was also attenuated. The use of digital simulations, for example, in computational fluid dynamics and, crucially, wind tunnel testing, where aerodynamic elements are advanced, has been subject to regulations. These major limitations require teams to make decisions about spending, testing, and new-part development very cautiously while ensuring that data and actual car performance are perfectly aligned. These measures were introduced to allow smaller teams with less sponsorship-backing and resources to have a close-to-level playing field with the major teams in Formula 1 (Barretto, 2020).

1.2 Aim, Questions, and Objectives

A problem arising from these circumstances is that teams must make upgrade and strategy decisions under considerable constraints. The approach of the engineering, logistics, and

business departments must adapt. Engineers and operational staff do not have the leeway to explore a wide range of potential upgrades or logistics options. It is imperative to validate as few design variations as possible, to allow focus to shift only on what is proven to work.

Existing research treats tyre degradation, lap time prediction, strategy, or cost cap mostly in isolation (Cappello & Hoegh, 2025). This project identifies a clear gap in these methods: the research treatment of race operation is fragmented, with little work combining performance and strategy modelling, upgrade effectiveness, plus cost cap and logistics constraints within a single framework. This often blurs the external perspective, as these studies do not account for the challenges inherent in such highly constrained environments. The relevance of this topic lies in the potential to support better decision-making for teams, especially midfield ones with a lower budget. Moreover, this study is informed by three semi-structured interviews with industry professionals, guiding the research by either confirming or disproving pursued methods. Their experiences act as a reliable, and primary source contribution, right from the world of Formula 1.

1.3 Scope and Structure

This thesis research answers the following question:

To what extent can data-driven methods quantify the impact of Formula 1 car upgrades and race strategies on on-track performance under cost cap and logistics constraints?

In order to provide meaningful and coherent research, this study further addresses the following sub-questions:

- 1. How reliably can tyre degradation and stint length be modelled to predict pit stop windows and strategy choices?*
- 2. How effectively can a combined predictive and simulation-based policy recommend race strategy decisions?*
- 3. Do in-season aerodynamic and mechanical upgrades result in measurable on-track performance gains, as evidenced by a difference in differences (DiD) analysis?*
- 4. What do cost cap and logistics constraints reveal about how teams prioritise development and operational decision?*

Together these define the main research question as follows: The first and second of the sub-questions assesses impact of the race strategies, and whether it is possible to simulate race event outcomes. The upgrade impact, as the third sub-questions aims to find out if mechanical upgrades bring measurable track-time, as shown by the data. Lastly, sub-question four will provide insight on how logistics and cost cap constraints affect these actions.

To address the research gap, this thesis contains three key sections in the form of part A, B, and C. Using quantitative and qualitative methods, it follows standard procedure, reviewing literature and defining existing knowledge and the research gap. It provides a deep analysis of the approach taken for research methodology, with part A addressing sub-questions 1 and 2, Part B addressing sub-question 3, while Part C addresses sub-question 4. The results of these elements are explored and detailed with carefully constructed tables, complemented by feedback from industry professionals. Lastly, this paper presents a discussion underlining the outcome, and business engineering implications after which it draws a conclusion and assesses future research directions.

2. Literature Review and Theoretical Background

2.1 Performance Modelling

Key findings show how multi-modal and high-frequency time-series analysis has a broader literature relevance for telemetry-style race data. In a study by Jiang et al. (2025), multi-modal time-series with attention-based architectures are presented as techniques to integrate heterogeneous inputs such as numeric sensor streams, categorical context, and event decisions when data is structured at a level such as driver-lap.

Panel-data econometric models with driver and circuit fixed effects have been used to explain variation in lap times (Pasz, 2025). Other approaches such as Brusik (2024) showcase a more complex, and data hungry method in performance modelling. The gradient boosting used in this thesis research project sits between these two edge cases. It differs mechanism wise compared to an econometric model but is less complex than a deep learning time-series model. This flexibility allows it to make predictions on unseen data and not be restricted to post-race analysis.

2.2 Race Strategy and Predictive Analysis

Recent contributions highlight how advanced statistical and time-series methods can be applied to motorsport data. Cappello and Hoegh (2025), for example, capture both a latent degradation process and observation noise in lap time data to model tyre degradation using a state-space framework. The outcome of their research explains that tyre wear can be represented as a hidden state evolving over race laps, which inherently drives changes in lap time and can be estimated from timing data.

Race strategy has been further examined through simulations, optimisation, and game theory. Monte Carlo methods have been applied to race simulations to incorporate probabilistic effects such as safety cars and incidents (Heilmeier et al., 2020). Technical sources confirm that scenario-based planning is standard in F1 strategy (Hatton, 2024; Aramco F1 Team, 2024).

Public APIs such as OpenF1 and FastF1 provide lap times, sector times, and tyre data for analysis. Using these data sources, the author developed [DeltaTrace](#), a web-based F1 data analysis platform built with HTML, CSS, and JavaScript, which forms part of the foundation for this project (Csermák-Scharbert, 2026).

2.3 Upgrade Effectiveness in Formula 1

The scarcity of public research on evaluating upgrade effectiveness partially sets up a future limitation explored in this research. Even though performance assessment is key in Formula 1, the difficulty in quantifying mechanical upgrade impact lessens the motive to attempt it.

On the other hand, qualitative information is abundant, as teams are mandated to release documents detailing which areas of the car were changed, and what effect it brought with it.



FIA FORMULA 1 WORLD CHAMPIONSHIP



Car Presentation – Barcelona-Catalunya Grand Prix SCUDERIA FERRARI HP

	Updated component	Primary reason for update	Geometric differences compared to previous version	Brief description on how the update works (min 20, max 100 words)
1	Front Wing	Performance - Local Load	Revised front wing footplate with new vanes arrangement and endplate diveplane addition, optimized wing elements spanwise loading distribution. New SM mechanism links and integration in the nose, revised nose shape with raised lower surface	Evolution of the front wing platform raced so far, with enhanced tip flow features coming with front wheel wake control benefits as well as increased aero balance capacity. More integrated SM mechanism links returns tidier centre line flow features. It also comes with flap lips modulation.
2	Front Wing Endplate			
3	Nose			
4	Floor Body	Performance - Local Load	Reduced keel volume. Redesigned front floor leading edge profiles and claws, front floor board elements optimisation (horizontal and vertical ones). Optimisation of rearward floor corner winglets, reprofiling of diffuser sidewall cutout. Updated diffuser winglet angle of attack and diffuser / tail expansion profile	Benefiting from an improved onset flow quality, this entire floor update is targeting overall aerodynamic load increase across the main car operating window as well as flow features robustness improvements together with improved flow topology and energy towards the rear corners
5	Floor Board			
6	Floor Edge			
7	Diffuser			
8	Sidepod/Coke	Performance - Flow Conditioning	Inflated sidepod shoulder and adapted cokeline	This geometry has been developed in conjunction with the floor update and manipulates front wheel wake together with the floor board update, as well as the rebalancing the front floor pressurisation

Figure 2.3. *Example of an FIA mandated upgrade document*

To prove performance gains of upgrades, this research showcases a quasi-experimental tool in the form of Difference-in-Differences (DiD). This method was chosen as the main candidate after Zhou et al., (2024) has proven its benefit of before and after evaluations, while considering the field's constant development alongside.

2.4 Cost Cap, Logistics, and Operational Constraints

Technical reports and public cost cap documents describe trade-offs in resource allocation and upgrade timing under the cost cap (Coleman, 2025). Precise documentation on how budget is allocated down to the dollar within a team is not public information, though all teams must publish year-end balance sheets, from which broad estimates can be made (FIA, 2024).

The gap arises from the problem how sources reveal the scale and the rules. The lack of a line-by-line allocation makes the quantitative conversion of budget into performance difficult, calling for a qualitative approach.

2.5 Research Gap

The elements described prior intertwine substantially and affect a team's performance over a season. Through telemetry data a team can assess weaknesses of their current package. Once recognised, development starts and new parts are strategically planned for an upcoming circuit where it can be well assessed in performance. During these decisions, teams consider where they will deliver this upgrade to lessen logistics cost but must also contemplate what cruciality level the upgrade possesses.

If it is necessary to manufacture it as soon as possible, all other development is pushed to the side, and the part is rushed to be designed, manufactured and shipped. These decisions cost teams critical resources in the form of restricted computational fluid dynamics (CFD) time and money. In case the need for this upgrade was an emergency, meaning if it remained unresolved, the team would have faced reliability issues, then it eats away at their yearly allowance, consuming allocated resources for future development in the season later (Cleeren, 2021; FIA 2024).

This process showcases all angles that ought to be considered when striving towards greater performance. However, there is limited work that integrates performance and strategy modelling, upgrade effectiveness, plus cost cap and logistics constraints into a single framework for the current regulatory era, which motivates the research design described next.

3. Conceptual Framework

3.1 Research Logic

To provide an example, imagine the following scenario. A team recognises how in the opening race of the season (Australia or Bahrain in recent years), they were lacking speed in high-speed corners compared to their competitors, losing 1-2 tenths of a second on qualifying lap. At this point in the season, upgrades are already in development, as data from pre-season testing has been processed. This previously unknown variable urges the team to get started on fixing this issue.

As revealed previously, this isn't done so easily. Through data, the team must identify which mechanical parts ought to be upgraded. That information is sent to the factory where development starts, with a greatly shortened lead time. Once the part is complete, it must be

shipped from Europe, inducing extra costs. In short, the timeline starts by modelling the pace of the car, then using that to evaluate strategy and upgrades, and lastly assessing the development path under constraints.

This example is possibly the simplest one a team might face during a season. It is imperative to understand how every new component serves to improve performance, and at the same time must obey to all constraints. This research strives to showcase the advantage of considering these segments collectively, and how an issue in one department can disturb all other ones.

Moreover, it prioritises generalisation to unseen conditions over in-sample fit. It aims for models that apply across a wide range of race conditions and therefore prioritises generalisation. This entails deliberate design choices, such as the usage of seasonal data splits and the use of simpler models, which may yield weaker in-sample metrics in exchange for more robust performance on unseen data. This allows better handling of regulation changes and race-to-race unpredictability.

3.2 Data-to-Model Framework

The research obeys the following pipeline. Data is extracted from open-source APIs, then sorted into a canonical structure labelled in different table setups. The cleaned dataset allows for features to be engineered into defining variables defining model understanding and accuracy. To support coherence, one shared feature layer feeds parts A and B to ensure they are drawn on the same lap-level evidence, while part C draws on a separate qualitative evidence base, but all three are integrated under the same constraint lens, as shown by Figure 3.1 below. Moreover, the project was made with live decision support in mind, to allow implementation on incoming real-time information. This live decision support functions through the implementation of online-safe features, meaning it only uses information available up to the current lap.

Lastly, it is vital to clarify the difference between relative versus absolute measurements. This research measures performance relative to the field, not in absolute terms, because access to teams' internal absolute benchmarks is classified. Therefore, a team's pace advantage is the field median minus their own time. As it is the nature of motorsport, this limitation is not so detrimental, as what matters most is where you stand relative to your competitors.

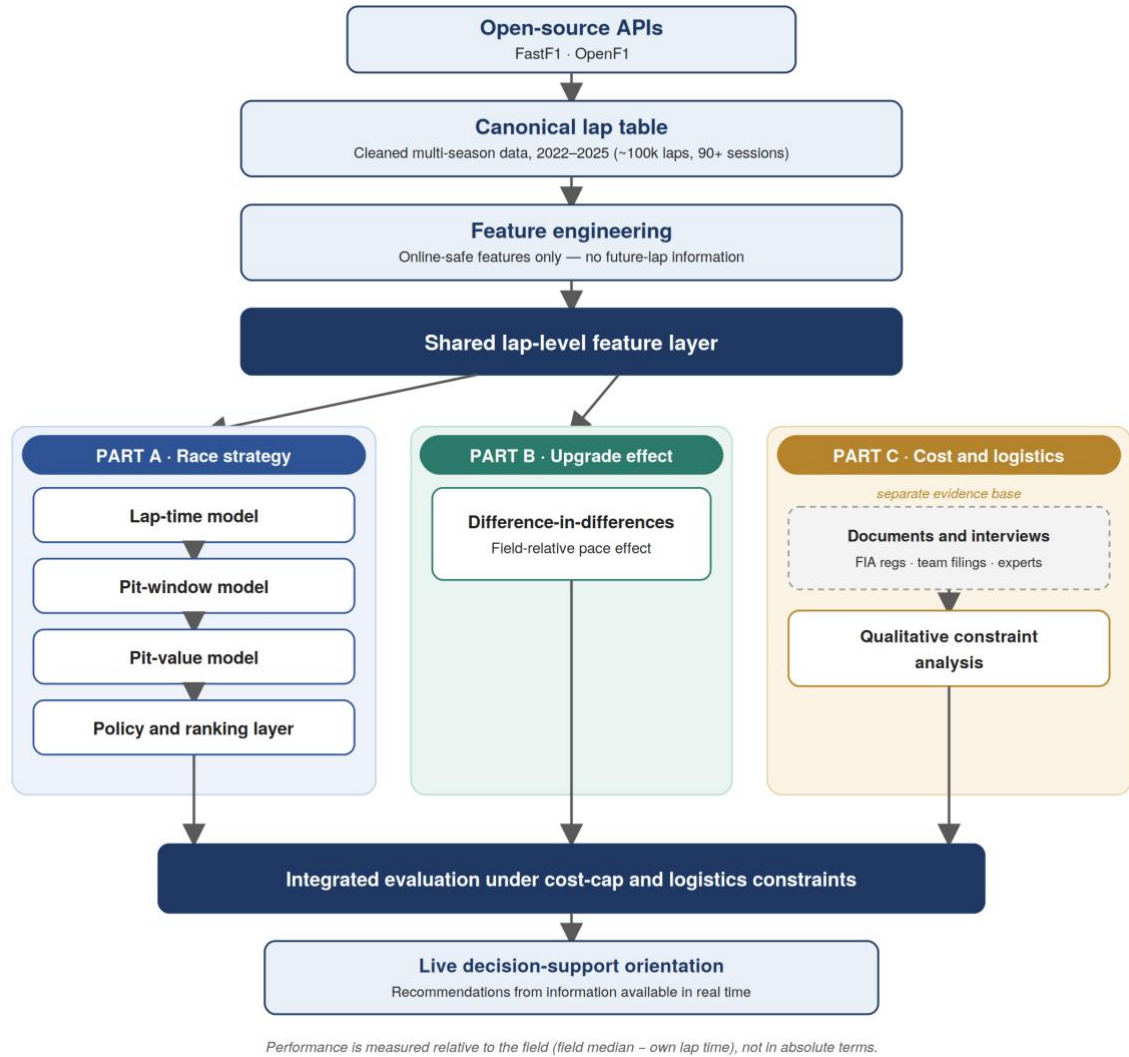


Figure 3.1. Data-to-model framework linking the shared lap-level pipeline (Parts A and B) and the qualitative constraint analysis (Part C)

