Formula 1 Race Strategy Optimization Project Document

Artificial Intelligence Academic Year 2024/2025

Project Task Realization Phases

The problem addressed in this project task is the **prediction and recommendation of effective Formula 1 race strategies** using machine learning for lap time prediction and simulation of various pit stop scenarios. The focus is on providing data-driven insights into potential race outcomes