3.3 Upgrade-Effectiveness Evaluation Framework

The DiD estimator is the key to achieving a possible assessment of upgrade effectiveness on raw data alone. It measures a team’s change against the field’s change over the same window, leading to general development being netted out. An absolute calculation of before and after performance would be misleading, as the whole field often develops in unison. This estimator uses the same clean lap data and pace definitions as Part A, indicating the kind of correlation research publications lack.

3.4 Business Engineering Decision Context

The investigated segments of what make up decision-making under constraint in Formula 1 is precisely an objective that closely relates to the business engineering coursework and programme objectives. This conceptual link between technical work and team decisions accurately showcases the informed choices made under cost cap and logistics constraints. Without the correct management of resources, engineering solutions will struggle alongside. It is only when taken together that the three parts describe the full picture of how a team should allocate their assets. Data science, operations, and strategy converge on one objective, and overall contribute to greatly improved performances.

4. Methodology

4.1 Research Design

This study uses a mixed quantitative and qualitative design suited around three connected parts introduced earlier. Part A develops and examines four supervised models (a gradient boosted lap time regressor, per circuit tyre degradation regressions, a logistic regression pit window classifier, and a ridge value regressor) combined into a strategy-ranking policy, summarised in table 4.1. Part B assesses upgrade effectiveness using a limited quantitative comparison where public data allows it, and is supported by qualitative interpretation where upgrade timing, track specificity, and confounding variables prevent clean causal measurement. Part C analyses cost cap and logistics constraints qualitatively, because the internal cost allocation, manufacturing, freight, simulation and resource-prioritisation data required for full quantified results, are not available in public datasets.

On another note, it is important that the three parts are not independent studies stitched together. They share one data pipeline, one canonical lap table, and one set of feature definitions, which allows the upgrade discussion in Part B and the cost cap framing in Part C to refer to same lap-level evidence that drives the strategy models in Part A. The overall logic is to firstly establish the extent which lap times, tyre behaviour, and pit-window choices can be modelled from public timing data, then use that foundation to identify where performance fluctuation signals are interpretable against the financial and logistical limits under which Formula 1 teams operate.

Part A attempts to follow predictive work in a supervised-learning workflow. Commencing with target definition. This means establishing a baseline for investigation, in the form of a single driver, or a teammate pairing. Feature engineering is then the consequent important step, which should always include features available at the time of prediction. As Formula 1 is an ever-changing sport, vehicle dynamics and capabilities change, requiring frequent surveillance. Closing stages of the workflow include fitting a given model on historical seasons and evaluating them on a held-out season. Upgrade work in Part B follows a simpler, field-referenced comparison logic. When an upgrade is introduced, the treated team's performance change is compared with the extended field over a race window, while keeping in mind that overlapping upgrades, track dependence, weather, tyre behaviour, and setup variation limit causal interpretation. Part C interprets and draws information on FIA financial regulation material and official cost cap descriptions, which explains the cap as a limit on certain performance related costs while leaving teams free to allocate resources within that limit (FIA, 2024). A summary of the models used can be found in the table 4.1 below.

Table 4.1. *Overview of models and methods across Parts A, B, and C.*

Model / method	Target or purpose	Key features / inputs
Lap-time model: histogram-based gradient boosting	Predict lap time (lap_time_seconds)	Tyre compound and age, stint, position and gaps, pace history, weather
Tyre-degradation model: degree-2 polynomial ridge (86 circuit $\times$ compound group models)	Predict pace deviation (lap minus clean-pace median) against tyre age	Tyre age (linear and quadratic terms), fitted per circuit and compound
Pit-window classifier: logistic regression (L2, class-weighted)	Probability the driver pits within the next three laps	Tyre age, pace delta, gaps to rivals, degradation features, opponent threat
Value regressor: ridge	Net time gain over six laps, pitting now versus staying out (counterfactual)	Candidate-lap pace, gaps, tyre and degradation state
Strategy policy: three- gate ranking system	Rank candidate pit laps using the models above	Combines classifier probability, value estimate, and a degradation gate
Difference-in- differences estimator	Isolate an upgrade's relative pace effect, netting out field-wide development	Pace advantage (field median - own clean median); treated team vs field, 3 races before / 4 after; 500- sample bootstrap

Model / method	Target or purpose	Key features / inputs
Qualitative document and interview analysis (no predictive model)	Establish which costs the cap binds and how logistics shapes development timing	FIA Financial Regulations, published team accounts, three expert interviews

4.2 Data Sources

Schaefer's (2026) FastF1 version 3.8.3 Python library has been chosen as the primary data source for this research. FastF1 is a package for accessing and analysing Formula 1 results, schedules, timing data, and telemetry, with data present as Pandas DataFrames and caching available for recurring API requests (Schaefer, 2026). This allows it to act as the foundation for a reproducible lap-level data pipeline, because similar extraction processes can rerun across multiple events and seasons using a consistent interface.

Alongside, OpenF1 was implemented as an optional complementary data source for future live or near-live extensions of the project. Based on implementation, licenses ought to be explored, as commercial activation using live data is restricted by Formula 1 rights. OpenF1 similarly provides historical telemetry, timing, and session data in JSON and CSV formats, including endpoints for car data, intervals, laps, pit events, position, race control information, stints, team radio, and weather (OpenF1, n.d.). However, OpenF1 was not integrated into the analytical pipeline for this thesis. The loader exists only as a stub, and all results reported in the thesis are built on FastF1 data and focused on data from the previous regulation set spanning 2022-2025. This is important to recognise, as replication does not depend on OpenF1 being active.

In Part C, financial regulation information is gathered from FIA Financial Regulations and official Formula 1 explanations of the cost cap. Stated in the regulations, the cost cap limits certain costs incurred by or on behalf of a team while leaving the team free to decide how to allocate resources within that cap. Set objectives of this cap include competitive balance to lessen budget allowance differences between teams bigger and smaller teams, support sporting fairness, and long-term financial stability (FIA, 2024). Formula1.com exhibits this cap as a limit on spending that impacts race car performance and identifies key exclusions such as marketing costs, race-driver fees, and the three highest-paid personnel (Barretto, 2020).

4.3 Data Collection

The dataset used for model construction is astoundingly rich and covers 101,297 laps in 4 years of race sessions. Vehicle GPS is available every 3 metres completed on the racetrack, alongside detailed telemetry of car behaviour. On the other hand, this implies dealing with incredibly large datasets. This research used local caching for a fixed, and highly reproducible data setup.

Furthermore, it is important to note why a four-season window spanning 2022-2025 was chosen. That window of opportunity makes use of the previous regulation set which is labelled as the ‘ground-effect’ era. Formula 1 by design is a sport that often experiences change, meaning certain models may only be applicable for a specific ruleset, while becoming obsolete for another. This is a key limitation and can be best mitigated by close monitoring and constructing a model suitable for new variables that may arise when rules change.

4.4 Data Cleaning and Processing

This is the stage where raw and messy session data is sorted into a clean and uniform table, reliable to use in all further steps of this research. This section explains five different cleaning decisions used.

First, is the construction of a 19-column canonical schema. This table is frozen and identical project-wide, to allow data samples to be gathered into one uniform group. FastF1 sessions don’t all look the same, as column names, available fields, and data types vary between events and seasons, such a reliable baseline is necessary to avoid missing variables, and discrepancies among sessions. As the primary step, every session gets converted into a single fixed table with the same 19 columns, in the same order, every time. Main benefit is that every feature is able to assume uniform input shapes and avoids breakage because of insufficient formatting.

Secondly, preprocessing validation checks were carried out before a converted session is accepted into the combined dataset. The method for such validation includes a set of sanity checks, including lap and stint numbering being internally consistent, tyre compounds matching known categories. If any session fails, it is noted down in a log and skipped rather than allowing it to contaminate clean sets.

After that, a ‘pit-out exclusion’ feature was implemented. A pit-out lap is the lap on which a driver rejoins the track from the pit lane, after making a tyre change. The values of such a lap reflect the slow speed travelled in the pit lane, and the cautious subsequent lap through which

a driver must warm up their tyres to optimal temperature, disallowing them to push. The out laps do not reflect the car's real pace, and depending on the track can be 10-20 seconds slower than a racing lap. Leaving it in would distort any analysis of green-flag running, and as a result, these laps are flagged and excluded from pace analysis.

Next objective was defining a 'valid green lap'. This single filter decides which laps count towards an overall sum of representative racing laps used for fitting the pace and tyre degradation models. A lap qualifies only if the following three hold. It wasn't deleted (official Formula 1 timings system invalidates some laps, for example track limit breaches), it isn't a pit-out lap, and it ran under green track status (no safety car, virtual safety car, or yellow flag in a designated sector). Once defined, this method helps to keep the pace signal clean and consistent.

Lastly, the research investigated the effect of weather on clean racing laps. As weather data arrives on its own clock, meaning at a specific time over the race, it cannot simply be slot into a per-lap table. To counter this, 'nearest-time merge' method was used, in order to pair each lap with the weather reading closest in time to that lap. That is how a lap gains air temp, track temp, wind among other attributes. Moreover, weather sits outside the frozen 19-column schema, as it has a different time base and isn't always available. When weather is unavailable, the columns are still created and filled with missing value placeholders, rather than left absent. This keeps the column intact and alive for the models, making it reliable for future applications.

Each of these rules strive to standardise data shape, or abandon information that would mislead a model, such as the pit-out laps. Together they support an immensely important baseline, allowing research on clean, and representative data.

4.5 Feature Engineering

This following section is perhaps the most crucial element of this whole research, as it is a detailed construction of the exact features that a predictive model will consider for strategy recommendations.

First and foremost, the research must establish a principle that disallows future-lap lookahead. Such a feature can be referred to as an 'online-safe' feature, meaning that at any given lap, it only uses information that existed by that lap. Nothing can be considered from a future lap. The reason why this feature is truly important, is because the strategy models are meant to

function live, mid-race, where future events are unknown. If information was known even from the next one or two laps, then in testing the model perform well, but would be obsolete in an actual race situation. To avoid this kind of leak, this must be established from the get-go. The following paragraphs will further explore groups of related features.

Tyre and stint features are tasked to describe the state of the current tyre. It is listed by compound, stint number, tyre age in laps, the lap index within the current stint, a first lap of the stint flag, and lagged tyre age values. As tyre state is the single biggest indicator of lap time change in a race, this group carries a significant part of the signal.

Position, delta gaps, and traffic features resemble where the car sits in the race. Running position, change in position since the last lap, field size, position percentile, gaps to leader and to other cars just ahead and behind, alongside derived indicators like clean-air flags, boxed in flags, and a rejoin-window estimate. This group captures crucial strategic context, as a pit stop needs to be assessed not by only tyre wear, but by traffic and track position. A driver may extend or shorten stints, in order to avoid coming out of the pit lane and encountering several slower cars.

The pace history feature group exhibits trends observable from a driver. It provides information on the previous one and two lap times, a rolling three lap mean, an expanding session mean, the lap-over-lap delta, and a signal comparing recent pace against the session average. This group further helps in recognising patterns when calculating tyre degradation time values.

Session context features help in providing an overall look at the state of the race, rather than an individual car. It looks at rates of track-status anomalies, the whole field's pit-out rate, the share of drivers that have already pitted, and the rate of position change across the field. In summary, instead of treating each lap in isolation, it looks at how the race unfolds. Through this approach, a wave of stops, or a chaotic phase can be recognised and acted upon.

Weather features are a unique set, as they measure air and track temperature, humidity, wind speed, a wet-race flag, and terms like change in track temperature since session start. In raw form, it is close to impossible to interpret and precisely assess how and to what extent it influences car performance. Even small changes in air pressure can affect power unit performance, while changes in temperature affect tyre warmup, degradation, and cooling rates. The metric this research aims to model best, is wind direction. It is of vital importance, as head or tail wind has different effect on aerodynamics, and as sometimes direction changes throughout a race, it can influence a drivers approach to corners. As wind direction is circular,

it is encoded as the sine and cosine of wind angle rather than raw degrees. When mapping each angle to its $(\sin\theta, \cos\theta)$ coordinates, they are placed on a unit circle, so 359 and 1 degrees fall almost on the same point, and the model interprets them as near identical. Otherwise, in raw degrees they would appear 358 degrees apart.

Tyre degradation is the bridge from fitted degradation model into the strategy model. It takes each lap, and reports the model's predicted pace deviation at the current tyre age, the observed deviation against that prediction, projected additional pace loss over the next five and ten laps, the penalty of the current tyre against a fresh one of the same compound, and the fitted linear and quadratic degradation terms for a certain situation. It may be a handful, but all these considerations must be observed, as all of them affect the decision whether to change tyres, or not.

Opponent threat features survey immediate strategic threat from nearby opponents. Using a position-keyed lookup across all drivers on the same lap, they pull tyre age from different situations. Firstly, the tyre age of the car directly ahead and behind, but also the tyre life difference, signalling an under- or overcut opportunity. These two continuous scores combine those differences with the gap to the relevant car.

Lastly, referring to section 4.4, a missing-value placeholder has also been put into place. Whenever a feature depends on data that is unavailable for a session, the column is still produced and filled with a placeholder, rather than dropped. This ensures identical sets of columns for reliable use.

4.6 Model Selection

To achieve a most accurate prediction model, a combination of models and classifiers have been used. First one being a lap time model 'lap_time_seconds', targeting the objective of predicting lap time from tyre state, position and gaps, pace history, and weather. In this research, two estimator methods have been tried, a ridge regression as an interpretable linear baseline, and a histogram-based gradient boosting regressor. After evaluation ([Section 5.2](#)), gradient boosting was selected as the primary model, because lap time depends on interactions a straight line is unable to represent. To clarify, the effect of tyre age depends on compound, while the effect of a gap depends on track position. Trees can capture these automatically.

Following this was ‘pace_deviation’ a model targeting to assess tyre degradation. It includes a lap time, minus the driver’s own clean pace median for that session, which strips out driver and circuit baseline so only degradation values remain. As tyre degradation is non-linear and accelerates as the tyre approaches its cliff of performance, a squared term was implemented to capture the curve in the form of a degree-2 polynomial ridge regression. It is important to note, that the reason for fitting separate models per event rather than a global one, is justified by degradation being circuit and compound specific. In the model the polynomial bias was dropped, as it makes two coefficients directly readable as the linear degradation rate (in the form of seconds per lap) and the quadratic acceleration.

The most important component of the predictive model is the classification system based on logistic regression with weighted classes. It is designed to tell the user at what point a pit stop is most optimal for a driver, which has two outcomes, either the driver pit or did not pit. It predicts the likelihood of the driver pitting on any given lap, by using tyre condition, gap to competitors, and race pace differences to underline pit window openings. Class weighing is used, because only ~9% of observation lead to a ‘pit soon’ event, and it stops the model from defaulting to ‘won’t pit’. An alternative approach using gradient boosting was also attempted, but performed much worse across season than the model defined above (see [Section 5.2](#))

Furthermore, a model evaluating net time gain from pitting now, versus staying out, measured over the following six laps has been implemented. The target is constructed as the difference between the realised and the alternative. In the case of this model, ridge has been chosen over boosting, as for the relatively few genuine pit decisions rows, boosting produced inflated estimates, over triggering the recommendation gate. It was evident that ridge’s shrinkage kept it more conservative. The reason for a six lap lookahead is that over four laps the pitlane time loss cannot be recovered, leading to the target being structurally negative.

Moving on to the final three model analyses, first, let’s look at the strategy policy layer aiming to turn the models into one recommendation. It combines the classifier and the value model over three gates. Any racing lap is only recommended if the pit probability, and expected-value score overcome a threshold, and the estimated rejoin position loss is within an acceptable window. All three must pass at once, raising precision. The decision threshold is calibrated to maximise the F0.5 score, which weights precision twice as heavily as recall. It is made sure to favour precision, as a strategist will act on flagged laps, making false alarms costlier, than a scarce miss. Alongside this binary ‘pit now’, the model produces a ranked watchlist of the most

optimal upcoming laps, which is precisely a form a strategist would use, and it is also why the top 3 hit rate is the main metric that matters.

The penultimate showcase in this section presents the Monte Carlo race simulation, which was used to stress test the decision under uncertainty. As this component generative rather than fitted, it is not evaluated with MAE or R^2 , which apply to other supervised models. From a given lap, two thousand runs per scenario are completed, where the model simulates the rest of the race. Each candidate strategy is considered, whether to stay out, or pit on current or upcoming few laps. Each predicted lap time is the degradation model's prediction, alongside a Gaussian noise draw, to make runs scatter realistically. The safety car is a per-lap Bernoulli event, meaning it compresses the whole field rather than helping one car in isolation. There was also an attempt to assess undercut and overcut dynamics in the form of a supporting opponent model. It is a simple model on purpose, as it projects rivals forward at their average pace with no pit decisions of their own. The limitations of this structure will be revisited later.

The key takeaways from this subsection ought to be how gradient boosting was used only for the lap time model, where it was backed up by accuracy results and notably helped. Regularised linear models are applied everywhere else, as predictions need to transfer across seasons. The superior cross-season transfer of the simpler model is itself a central methodological finding. On this sample of ground-effect regulations, the more constrained models generalised more reliably than the flexible ones, and as interpretability has real value in a strategy context for an engineer, it is the correct approach to pursue.

4.7 Model Training and Validation

The following subsection strives to answer two questions. How was the data split, and how were the models tuned?

The data was divided by season, not randomly in a season-based split. First batch from the timespan of 2022-2024 (approximately 75,000 laps) realised for training, 2025 used as the validation test (approximately 27,000 laps), and 2026 upheld as an external set. As the 2026 season has only concluded 7 races at the point of publication, sample size is lacking for detailed and renewed evaluation. The justification behind this data-split approach is the avoidance of leakage. A random split would likely scatter laps from the same race into both training and validation, which would give the model the ability to see a part of the race it is simultaneously

being trained on. Such an event would lead to inflated scores while possibly hiding the exact overfitting worth investigating. A season split forces the prediction of a year it never trained on, providing an honest test for a forecast-like tool.

The 2026 season has been treated separately, as it cannot simply be discounted to just ‘more validation’. It is used as an extreme stress test of how the model holds up when major underlying systems shift and datapoints renew, in this case along with the immense 2026 chassis, and power unit changes. The cars behave drastically different, even losing key attributes such as the drag reduction system (DRS), and 25 to 30% of downforce (Barretto, 2024). This is all without considering how the data variables themselves change. Overtake mode, straight mode, and overall energy management greatly affect race outcomes, and they don’t even exist in pre-2026 data samples. Overall, training on an abundant regulation set is advantageous, as it provides massive datasets, but to perfect a strategy model capable of reliable prediction across seasons, close monitoring is necessary with regular updates to any change in data acquisition (Barretto, 2024).

Furthermore, hyperparameters have been tuned against the 2025 validation season. The final configuration for the lap time gradient boosting model came from a search over the number of boosting iterations, learning rate, tree size, and L2 interactions. The biggest takeaway was the low learning rate with a moderate number of iterations with constrained trees, along and raised validation error. The linear models’ regularisation strength has been checked in a similar fashion.

4.8 Evaluation Metrics

Before discussing quantified metrics, this subsection showcases the role of each appropriate evaluation metric in the case of this research.

Continuous models play a major part in lap time and tyre degradation assessment. As these models deal with predicting time losses or gains (in seconds) throughout a race event, they foster mean absolute error (MAE) evaluation, reading directly as the average prediction error. Root mean squared error (RMSE) sits just beside, as it squares errors before averaging to penalise sizeable misses more. When combining the two methods, and taking the gap between RMSE and MAE, we are able to measure whether a few incorrect decisions are inflating the error. Last in this group, R^2 reports the share of lap time variance.

As the binary classifier, Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) measure the quality of how well the model ranks an actual pit-soon lap above a non-pit lap. One of its biggest strengths is that it's robust to heavy class imbalance.

The primary operation product in this research, the strategy model, is where perhaps the ranking metrics matter most. The primary metric being the top 3 hit rate, it exhibits how often the lap on which the driver actually pitted on appears in the model's top candidates. The feasibility of it is justified, as a strategist works from a short watchlist, not just a simple binary alarm. Mean reciprocal rank (MRR) further refines it by rewarding the true lap for ranking higher.

4.9 Upgrade-Effectiveness Methodology

The evaluation of upgrade performance takes a different approach compared to Part A. Though attempted in this research, valuable conclusions could not have been drawn from a quantified data model assessing performance growth from documented physical upgrade packages in the given time window. The hypothesis described a potential way to measure whether an upgrade actually improved a car, shown in a way that is not fooled by the rest of the field developing at the same time.

A naïve before and after comparison attributes all of team's pace change to an upgrade. Moreover, as every team develops continuously over a season, a team that kept pace with general development would exhibit improvement in absolute terms. In the quantification attempt to estimate causal effects, difference-in-differences (DiD) method was used, by taking a double difference, taking the team's before to after change, minus the field's changes over the same races. For each valid green lap, a team's pace advantage is defined as the field's session-median lap time, minus the team's own median clean lap time. The main assessment would be the change in value, positive meaning faster than the rest of the field. A comparison structure was also constructed, the treated group being the upgrading team, and control group being other teams in the same observable time frame. This interval is three races before, and four races after an upgrade. Laps were grouped to a team per event, allowing them to be one stable unit.

Confidence intervals derived from bootstrap resampling in each window, 500 iterations overall. Effects are flagged significant at 95% when its interval excludes zero. The key setback of the

model is that it is only as good as the log it relies on. Teams often bring new packages in stages, which is detrimental for the accurate assessment of any single upgrade. Meaning that unfortunately overall, a quantified version of this assessment arrived at no feasible outcome.

In an interview conducted with motorsport engineering professional Steve Clark, he raised an important question. “How do you define an upgrade?”. Upgraded aerodynamic parts may lead to time gain on track, but that is merely scratching the surface. Teams may bring updated mechanical parts to improve overall aero-balance and driveability, which does not automatically imply a performance gain. As it is the nature of the sport, each team constantly strives to go faster with good reliability and consistency, but assessment cannot be simplified into “did we go any faster?” statements. Such limitations will be further discussed in later stages of this report.

4.10 Qualitative Methods and Constraint Analysis

This subsection is the methodology backbone of Part C of this research, exploring logistical and cost cap constraints with the intensive use of qualitative data. Why is that the case? The quantities that matter most for cost cap and logistics decisions, research and development allocation, wind tunnel and CFD restrictions, freight and personnel logistics, simply aren’t recoverable from public lap data. It would be a false approach to model something at the lap level, which lap data doesn’t attain. A deeply rooted, structured qualitative analysis is the most appropriate method for this section.

This approach combines FIA Financial Regulations figures and the included per-season cap value, set against the operational cost of balancing strategic budget allocation. Even though minor individual purchases are unavailable to the public, yearly spendings of a Formula 1 team are observable, as they are published within 9 months after the conclusion of each season. This allows a look into overall summations per department, for example catering or logistics, but they wouldn’t be precise enough for a quantified model.

Three semi-structured interviews further aid defining how qualitative information may contribute, and what constraint it brings with it. Catrin Mueller from DHL, the long-standing logistics partner of Formula 1 gave insight on freight constraints and the confidentiality boundary around cost data. Mueller’s perspective provides support for Formula 1 team decision-making under logistical and financial restraints. Steve Clark, motorsport engineering

professional helped in understanding what decisions are top priority during a race event, and what metrics teams will critically evaluate. Finally, Alex Powell, software engineer from Soft Pauer guided this research in key directions including model construction and the commercial-strategy tooling landscape.

These opportunities served two analytically distinct functions. Experts supplied constraints and operational realities that public data can only speculate on. Catrin Mueller's confirmation of per-kilogram freight costs and total seasonal logistics spendings are commercially confidential underlines why Part C had to remain qualitative. Not as a limitation, but rather a reflection on classification. Further logistical strain may be present, as each team chooses their own 'high priority' parts, and will allocate resources accordingly. Moreover, engineering testimony was used to check whether the model's initial signals and its boundaries were realistic. Steve Clark's testimony on what constitutes as a meaningful upgrade, and Alex Powell's framing of the quantities that matter most in strategy confirmed the features which the model leans on. Tyre degradation relative to a fresh baseline, gap to competitors, and stint age are included in variables teams actually monitor.

Equally important is what these interviews confirmed the model cannot do. That being safety car timing and the isolation of a single upgrade's were found to resist quantification. In general, it is well-defined that regulations set the boundary of what is permitted, public accounts give scale to the cost and development, while interviews supply the operational details and external validity check that no source can provide alone.

Overall, limitation still exist within this qualitative approach. Three experts' feedback though extremely valuable, is still a small sample. One is a commercial partner whose framing on certain topics is promotional, meaning the sustainability material is handled as out of scope. As with other conversational information, practitioner judgment and confidentiality bound what can be claimed. These interviews are used for support and validation, rather than standalone proof.

5. Results

5.1 Dataset Overview

To best evaluate the results of this research, it is imperative to understand the details of the concrete base, upon which every outcome is derived. As mentioned in section 4.3 the final dataset of the 2022-2025 regulation set, amounted to 101,297 laps over 25 unique circuits. During this four-year interval, the 2022-2024 split saw 74,605 laps in its structure, while the 2025 season set accounted for 26,692 laps. These laps covered all conditions encountered in Formula 1 race conditions. For this reason, it was necessary to filter out the green racing laps (uninterrupted competitive laps), as this allows the model to be trained best on normal race performance.

Tyre compounds included the whole range of soft, medium, hard, intermediate and wet tyres. Laps completed in wet weather conditions amounted to 3,982 laps, while this number evaluated on wet/intermediate laps accounted for 6,406. The difference comes from data attributes, as even though the track is not declared wet by the FIA and the 'is_wet_race' flag, drivers and teams still may choose to run wet tyres. As the data is split by season, the model is evaluated across varied circuits and conditions rather than on a simple sample. Several laps were discarded due to different factors leading to a reduction in lap number, from 101,297 to 87,032. This number now represents the true laps on which this research is based. In table 5.1a, all categories based on which laps were pruned from the clean set can be precisely extracted

Table 5.1a. *Lap-level data attrition (101,297 → 87,032).*

Pruning category	Laps removed	Cumulative remaining	% of total
Raw dataset (starting point)	-	101,297	100.0%
Track-status anomaly laps (SC / VSC / red flag)	-9,559	91,738	-9.4%
Pit-out laps (tyre_life < 2)	-3,433	88,305	-3.4%
Deleted laps (lap time invalidated by stewards)	-1,239	87,066	-1.2%
Both pit-out and deleted (overlap, already counted)	-34	87,032	-0.03%
Valid green-flag laps (model training set)		87,032	85.9%

On the other hand, as no dataset is truly perfect, occasional missing values were present and accounted for, to avoid breakage of core functionality. There was a total of 85,277 missing values across 32 columns, in other terms, about 0.98% of all cells. In table A.1 of the appendix is a summary of structural gaps, not random dropout, meaning they are filled with median or 0 imputation, or left as NaN to be handled by the model's missing-value mechanism, such as the histogram-based gradient boosting machine (HistGBM).

5.2 Model Performance

Lap time model

The key results of this subsection are best presented in their respective tables, as highlighted below.

Table 5.2a. *Lap-time prediction (HistGradientBoostingRegressor).*

Metric	Train (2022-2024)	Validation (2025)
MAE (s)	0.468	0.778
RMSE (s)	0.858	1.725
R ²	0.994	0.973
Rows	63,905	23,041

Showcasing the lap time estimator, a validation MAE of 0.778s was achieved. This three-quarters of a second can be considered a strong estimation, as Formula 1 lap times run from 70 to 110 seconds, putting the result well inside 1% of a lap. As the gradient boosting regressor in such an application has an inherent theoretical limit, improvements throughout this research that have gradually lowered this value from initial reports of 2.43s to 1.89s before the final value show clear positive results. Further enhancements would require the usage of a more complex model, but that was outside the realistic scope for the given research window.

On the other hand, training data MAE gave a resilient 0.468s, meaning that the model is more accurate on training data, than the unseen 2025 season, which is an ordinary occurrence. The key result to look for is the gap between these values, which in this case is moderate. Though validation error is larger, it is of the same order of magnitude, not double or triple. This close result is telling of a model successful in generalisation, instead of exhibiting memorisation.

To unpack the RMSE value of 1.725s, its need to clearly define what it is doing. From the low MAE, but 2.2 times higher RMSE, it is apparent that most laps are predicted tightly, but a minority are missed by a big margin, increasing the RMSE result. The explanation for that minority can be derived from safety car neutralisation (e.g. a VSC is over during the lap, and still being flagged green), traffic, driver error, or weather transitions, where lap time stops being a smooth function of pace. This result tells us where the model's uncertainty concentrates and will be analysed further in section 5.3.

As the last companion to MAE, it is observable in table 5.2a an R^2 of 0.973 was calculated. The reason for such a high value is partly accredited to the one-hot encodings of event, driver, and team that absorb a lot of structural variance. Different circuits have very different baseline lap times, that prompts the model to start with an easy share of variance already accounted for.

Tyre degradation model

The two primary results that need detailed interpretation from this model is the negative degradation rate of -0.044s/lap, and negative R^2 value of -4.54. Before jumping to quick conclusions, the underlying structure must be explained in detail. This model is not a single model, but rather 86 group specific polynomial regressions, one per circuit and compound combination, with a global fallback. Each of them produces a predicted lap time as a function of tyre age. The output feeding the strategy classifier is not the raw prediction, but rather a derived feature named: `deg_penalty_vs_fresh`. It defines the difference between the current predicted pace, and what performance a fresh tyre on the same circuit would achieve.

Table 5.2b. *Tyre degradation (polynomial Ridge, per circuit $\times$ compound).*

Metric	Train	Validation
MAE (s/lap)	1.113	1.750
RMSE (s/lap)	2.392	5.407
R^2	0.412	-4.54
Group models fitted	86	-
Global degradation rate (s/lap)	-0.044	-

Of the 86 group models, 28 (33%) exhibit a positive linear degradation rate, implying lap times worsening with tyre age. 58 of these models (67%) show a negative linear term. That demonstrates the tyre warm-up phase being more dominant over the observed stint length. On the other hand, 62 of the 86 groups (72%) have a positive quadratic coefficient, exhibiting an upwards curving trend, implying worsening lap times over a stint. The -0.044s/lap aggregates over both phases and should within itself is not to be considered as proof that tyre life improves over time. This shows that the tyres are cold at the start of the stint and grip improves as the rubber heats up, meaning that the first segment of the stint will always be faster than prior laps. It is this event that the negative linear term captures, while the quadratic term captures the subsequent tyre degradation cliff.

Tackling the alarming looking R^2 value of -4.54 , it thankfully has a straightforward explanation. R^2 is not the most optimal metric for a per-group model evaluated on a fully run Formula 1 season, as the baseline (mean lap times) massively vary throughout the year thanks to unique circuit layouts. The key takeaway should be that the MAE (1.75s/lap on validation) is an acceptable metric, as the model's goal is to provide relative features, and not absolute lap time predictions. The 1.75 s/lap error does not carry over 1.75 s error in the `deg_penalty_vs_fresh` feature, because the baseline shifts together with the prediction. Because the feature is the difference between two predictions from the same model (current tyre age versus fresh), any systematic offset in that model is present both times, and cancels in the subtraction, leaving the relative penalty largely unaffected.

It is important to note that the per group method requires sufficient historical data for each circuit and compound pair. Groups with fewer than 8 observations automatically fall back to the global model, which means unusual compounds at new circuits (e.g. a wet tyre at a circuit where it barely rains) are fitted on thin data.

Pit Classifier

The ROC-AUC metric measures how well the model ranks laps. The model in its current state will assign a higher probability for a true pit-soon lap, compared to no pit. The ROC-AUC of 0.740 is meaningful given class imbalance, as only about 8.7% of laps result in a pit stop.

Table 5.2c. *Pit classifier (logistic regression, L2). Target: will the driver pit within the next 3 laps?*

Metric	Train (2022-2024)	Validation (2025)
ROC-AUC	0.809	0.740
Average Precision	0.308	0.220
Recall	0.766	0.707
Precision	0.229	0.182
F1	0.353	0.289
Accuracy	0.727	0.698
Rows	61,247	21,970

The average precision (AP) value of the classifier is 0.220. With a severe class imbalance, it is expected that this value is on the lower end. It is approximately 2.5 times better than the base rate of 0.087 of a random classifier. That is the more meaningful comparison, not against 1.0.

Value Regressor

The value regressor achieves a validation MAE of 10.318s and an R^2 of 0.382. This model assesses the net time gain over six laps from pitting now versus staying out. This is a counterfactual quantity, only seen for one branch of each decision, so the absolute error reflects inherent outcome variance rather than model discrepancy. The regressor is tasked to rank candidate laps rather than forecast exact gains. Ranking quality is evaluated at the policy level.

Table 5.2d. *Value regressor (Ridge). Target: net time gain over the next 6 laps from pitting now versus staying out (counterfactual).*

Metric	Train	Validation
MAE (s)	5.883	10.318
RMSE (s)	14.320	21.974
R^2	0.533	0.382
Rows	1,310	438

Strategy Policy

The primary segment to highlight from this model, is the top 3 hit rate. The pooled value turned out to be 0.503. This algorithm ranks all candidate laps by their predicted pit values. The ‘top

3 hit rate' implies the lap the driver *actually* pitted on appears in the model's top 3 ranked laps in 50.3% of driver sessions. From an operational standpoint, that means a strategist would have the real pit lap flagged about half the time.

Table 5.2e. *Strategy policy (three-gate ranking system), 398 driver-sessions, 2025. (Dataset inspection revealed 3 races out of 24 to be incomplete for sufficient analysis)*

Metric	Train	Validation (pooled)	Live-scored (per-race mean)
Top 3 hit rate	0.587	0.503	0.347
Top 1 hit rate	0.394	0.362	-
MRR	0.517	0.465	0.331
Precision	0.253	0.173	-
Recall	0.328	0.262	-
Recommendation rate	3.8%	4.0%	-
Driver-sessions / groups	1,060	398	21 races

To continue, the difference between pooled versus live scoring metrics ought to be explained. The higher value is given by pooling 398 driver sessions from the 2025 validation season into one calculation. The 0.347 figure is extracted by averaging the hit rate per race across 2025. The dissimilarity is an outcome of the per-race value weighing each race equally and is more sensitive to races in which the model scores zero, hence the low number. As the pooled rate weights races by number of pit decisions, it serves a higher value.

A further evaluation point is the mean reciprocal rank (MRR). The rank of the actual pit lap is taken from the model's list, and then its reciprocal is taken ($1/\text{rank}$) per race session. The score of 0.465 means that on average the real pit lap is ranked approximately second ($1/2=0.5$). It can be considered a softer version of the top 3 rate, providing partial credit for near misses.

Lastly presented in this subsection and results of this model, is a comparison between a gradient boosting machine (GBM) and logistic regression, and their values of 0.297 and 0.503 respectively. When testing GBM classifier instead of logistic regression, the rate drops to 40% of its prior precision. This outcome is interpreted in section 6.2.

5.3 Model Error Analysis

Moving on from aggregate accuracy to the distribution of error, this subsection delves into finding out where the model struggles with precision.

Table 5.3a. *Per-race top 3 hit rate, 2025 season (calendar order excluding the 3 races with insufficient datasets).*

Race	Race type	Top 3 hit rate
Australian GP	SC/VSC	0.000
Chinese GP	Clean	0.053
Japanese GP	Clean	0.200
Bahrain GP	SC/VSC	0.278
Saudi Arabian GP	SC/VSC	0.000
Miami GP	SC/VSC	0.091
Emilia Romagna GP	SC/VSC	0.000
Monaco GP	SC/VSC	1.000
Spanish GP	SC/VSC	0.421
Canadian GP	SC/VSC	1.000
Austrian GP	SC/VSC	0.118
British GP	SC/VSC	0.688
Belgian GP	SC/VSC	1.000
Hungarian GP	Clean	0.250
Dutch GP	SC/VSC	0.200
Italian GP	Clean	0.188
Azerbaijan GP	SC/VSC	0.500
Singapore GP	Clean	0.750
United States GP	SC/VSC	0.158
Mexico City GP	SC/VSC	0.167
Sao Paulo GP	SC/VSC	0.222

The intuitive hypothesis is that safety car races wreck the policy with chaos, unscheduled stops, and unpredictable timing. On the other hand, those session with a SC/VSC interruption led to similarly accurate top 3 hit rate results, as clean racing laps (0.200 clean vs 0.211 SC/VSC). Observable also is the fact that some races score above 0.50 while others score the 0.19 metric.

The model's performance shows that as strategies become more complex with a higher number and more spread-out pit stops, that is where prediction metrics begin to struggle.

Table 5.3b. *Summary by race type.*

	Clean (5 races)	SC/VSC (16 races)
Median top 3 hit rate	0.200	0.211
Mean top 3 hit rate	0.288	0.365
Standard deviation	0.268	0.366

Analysing the Monaco/Canada/Belgium/Azerbaijan/Silverstone versus Singapore split is key evidence. Observable in Figure 5.3 below, five of those six races are SC/VSC races and scored at or above 50%, and correlate with good scores, as it tends to compress the drivers' stops into a similar window. The natural hypothesis is rejected by the data, thanks to the near identical medians, meaning the real driver is pit-window concentration. Furthermore, it must be noted that safety car races aren't systematically easier, they are systematically more spread out (hit rate variability), which can be observed by the mean, as it is pulled around by that spread (0.288 clean, 0.365 SC/VSC).

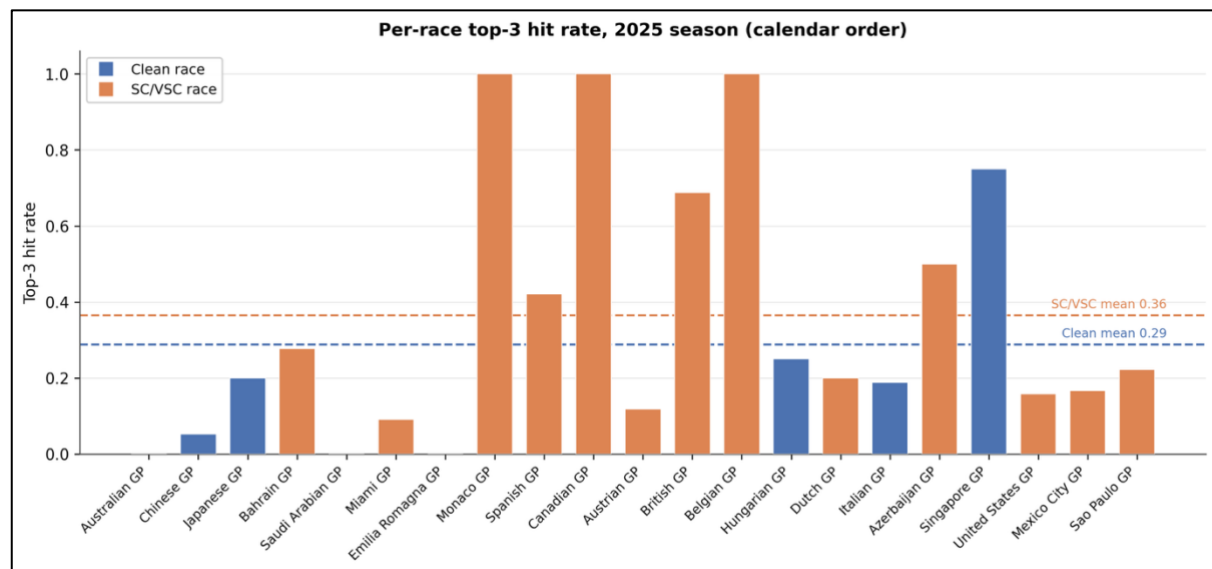


Figure 5.3. *Per-race top 3 hit rate, 2025 season*

5.4 Feature Importance

The most important and actionable feature that the model leans on is `deg_penalty_vs_fresh`. This feature is the core baseline for the assessing the drawbacks or advantages of pit timing and encompasses real world pit decisions. This result is of great importance, as it is evidence on how the model learned something sensible.

Moreover, there was an attempt to include opponent threats features, such as undercut/overcut danger, but it was deemed unhelpful in its contributions, as the result was collinear with gap and tyre life features already present. The -0.5 percentage points on top 3 hit rates only indicate noise, not anything useful for prediction. Opponent features are derived from `gap_behind_seconds` and `tyre_life`, already found in the model, meaning no new information transportation that the linear classifier wasn't already using. The undercut/overcut threat is real but is captured by simpler features.

Table 5.4a. *Feature groups by category.*

Group	Features	Examples	Role
Pace / lap time	6	<code>lap_time_lag_1_driver</code> , <code>lap_time_rolling_mean_3_prior</code> , <code>recent_pace_change_seconds</code>	Core pace signal
Tyre state	5	<code>tyre_life</code> , <code>tyre_life_delta</code> , <code>compound_*</code> (one-hot)	Age and compound encoding
Degradation signals	7	<code>deg_penalty_vs_fresh</code> , <code>deg_linear_rate</code> , <code>deg_pace_loss_5laps</code> , <code>deg_pace_loss_10laps</code>	Tyre condition vs fresh baseline
Race context	6	<code>lap_number</code> , <code>lap_number_progress</code> , <code>stint</code> , <code>stint_lap_index</code> , <code>tyre_life_to_lap_ratio</code>	Position in race / stint
Track position / gap	6	<code>position</code> , <code>position_percentile</code> , <code>gap_ahead_seconds</code> , <code>gap_behind_seconds</code> , <code>gap_to_leader_seconds</code>	Relative field position
Weather	9	<code>air_temp_celsius</code> , <code>track_temp_celsius</code> , <code>humidity_pct</code> , <code>wind_speed_kph</code> , <code>is_wet_race</code>	Ambient and surface conditions
Season-lagged priors	5	<code>lap_time_expanding_mean_prior</code> , <code>lap_time_delta_driver</code>	Cross-stint and cross-race baseline
Sector shares	3	<code>sector_1_share</code> , <code>sector_2_share</code> , <code>sector_3_share</code>	Circuit-section pace distribution

Group	Features	Examples	Role
Traffic	3	traffic_gap_min_seconds, cars_ahead, cars_behind	Close-racing context
Opponent-threat	6	opponent_tyre_life_delta_behind, opponent_tyre_life_delta_ahead, field_pit_stop_share	Undercut / overcut threat
Categorical (encoded)	4	event_name, driver, team, compound_clean	Circuit / driver / team identity

Table 5.4b. *Key findings on feature importance.*

Feature	Finding
deg_penalty_vs_fresh	Most actionable feature. Quantifies the performance cost of staying out versus pitting for a fresh tyre of the same compound. Adding the 7 degradation features lifted policy top 3 hit rate by +3.7 pp.
Lap-time lags (lag_1, lag_2, rolling_mean_3)	High weight to capture the recent pace trend within a stint.
tyre_life	Significant as stint age is a core signal independent of the degradation features.
gap_behind_seconds	Captures undercut threat from a chasing car.
Opponent-threat features (6 total)	Minimal incremental impact as -0.5 pp top 3 hit rate (within noise). Already encoded by existing gap_behind_seconds and tyre_life features; collinear with these in the linear classifier. Kept in but contributes negligible marginal signal.
Weather features (9 total)	Moderate as it's primarily useful for wet/intermediate compound sessions.
is_wet_race	Binary; reduces model confidence in degradation predictions for abnormal conditions.

Table 5.4c. *Degradation feature detail.*

Feature	What it captures
deg_penalty_vs_fresh	How much slower the current tyres are versus a fresh set of the same compound at this circuit.
deg_linear_rate	Circuit-compound fitted linear degradation rate (s/lap).
deg_quadratic_rate	Quadratic acceleration term captures the late-stint cliff.
deg_pace_loss_5laps	Projected additional pace loss from staying out 5 more laps.
deg_pace_loss_10laps	Projected additional pace loss over 10 more laps.

Feature	What it captures
deg_stints_on_compound	How many stints the driver has already run on this compound.
deg_relative_age	Tyre life as a fraction of the typical stint length for this compound at this circuit.

5.5 Upgrade-Effectiveness

A sum of 12 upgrades were tested, 4 were marked significant at 95% (1 positive, 3 negative), while most showing no statistically significant effect. On the other hand, the biggest true positive (team gained pace relative to the field around the upgrade) result of +0.863 with a CI of [+0.246, +1.486] is a standout case, showing good promise in model performance. The reason that statement may be declared, is because as mentioned in other sections of this report, the feedback from industry professionals was clear that defining an upgrade package and evaluating it purely from an observable on-track approach, is extremely difficult. Circuit specific corners, weather conditions, and a plethora of other elements affect car performance. To contrast this point of view, the McLaren finding is very positive. That upgrade package and overhaul of race car ejected the team from the lower midfield, finishing around 15th/16th to podium contention. It also marked the first step in their development towards an eventual 2024 constructor's championship. The fact that the model was able to evaluate that package positively, is a great result.

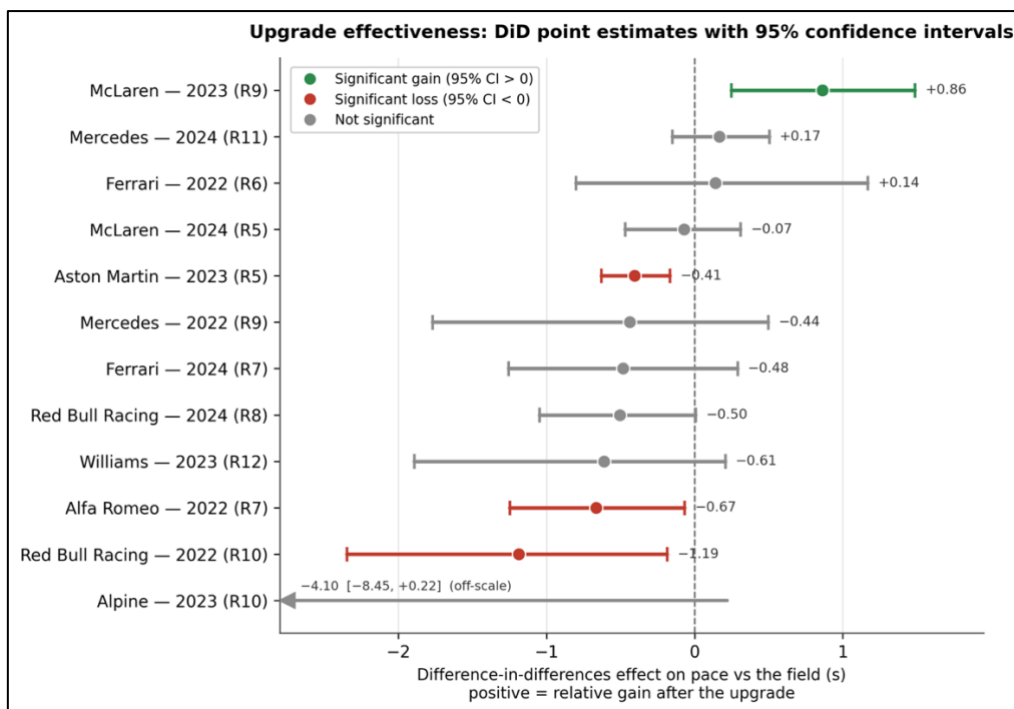


Figure 5.5. Upgrade effectiveness: DiD point estimates with 95% confidence intervals

Alternatively, the model also resulted in significant negative effects, such as the packages for Aston Martin 2023 R5, Alfa Romeo 2022 R7, Red Bull 2022 R10. Even though there are examples of upgrades doing more harm than good for a team, these instances are not such occurrences. Furthermore, several estimates have very wide CIs, signalling low precision for those cases.

Table 5.5a. *Summary of upgrade-effectiveness results.*

	Count
Upgrades tested	12
Significant at 95% CI	4
Number of which clearly positive	1 (McLaren 2023, Austria)
Number of which significant negative	3 (Aston Martin 2023, Alfa Romeo 2022, Red Bull 2022)
Not statistically significant	8
Positive point estimate (not sig.)	2
Negative point estimate (not sig.)	6

5.6 Cost Cap and Logistics Findings

To bring a clear perspective on such findings, it must first be established what public accounts do and do not reveal. Unfortunately, public data is not decomposed into department-level spendings and as a result, this section will analyse only the summarised spendings seen in Mercedes-Benz Grand Prix Limited's 2024 financial data.

To answer spending focus backed by actual data, there are headline numbers to analyse. Revenue grew 16% to £636 million (\$860 million) for the year ended 31 December 2024, which is up from £546.5 million in 2023, with sponsorship and licensing accounting for 60% of that income. Capped racing operation of roughly \$135-160 million means that the regulated, performance relevant spending is well under a quarter of the company's total revenue (Mercedes-Benz Grand Prix Limited, 2025)

Earnings further climbed from £113.8 million to £160.6 million. With just over £40 million in tax, net profit was £120.3 million. 2023 filings also documented staff numbers rising to 1,289 and staffing costs reflecting £127 million, with capitalised race car development rising to £52.2 million. The 2024 documents carry equivalent figures. This salary sum does include all 1,289

members, but not the whole £127 million contributes towards the cost cap (Mercedes-Benz Grand Prix Limited, 2025).

One caveat in this analysis, is the unrepresentative showing which Mercedes' financial data provides. The team is an engine manufacturer supplying power unit components to other F1 teams (McLaren, Williams, pre-2026 Aston Martin, and post-2026 Alpine), and receives heavy sponsorship and prize money, placing them in the top end, and overstating what a customer midfield team looks like (Mercedes-Benz Grand Prix Limited, 2025).

Extractable assets include capitalised car development, total staff cost and head count, alongside revenue mix, but other numbers such as capped budget splits across aerodynamics, logistics, manufacturing, and simulation departments remain unknown. Investigating the FIA Financial Regulations, which defines the cap, we can in the least arrive at approximations by understanding what is not included.

Thanks to Catrin Mueller from DHL, the following sections will provide information on freight and transportation data relevant to everyday functioning of a team.

Freight dimensions, just like a Formula 1 car are engineered with hard, physical constraints in mind. Mueller highlighted how it gets precise to the point, where equipment is packaged to fit specific aircraft and move multimodally across seas, air, and road, clearly driven by cost savings.

Another crucial point Mueller made in upgrade timing, is the logistics lead times, and how they shape it. Teams rely on DHL's lead time, packing, and routing guidance to decide which parts to bring and when, further emphasizing how development decisions are made under logistical constraints. Granular costs which may provide enough backing for a quantitative model is unobtainable, encouraging the inclusion of qualitative findings and rough approximations. This answer also shows that bottom-up logistics costs are closed to outside research.

This takeaway explains how it's directional, not numeric. The key contest that consumes a team's allocation is development, with logistics being a real operational constraint that partly sits inside the cap, shaping when development reaches the track, while in-race strategy is near-costless.

Table 5.6a. *Treatment of major cost categories under the Formula 1 cost cap.*

Cap base: \$135m for 2024-2025 on a 21-race calendar, plus \$1.8m per additional race (about \$5.4m for a 24-race season) and an inflation indexation, putting the practical limit nearer

\$150-160m. The power unit is capped separately. From 2026 the cap rises to roughly \$215m and folds in several previously excluded items (FIA, 2024; Barretto 2020).

Cost category	Treatment	Notes
Car and development		
Chassis, aerodynamic surfaces, and all car components	Included	Steering wheel to the tyres, the core of relevant costs.
Research, design, and development	Included	The principal performance battleground inside the cap.
Wind tunnel and CFD testing costs	Included	Spending counts toward the cap, usage is also separately restricted by the Aerodynamic Testing Restrictions.
Spare parts	Included	-
Fuel and lubricants	Included	-
Power unit / engine	Excluded	Outside the team cap; governed by a separate power-unit cost cap (\$95m through 2025).
Heritage car programmes	Excluded	-
Personnel		
Operational staff salaries (engineers, mechanics, aerodynamicists, designers, factory and operations)	Included	Most of the workforce sits inside the cap, and salaries are usually its single largest component.
Three highest-paid staff	Excluded	Typically, the team principal and chief technical officer.
Driver salaries and retainers	Excluded	Not ‘staff’ in the accounting sense; excluded regardless.
Legal, HR, and finance functions	Excluded	Back-office departments are left out.
Marketing (staff and expenditure)	Excluded	-
Sustainability initiative costs	Excluded	-
Operations and logistics		
Garage and operational equipment	Included	-
Race operations	Included	-
Freight / transport of car equipment	Included	Treated as a relevant cost in most readings.
Team personnel race travel	Excluded	Distinct from equipment freight, travel costs are an exclusion.

Cost category	Treatment	Notes
Commercial and overhead		
Property costs	Excluded	-
Entry and licence fees	Excluded	-

6. Discussion

6.1 Answer to Research Question

To what extent can data-driven methods quantify the impact of Formula 1 car upgrades and race strategies on on-track performance under cost cap and logistics constraints?

It was this question that motivated the pursuit of this entire research. The results show, it is possible to a meaningful but bounded extent. Lap performance is highly predictable, especially if trialled through complex models. Strategy timing is less successful and exhibits a real ceiling. The simpler the strategy in a race event, the better the results regarding pit stop windows. Upgrade effectiveness is quantifiable to a certain extent but is generally noisy. A realistic approach must be considered for this quantification, as feedback from industry experts underlines the difficulty of measuring performance of a single upgrade package. Cost cap restrictions and its effect on logistics and upgrade planning is best evaluated qualitatively.

Understanding the major areas of spending in a competitive season tells exponentially more than attempting to quantify such a constraint. Using this method, better assessments can be made to determine which regions were more resource intensive compared to others, providing a clearer view of where teams were forced to restrict development.

6.2 Interpretation of Model Results

The overarching system consists of four independent components, including a lap time baseline, a degradation model, a pit classifier, and a strategy policy, all evaluated on the 2025 season. The most significant finding is that a regularised linear classifier outperforms a gradient-boosted alternative at the policy level, and that the margin between them is traceable to a single structural property of the data, which is the regulation era consistency.

Lap time model

A validation MAE of 0.78s compared to on average 90-second laps represents a relative error under 1%. This outcome confirms that the feature set made up of tyre compound, tyre age, circuit identity, cumulative race time, and weather captures the dominant sources of lap time variance. There exists a sizeable gap between train MAE and validation MAE, which is to be expected, since training was done on the 2022-2024 dataset, and was evaluated on the 2025 season with new circuit layouts and car development. This gap ought not to be considered as evidence of overfitting, but rather of genuine distribution shift across seasons within the same regulation era.

Logistic regression versus gradient boosting

The choice of classifier turned out to attain a disproportionate effect on policy performance. Logistic regression with L2 regularisation achieved a top 3 hit rate of 0.503, on the 2025 validation season, whereas a gradient boosted machine with calibrated probabilities achieved only a score of 0.297, meaning a 21-percentage point deficit. This phenomenon is explained by a GBM's function of learning to associate specific event-compound combinations with pit decisions, capturing season-specific strategy patterns that do not transfer to 2025.

On the other hand, a logistic regression's linear decision boundary is constrained to use only the continuous features, such as tyre age, pace delta, gap to competitors. It is a generalisation asset that prevents the model from memorising patterns that will not occur, that causes the discrepancy between the two methods. It is a principle that costs in-sample performance for robustness.

Degradation model

As the metrics from this model are at a first glance surprising, it must be placed into the correct context. The validation MAE and negative global degradation rate reflect the tyre warm-up phase dominating at low stint ages. 62 out of 86 groups show a positive quadratic coefficient, implying that pace does worsen over lengthy stints, which is as expected. More importantly, the model is not evaluated in isolation. Its output is assessed by the strategy classifier as seven derived features. When those features were added, the top 3 hit rate improved by an observed

3.7 percentage points, providing quite direct evidence that polynomial fits are capturing actionable tyre state signals.

Safety car/virtual safety car findings

Natural strategy instincts interpret the hypothesis that SC/VSC interruptions ought to degrade model performance, as a result of compressing the field and disrupting normal pit timing. The analysis based on the 2025 season does not support this. Safety car presence alone does not explain model performance. Both good and bad results are present during safety car races from the model. The differences happen because of the distinguishing factors of each individual SC event. When a safety car creates a single clear pit window, the features correctly flag it. Alternatively, when a safety car triggers simultaneous chaos, during which most of the field pits reactively within two or three laps, the decision gets decided by opponent behaviour.

This decision is driven by opponent behaviour rather than the driver's own tyre state, and the lap feature inputs carry little signal. Realistically, no lap feature model can anticipate a pit decision as a response to what other drivers do in real time.

As the Monte Carlo simulator inherently assumes green flag conditions for the race after an interruption, it understates uncertainty in strategy scenarios, because SC probability is not modelled forward. Opponents may pit, and change pace when they feel threatened, leaving the simulated position outcomes as approximations rather than forecasts.

Concluding this section, it is notable that the results suggest that a linear, regularised architecture is the more appropriate choice for cross-season generalisation of strategy, given the regulation-era structure. Therefore, the primary ceiling on policy performance is not model capacity but an inherent unpredictability of safety car timing.

6.3 Interpretation of Upgrade Effectiveness

As reported before, 4 out of 12 results were deemed significant. This segment discusses why most are null and what the exceptions mean.

Quantifying performance before and after documented upgrade packages led to the conclusion that most upgrades show no significant relative effect. In a continuously developing field, the expected minimum is to keep pace with the grid. As a result, a clearly measurable relative jump

is rare. Here is where McLaren and their Austria upgrade come in and exhibit such an occasion. As detailed before, this change came with such a massive performance improvement, especially compared to McLaren's then-rivals, the midfield teams, it can clearly be shown by data.

The best way to evaluate the partial quantifying attempt, is to confront the method against reliable literature, putting DiD against prior work on upgrade impact, and race reports. Research metrics can only provide an estimated range of effectiveness, not a precise value. As assessment solely through data is naïve, a combination of quantitative and qualitative results ought to be evaluated together. When taking a look at 'negative' upgrade effects, it doesn't mean the team lost performance with the upgrade, it may also mean they lost ground compared to the field, who perhaps also brought one. Simultaneous race car improvements further contribute to difficult single-team assessment.

6.4 Business Engineering Implications

To best assess the business engineering implications of this research project, the different key parts being A, B, and C must finally become one entity. It is this section where this research's results are finally presented as they are meant to be, integrated with each other. It has been observed that performance is predictable, strategy timing partially so, upgrades are situational and quantifying them is challenging, and in-race cost is trivial to the cost cap, leaving biggest spending margins on upgrades, travels.

Race day strategy and car development operate all from the same cost cap but take up its allocation in fundamentally different proportions. In-race strategy is operationally near-costless, meaning it does not meaningfully constrain strategy decisions, and is not a key consideration for teams watching their spendings. It is the in-season car development eating away at the budget. Part B has shown that measurable and statistically significant upgrade gains are rare, but it under no circumstance makes upgrades meaningless. If a team doesn't innovate and improve, they get left behind in no time. Part C establishes how real expenditure comes in the form of R&D, wind tunnel time, and logistics, not in race operations. Therefore, it can be declared that the binding constraint on a team's competitiveness is defined by how effective they are in converting the capped development budget and time into pace. Strategy quality and development quality are distinct problems, under the same budget, but the latter suffering more.

Understanding the importance between development and strategy, the practical consequence for allocation is clear. Harder to reverse and more expensive decisions ought to be prioritised such as upgrade packages, logistical infrastructure, and personnel over cheap and repeatable items. The time when race events themselves are dangerous to higher costs, is when mechanical parts get damaged during the race. Such mishaps may have bigger implications than first expected. If a driver damages body work, like a front wing, a spare will be used, but that implies a new one should be manufactured in its place. On occasions when a team brings new parts, and those get damaged, it is often the case that to manufacture new parts of the new spec, all focus and resources must focus on it. New parts often only have one spare, which if used up, quickly changes priority and lead times for the team.

Formula 1 under the cost cap is an allocation constrained problem, having to consider a fixed budget, competing investment demands, and marginal performance returns on each dollar spent. The findings here suggest that teams don't just have to consider costs present from R&D, manufacturing, and testing, but must also closely collaborate with DHL and other logistics partners to get new parts where they need to be as soon as possible. Small mishaps during triple headers can have large implications and exhibit how the smallest miscalculations can cause large setbacks in transportation, plaguing operation down the line.

6.5 Limitations

An attempt has been made to quantify upgrade effectiveness, and though successful on occasion, it is not feasible to rely on just the data, as a new aerodynamic package has an observable performance increase window of only a few races, or before a new one is introduced. The most optimal way would be to compare the difference before and after a new upgrade on the same track, but a unique track is only raced once per season. After a year, a single upgrade cannot be evaluated, as packages arrive several times during that time frame.

The trade-off this study has is also its opportunity for reproducibility. As it runs on public data, it can support future research, but is inherently bound to information that is knowable. No internal telemetry or cost breakdowns, no team-side precise truth, implying that absolute benchmarks and granular allocation stay out of reach.

Learned relationships in machine learning are challenged every regulation cycle. Without retraining, these relationships will not carry over across a regulation boundary. It is one thing

to build a model customised for new rules, but the drip-feeding of data further disallows quick retraining. After a year, the only available dataset consists of around 21 races, all on unique tracks and weather conditions. All these constraints are limits to external validity.

Furthermore, quantitative weaknesses also plague some of the individual models. The counterfactual value regressor's 10.318s validation MAE is the noisiest component. The target is meant to analyse a six lap trade-off between pitting or staying out, including a pit lane loss of 22-25s, meaning the error reflects inherent outcome variance rather than a mis-specified model.

The evaluation of the degradation model returned a negative R^2 , but the metric is inappropriate for the model's actual role, as it produces relative features rather than absolute lap time forecasts. The reason for its implementation is defined by its measurable contribution to lifting the policy.

Lastly, as the key component of this research, we must analyse its limitations. The top 3 accuracy of the strategy policy is found to be capped below 1, which is unfortunately a structural binding constraint and not fixable with better features. The main explanations come from interview evidence stating that safety car timing is close to random. Such interruptions do absolutely affect strategy decisions, but as their likelihoods are directly connected to driver-error or mechanical part failures, or many other events, they cannot be quantified.

6.6 Future Research Opportunities

To further enhance this research topic, several enhancement routes can be taken, especially with a bigger development window. A v2 strategic opponent model has the potential greatly increase pit stop timing prediction precision, as current over/undercut events are underestimated. Furthermore, a potential OpenF1 live integration may be explored, but usage of a live prediction model has legislation restrictions, and as a result its use must be fully assessed.

As the current 2026 season has only concluded 7 races at the time of this publication, there are more optimal opportunities further on to test the model and correct it to better suit the new era of regulations. The more time passes, the more concrete results one can achieve. Lastly, there may be changes made to budget publications, implying a better look and assessment opportunities for cost cap spendings.

This research aimed to perform well solely on public information and used internal interview outcomes only to reassure and improve accuracy. It is also advised to treat regulation eras as distribution boundaries, rather than pooling across periods.

7. Conclusion

7.1 Main Findings

Data-driven methods quantify performance and strategy modelling, upgrade effectiveness, cost cap and logistics constraints to a meaningful but bounded extent, with its value emerging when interpreted together.

Part A of this paper found answers for sub-questions one and two. Lap performance is highly predictable based on stable conditions and minimum session interruption. The top 3 hit rate strategy policy is constrained by a theoretical limit and would benefit from a more accurate threat-analysis feature. Logistic regression as the spine of this model beat the GBM on cross-season transfer.

Part B attempted the quantification of mechanical upgrades, proving 4 out of 12 significant at 95%, while most were null. McLaren's B-spec car upgrade stood out as the most promising example, exhibiting partial reliability.

Part C deemed the fourth sub-question's question about in-race strategy events near costless, and considered development as the binding constraint on competitiveness, and underlining it as the most critical objective of resource allocation.

7.2 Research Contribution

Inspiration for pursuing this research stemmed from the motivation of discovering how a Formula 1 team functions as a whole. It is evident from this paper, that to maximise performance, a team must run as a well-oiled machine. Prior work treats lap time prediction, strategy and cost cap in isolation.

By running these elements on one shared lap level baseline this paper proved how important each part is to the other. One cannot succeed without information from another. The

methodological finding followed a structure applicable for a whole regulation-era by being a simpler regularised model. This robust decision exhibits a transferable lesson beyond F1.

Moreover, the public-data-only constraint contributes to reproducible results without the use of privileged team data, even if it was accessible to the researcher. The results can benefit any smaller motorsport team who is looking to join a series with cost constraints, as it is the main element that will then have constraining effects on development, upgrades, and transportation. It will further benefit student teams around the world in series such as Formula Student, as the integrated framework showcases how every aspect explored is connected, and can provide better understanding right from the get-go of a student project. Lastly, the author hopes that the integrated single-framework methodology provides a lesson that transfers beyond Formula 1.

7.3 Practical Implications

Evident from this research, the allocation argument has a clear answer. Teams should prioritise expensive and more development-heavy decisions regarding upgrade packages, logistics, and personnel, over cheaper, repeatable costs. This does not imply that they should be disregarded, it only signals that strategy decisions and actual on-track running eat away at the allocation in a small manner. From a cost perspective, it is not as resource intensive as development, logistics, and team management.

7.4 Final Reflection

Attempting to fully quantify all elements investigated in this domain is a naïve approach. Even though Formula 1 is an engineering sport, it includes great amounts of the human element. The drivers, strategists, mechanics, they all contribute towards performance. Alongside these exist environmental factors such as temperature, wind, humidity possibly inducing fluctuation in data and evaluation. These are examples why the strategy model in this research is meant as a strategy aid, not a full replacement for a strategist. Interviews with industry professionals further underline this element and reveal how quantitative and qualitative evidence must be read together to provide a full picture. The research outcome may be bounded in its application as it aims to function on unseen data, it provides a sidekick for when difficult decisions may have to be made.

Future Perspective

This study has the ground-effect era of Formula 1 as its spine. Methodologies explored are fitting that 2022-2025 era by design, leaving space for further modifications fit for the 2026 regulations. As the current regulations evolve and grow in sample size, the same pipeline's predictive power can be strengthened. However, to create a most precise model, modifications will surely have to be made to fit the attributes of the new changes and its mechanical overhaul.

The trajectory of this project, acting as a foundation for live decision support rather than a post-race strategy analysis tool allow for more complex use, potentially outside of Formula 1 with minor alterations. Richer telemetry and longer samples will sharpen the strategy policy beyond its current bounded ceiling. Furthermore, evolving cost cap transparency may arrive in the future and may push the field from estimation toward measurement over time.

References

- Aramco F1 Team, A. M. (2024, November 14). *Aston Martin F1 Team*. Astonmartinfl.com.
<https://www.astonmartinfl.com/en-GB/news/feature/strategically-speaking-f1-strategy-explained-with-tom-mccullough>
- Barretto, L. (2020, May 27). *The 2021 F1 cost cap explained – what has changed, and why?* | Formula 1®. Wwww.formula1.com. <https://www.formula1.com/en/latest/article/the-2021-f1-cost-cap-explained-what-has-changed-and-why.5O1Te8udKLmkU14PyVZtUJ>
- Barretto, L. (2024, June 6). *FIA unveils Formula 1 regulations for 2026 and beyond*. Formula 1® - the Official F1® Website. <https://www.formula1.com/en/latest/article/fia-unveils-formula-1-regulations-for-2026-and-beyond-featuring-more-agile.75qJiYOHXgeJqsVQtDr2UB>
- Brusik, F. (2024). *PREDICTING LAP TIMES IN A FORMULA 1 RACE USING DEEP LEARNING ALGORITHMS A COMPARISON OF UNIVARIATE AND MULTIVARIATE TIME SERIES MODELS*. <https://arno.uvt.nl/show.cgi?fid=180319>
- Cappello, C., & Hoegh, A. (2025). *A State-Space Approach to Modeling Tire Degradation in Formula 1 Racing*. ArXiv.org. <https://arxiv.org/abs/2512.00640>
- Coleman, M. (2025, October 28). How Formula 1's cost cap works: Rules, penalties and past breaches explained. *The New York Times*.
<https://www.nytimes.com/athletic/6757239/2025/10/28/formula-1-cost-cap-explained/>
- Csermák-Scharbert, B. (2026). *DeltaTrace*. Delta-Trace.com. <https://delta-trace.com>
- FIA. (2023). *Decision documents*. Federation Internationale de L'Automobile.
<https://www.fia.com/documents/championships/fia-formula-one-world-championship-14>

FIA. (2024). *Formula 1 Financial Regulations*.

https://www.fia.com/sites/default/files/fia_formula_1_financial_regulations_-_issue_21_-_2024-07-31_0.pdf

Filip Cleeren. (2021, April 22). *Shovlin: Cost cap makes big F1 accidents “quite a concern.”*

Autosport.com; Autosport. <https://www.autosport.com/f1/news/shovlin-cost-cap-makes-big-accidents-quite-a-concern/6397711/>

Hatton, G. (2024, July 4). *F1 strategy: How Formula 1 teams determine the fastest race*

strategy. Raceteq.com. <https://www.raceteq.com/articles/2024/07/how-formula-1-teams-determine-the-fastest-race-strategy>

Heilmeier, A., Graf, M., Betz, J., & Lienkamp, M. (2020). Application of Monte Carlo

Methods to Consider Probabilistic Effects in a Race Simulation for Circuit

Motorsport. *Applied Sciences*, 10(12), 4229. <https://doi.org/10.3390/app10124229>

Jiang, Y., Ning, K., Pan, Z., Shen, X., Ni, J., Yu, W., Schneider, A., Chen, H., Nevmyvaka,

Y., & Song, D. (2025). *Multi-modal Time Series Analysis: A Tutorial and Survey*.

ArXiv.org. <https://arxiv.org/abs/2503.13709>

Mercedes-AMG F1 Team. (n.d.). *INSIGHT: The Trackside Engineers*. Mercedes-AMG

PETRONAS F1. <https://www.mercedesamgf1.com/news/insight-the-trackside-engineers>

Mercedes-Benz Grand Prix Limited. (2025, June 19). *MERCEDES-BENZ GRAND PRIX*

LTD filing history - Find and update company information - GOV.UK. Find-And-

Update.company-Information.service.gov.uk. [https://find-and-update.company-](https://find-and-update.company-information.service.gov.uk/company/00787446/filing-history)

[information.service.gov.uk/company/00787446/filing-history](https://find-and-update.company-information.service.gov.uk/company/00787446/filing-history)

Mitchell, S. (2022, November 18). *How Data Works in Formula 1*. Racecar Engineering.

<https://www.racecar-engineering.com/articles/how-data-works-in-formula-1/>

OpenF1 API | Real-time and historical Formula 1 data. (n.d.). Openf1.org.

<https://openf1.org/>

Pasz, S. (2025). *Determinants of Lap Times in Formula 1: An Econometric Analysis of the 2019-2025 Seasons Bachelor's Thesis.*

<https://dspace.cuni.cz/bitstream/handle/20.500.11956/202790/130438956.pdf?sequence=1>

Schaefer, P. (2026). FastF1 (Version 3.8.3) [Python package]. PyPI

<https://pypi.org/project/fastf1/>. GitHub. <https://github.com/theOehrly/Fast-F1>

Zhou, M., Huang, S., Tu, W., & Wang, D. (2024). Machine learning-based causal inference for evaluating intervention in travel behaviour research: A difference-in-differences framework. *Travel Behaviour and Society*, 37, 100852.

<https://doi.org/10.1016/j.tbs.2024.100852>

Appendices

Appendix A - Tables

Table A.1. *Missing values by column*

Column	Missing values	% of rows	Cause	Handling
gap_balance_seconds	12,569	12.4%	Only defined for mid-field positions	HistGBM handles natively
gap_behind_seconds	7,077	7.0%	Last car on track has no car behind	NaN
gap_ahead_seconds	7,077	7.0%	Leader has no car ahead	NaN
tyre_life_delta	6,226	6.1%	First lap of stint, no previous lap to diff	NaN
tyre_life_lag_1	6,226	6.1%	First lap of stint	NaN
recent_pace_change_seconds	5,682	5.6%	Requires ≥ 2 prior laps in stint	NaN
lap_time_lag_2_driver	4,945	4.9%	Fewer than 2 prior laps in session	NaN
lap_time_delta_driver	4,051	4.0%	First lap of session	NaN
lap_time_lag_1_driver	3,184	3.1%	First lap of session	NaN
sector_1_seconds / sector_1_share	2,101	2.1%	Sector splits not always telemetered	NaN
rolling_minus_expanding_pace_seconds	2,041	2.0%	Requires rolling window of ≥ 3 laps	NaN
lap_time_rolling_mean_3_prior	2,041	2.0%	Requires 3 prior laps	NaN
gained_positions_last_lap	1,931	1.9%	First lap: no prior position	NaN
lost_positions_last_lap	1,931	1.9%	First lap, no prior position	NaN
position_delta_driver	1,931	1.9%	First lap	NaN
lap_time_expanding_mean_prior	1,833	1.8%	First lap of session	NaN
traffic_gap_min_seconds	1,585	1.6%	No nearby cars in window	NaN
cumulative_race_time_seconds	1,546	1.5%	Missing telemetry (null lap time)	Row excluded from regression
gap_to_leader_seconds	1,546	1.5%	Missing telemetry	Row excluded from regression
lap_time_seconds	1,546	1.5%	Missing telemetry	Row excluded from regression
tyre_life_to_lap_ratio	1,081	1.1%	Null tyre_life	NaN

Column	Missing values	% of rows	Cause	Handling
tyre_life	1,081	1.1%	Tyre data unavailable for lap	NaN
compound	872	0.9%	Compound not reported by API	Encoded as unknown category
leader_cumulative_race_time_seconds	683	0.7%	Leader telemetry gap	NaN
stint_lap_index	446	0.4%	Stint boundary ambiguity	NaN
stint	446	0.4%	Stint boundary ambiguity	NaN
sector_3_seconds / sector_3_share	289	0.3%	Sector telemetry gap	NaN
sector_2_seconds / sector_2_share	176	0.2%	Sector telemetry gap	NaN
cars_ahead	142	0.1%	Position data unavailable	NaN
position_percentile	142	0.1%	Derived from position	NaN
position	142	0.1%	Timing data unavailable	NaN
cars_behind	142	0.1%	Position data unavailable	NaN

Table A.2. *Difference-in-differences estimates for 12 upgrade packages (FIA, 2023).*

Team	Season	Round	DiD effect (s)	95% CI	Significant?	Direction	Upgrade description
McLaren	2023	9	+0.863	[+0.246, +1.486]	Yes	Positive	MCL60 B-spec front wing and floor revision (Austrian GP)
Mercedes	2024	11	+0.168	[−0.151, +0.503]	No	Positive	W15 front wing and floor concept revision
Ferrari	2022	6	+0.140	[−0.801, +1.167]	No	Positive	New underbody floor and beam wing update
McLaren	2024	5	−0.072	[−0.471, +0.309]	No	Negligible	MCL38 aero continuation package and floor revision
Aston Martin	2023	5	−0.405	[−0.631, −0.168]	Yes	Negative	AMR23 major aero overhaul

Team	Season	Round	DiD effect (s)	95% CI	Significant?	Direction	Upgrade description
							incl. floor and sidepod
Mercedes	2022	9	-0.439	[-1.770, +0.496]	No	Negative	Zero-sidepod concept with revised cooling layout
Ferrari	2024	7	-0.483	[-1.256, +0.289]	No	Negative	SF-24 diffuser and rear wing update
Red Bull Racing	2024	8	-0.504	[-1.048, +0.007]	No	Negative	RB20 floor revision and diffuser
Williams	2023	12	-0.612	[-1.893, +0.206]	No	Negative	New sidepod and rear suspension geometry
Alfa Romeo	2022	7	-0.665	[-1.248, -0.068]	Yes	Negative	New front wing and floor package
Red Bull Racing	2022	10	-1.186	[-2.347, -0.186]	Yes	Negative	Revised floor and diffuser with new edge winglets
Alpine	2023	10	-4.098	[-8.447, +0.221]	No	Negative	A523 aero update with revised diffuser

Appendix B - Data Dictionary

All columns present in the final feature dataset (historical_featured_laps_2022_2025.csv). The first 19 columns form the canonical schema. The frozen contract produced by the preprocessing pipeline. All remaining columns are engineered features added downstream.

Part 1: Canonical Schema (19 columns)

These columns are produced by preprocess.py and never change. They form the stable data contract between raw session loading and all downstream feature engineering.

Column	Type	Description
season	int	Championship year (2022-2025)
round_number	int	Round number within the season
event_name	str	Grand Prix name (e.g. 'British Grand Prix')
session_name	str	Session type - always 'R' (Race) in this dataset
driver	str	Three-letter FIA driver code (e.g. 'VER', 'NOR')
team	str	Constructor name (e.g. 'McLaren', 'Red Bull Racing')
lap_number	int	Lap number within the race
lap_time_seconds	float	Lap duration in seconds; null if telemetry unavailable
sector_1_seconds	float	Sector 1 time in seconds; null if unavailable
sector_2_seconds	float	Sector 2 time in seconds; null if unavailable
sector_3_seconds	float	Sector 3 time in seconds; null if unavailable
stint	float	Stint number (1 = first tyre set, 2 = after first stop, etc.)
compound	str	Raw tyre compound label from API (may include nulls)
tyre_life	float	Number of laps completed on the current tyre set
is_pit_out_lap	bool	True if this is the first lap after a pit stop (tyre_life < 2)
track_status	int	Encoded track condition flag (1 = green, 4x = SC, 6x = red flag)
position	float	Race position at end of lap; null if timing gap
is_deleted	bool	True if lap time was deleted by stewards (track limits etc.)
source	str	Data source identifier (e.g. 'fastf1')

Part 2: Engineered Features

Tyre and Stint Features

Column	Type	Description
stint_lap_index	float	Lap count within the current stint (resets to 1 after each pit stop)
is_first_stint_lap	bool	True if this is the first lap of a stint
tyre_life_lag_1	float	Tyre life on the previous lap (null on first stint lap)
tyre_life_delta	float	Change in tyre life lap-over-lap (null on first stint lap)

Column	Type	Description
compound_clean	str	Normalised compound label: SOFT / MEDIUM / HARD / INTERMEDIATE / WET / UNKNOWN
compound_is_soft	bool	One-hot: True if compound_clean = SOFT
compound_is_medium	bool	One-hot: True if compound_clean = MEDIUM
compound_is_hard	bool	One-hot: True if compound_clean = HARD
compound_is_intermediate	bool	One-hot: True if compound_clean = INTERMEDIATE
compound_is_wet	bool	One-hot: True if compound_clean = WET
compound_is_unknown	bool	One-hot: True if compound label is missing or unrecognised
fresh_tyre_run	bool	True if driver is on a fresh (unused) tyre set
tyre_life_to_lap_ratio	float	Tyre life divided by current lap number - fraction of race on current tyres

Race Position and Gap Features

Column	Type	Description
position_delta_driver	float	Change in race position since last lap (negative = gained)
gained_positions_last_lap	bool	True if driver gained at least one position
lost_positions_last_lap	bool	True if driver lost at least one position
field_size	int	Number of classified drivers on this lap
position_percentile	float	Race position expressed as a percentile (0 = leader, 1 = last)
cars_ahead	float	Number of cars ahead of driver
cars_behind	float	Number of cars behind driver
has_made_pit_stop	bool	True if driver has already made at least one pit stop this race

Sector Shares

Column	Type	Description
sector_1_share	float	Sector 1 time as a fraction of total lap time
sector_2_share	float	Sector 2 time as a fraction of total lap time
sector_3_share	float	Sector 3 time as a fraction of total lap time

Cumulative Race Time and Gap Features

Column	Type	Description
cumulative_race_time_seconds	float	Total elapsed race time for this driver up to and including this lap
leader_cumulative_race_time_seconds	float	Cumulative race time for the race leader on this lap
gap_to_leader_seconds	float	Time gap to the race leader
gap_ahead_seconds	float	Time gap to the car immediately ahead

Column	Type	Description
gap_behind_seconds	float	Time gap to the car immediately behind
traffic_gap_min_seconds	float	Minimum gap to any car within a 5-second window (traffic exposure)
gap_balance_seconds	float	Difference between gap ahead and gap behind - positive = more exposed to car behind

Traffic Features

Column	Type	Description
has_clean_air_ahead	bool	True if gap ahead exceeds the clean-air threshold
has_clean_air_behind	bool	True if gap behind exceeds the clean-air threshold
close_car_ahead	bool	True if a car is within DRS / close-racing range ahead
close_car_behind	bool	True if a car is within close-racing range behind
traffic_density_window_5s	int	Number of cars within a 5-second time window around driver
is_traffic_boxed_in	bool	True if driver has close cars both ahead and behind
rejoin_window_open	bool	True if a gap exists for clean re-entry after a pit stop

Weather Features

Column	Type	Description
air_temp_celsius	float	Ambient air temperature at time of lap (°C)
track_temp_celsius	float	Track surface temperature at time of lap (°C)
humidity_pct	float	Relative humidity (%)
rainfall	bool	True if rain was recorded during the lap
wind_speed_kph	float	Wind speed (km/h)
wind_direction_degrees	int	Wind direction in degrees (0-360)
track_temp_delta_from_lap1	float	Track temperature change since lap 1 of the race (°C)
track_temp_rolling_mean_3	float	Rolling 3-lap mean of track temperature
wind_direction_sin	float	Sine encoding of wind direction (avoids 0°/360° discontinuity)
wind_direction_cos	float	Cosine encoding of wind direction
is_wet_race	bool	True if any lap in the session recorded rainfall - session-level flag

Pace History Features

Column	Type	Description
lap_time_delta_driver	float	Change in lap time vs previous lap (positive = slower)
lap_time_lag_1_driver	float	Driver's lap time on the preceding lap
lap_time_lag_2_driver	float	Driver's lap time two laps prior

Column	Type	Description
lap_time_rolling_mean_3_prior	float	Rolling mean of driver's 3 most recent lap times (excluding current)
lap_time_expanding_mean_prior	float	Expanding mean of all prior laps in the stint - the driver's session baseline
recent_pace_change_seconds	float	Difference between rolling_mean_3 and a longer window - detects pace fade
rolling_minus_expanding_pace_seconds	float	Short-window mean minus long-window mean - positive = recent laps slower than trend

Race Progress Features

Column	Type	Description
session_lap_number_max	int	Total number of laps in the race
lap_number_progress	float	Lap number as a fraction of total race laps (0 = start, 1 = final lap)

Track Status and Session Context Features

Column	Type	Description
has_track_status_anomaly	bool	True if track_status contains an SC / VSC / red-flag code
is_valid_green_lap	bool	True if lap is a clean green-flag lap eligible for model training
session_track_anomaly_rate_so_far	float	Fraction of laps in this session that have been anomalous up to this lap
session_green_flag_rate_so_far	float	Fraction of laps in this session that have been green-flag up to this lap
session_pit_out_rate_recent_3	float	Fraction of the 3 most recent laps that were pit-out laps (field-level)
session_field_pit_stop_share_so_far	float	Fraction of all drivers who have pitted at least once in this session
session_position_change_rate_so_far	float	Average rate of position changes across all drivers up to this lap
session_position_change_rate_recent_3	float	Position change rate over the 3 most recent laps
session_abs_position_delta_mean_so_far	float	Mean absolute position change per lap across the field so far

Appendix C - Model Configuration and Hyperparameters

Final hyperparameters for all five models as used in the thesis results. Parameters were selected through manual search on the 2025 validation season. All models are implemented in `src/fl_thesis/modeling/`.

Model 1 - Lap-Time Prediction

Algorithm: HistGradientBoostingRegressor (scikit-learn)

File: src/fl_thesis/modeling/baseline_regression.py

Target: lap_time_seconds

Parameter	Value	Rationale
max_iter	700	Plateau in validation MAE between 600-900 iterations; beyond 1,000 overfits 2022-2024
learning_rate	0.01	Key tuning finding: reducing from 0.05 → 0.01 was the largest single improvement
max_leaf_nodes	63	127 leaves consistently worsened validation MAE; 63 + L2 is the correct trade-off
min_samples_leaf	20	Guards against splitting on very small groups
l2_regularization	0.1	Shrinks leaf values; complements max_leaf_nodes restriction
Categorical features	Ordinal (native)	event_name, driver, team, compound_clean passed as categorical dtype
Missing value handling	Native NaN splits	HistGBM learns optimal direction for missing values; no imputation step

Ridge baseline (for comparison only, not primary model): alpha = 1.0, with StandardScaler + median imputation in pipeline.

Model 2 - Tyre Degradation

Algorithm: Degree-2 polynomial Ridge regression, one model per circuit × compound group

File: src/fl_thesis/modeling/tyre_degradation.py

Target: Pace deviation from driver's session median (s)

Parameter	Value	Rationale
polynomial_degree	2	Captures the warm-up (negative linear) and degradation cliff (positive quadratic)
include_bias	False	Intercept is fitted separately by Ridge; avoids double-counting
alpha (Ridge)	0.5	Moderate regularisation; prevents overfitting in small groups
fit_intercept	True	Allows per-group baseline offset
min_laps_per_group	8	Groups with fewer laps fall back to the global model
Global fallback	Yes	One global Ridge model trained on all laps; used for unseen circuit × compound pairs
Total group models fitted	86	Unique circuit × compound combinations meeting the 8-lap minimum

Pipeline per group: PolynomialFeatures(degree=2, include_bias=False) → Ridge(alpha=0.5)

Model 3 - Pit Classifier

Algorithm: LogisticRegression with L2 regularisation (scikit-learn)

File: src/fl_thesis/modeling/strategy_baseline.py

Target: will_pit_within_3_laps (binary, positive class ≈ 8.7% of laps)

Parameter	Value	Rationale
C (inverse regularisation)	1.0	Validated at C=0.3, 1.0, and 3.0; C=1.0 selected - minimal difference at this scale
class_weight	'balanced'	Corrects for the 8.7% positive class rate; without it the model ignores pit laps
max_iter	1,000	Sufficient for convergence on this feature set
solver	default (lbfgs)	Appropriate for L2 penalty with moderate feature count
penalty	L2	L2 regularisation provides cross-season generalisation; validated against GBM alternative

Alternative tested and rejected: HistGradientBoostingRegressor + CalibratedClassifierCV(method='isotonic', cv=5): top-3 policy hit rate 0.297 vs LogReg 0.503. GBM memorises season-specific event × compound patterns that do not transfer to 2025. LogReg selected as final classifier.

Model 4 - Value Regressor

Algorithm: Ridge regression (scikit-learn)

File: src/fl_thesis/modeling/strategy_value.py

Target: net_pit_gain_over_next_6_laps_seconds (counterfactual)

Parameter	Value	Rationale
alpha	1.0	Default; dataset is small (1,310 train rows) so regularisation is important
Missing value handling	Median imputation in pipeline	Ridge does not handle NaN natively
Feature set	57+ numeric + 4 categorical (encoded)	Includes all degradation signals and gap features
Look-ahead window	6 laps	Replaced 4-lap window; 6-lap window allows genuine positive predictions after pit-lane loss

Why 6 laps: A pit-lane loss of 22-25 seconds cannot be recovered within 4 laps of pace gain. The 4-lap window produced structurally negative targets and suppressed policy top-3 hit rate to 18.5%. The 6-lap window lifted it to 50.3%.

Model 5 - Strategy Policy (Three-Gate Ranking System)

File: src/fl_thesis/modeling/strategy_policy.py

Type: Rule-based scoring combining Models 3 and 4 outputs; not a trained model.

The policy assigns a score to each lap and ranks candidates within a driver-session. A lap reaches the watchlist only if it passes all three gates:

Gate	Threshold	Description
Gate 1: Pit probability	≥ 0.70	Classifier (Model 3) probability of pitting within 3 laps
Gate 2: Predicted gain	≥ -10.0 s	Value regressor (Model 4) net gain; set low to allow full scoring range

Gate	Threshold	Description
Gate 3: Expected value	≥ -2.0 s	Composite score: $\text{pit_probability} \times \text{predicted_gain} - \text{rejoin_penalty} - \text{traffic_penalty}$

Penalty parameters:

Parameter	Value	Description
rejoin_penalty_weight	0.2	Fraction of predicted re-entry gap loss applied as penalty
traffic_penalty_seconds	0.75	Fixed penalty per close car in the rejoin window

Watchlist thresholds (broader second-pass for the top-5 candidates):

Parameter	Value
pit_probability_threshold	0.35
predicted_gain_threshold	4.0 s
expected_value_threshold	2.0 s
top_k	5 candidates

Model 6 - Upgrade DiD Estimator

Algorithm: Difference-in-differences with bootstrap confidence intervals

File: src/fl_thesis/modeling/upgrade_effects.py

Target: Relative team pace change (seconds) around upgrade introduction

Parameter	Value	Description
window_races_before	3	Number of races before upgrade round used as pre-period
window_races_after	4	Number of races after upgrade round used as post-period
n_bootstrap	500	Bootstrap iterations for 95% confidence interval estimation
min_laps_per_window	30	Minimum clean green-flag laps required in each window; below this the upgrade is skipped
CI level	95%	Standard; significant = CI entirely above or below zero

Pace metric: Median clean race lap time for the team vs median for all other teams in the same window. The DiD subtracts the field's change from the team's change, removing general seasonal development trends.

Appendix D - Code Repository and Availability Statement

The full source code for the data pipeline, feature engineering, and all models described in this thesis is available at:

<https://github.com/balintcsc/F1-Data-Model>

This repo includes preprocess.py, the modelling code itself: src/f1_thesis/modelling. Lap timing and data were gathered through the open-source FastF1 API, with information on upgrades came from FIA car presentation submissions. Feature engineering breakdown is present in Appendix A. The repository is public for 1 month after research publication, and after is accessible on request.

Appendix E - Generative AI Usage Declaration

In accordance with the SBE guidelines on the use of generative AI, I declare that I used a generative AI assistants (Claude, Anthropic; Perplexity, Perplexity AI) as a support tools during the preparation of this thesis. Its use was limited to the following:

Perplexity: Structuring, outlining of report and collection of research papers for initial topic exploration. I used the tool to help organise the overall structure of the report and to sketch a section outline, which I then revised and populated myself.

Claude: Table creation. I used the tool to format and lay out tables. All underlying data were collected, computed, and provided by me; the assistant arranged the data I supplied and did not generate, estimate, or alter any values.

Language and code editing. I used the tool for grammar and spelling checks alongside minor phrasing corrections, without adding new content or arguments. The tool was also used in a limited way to support code development (bug fixing, and clarification of language features).

All algorithmic design, analysis, results, interpretation, and written argumentation are my own original work. I verified the accuracy and appropriateness of any AI-assisted output and take full responsibility for the final content of this thesis.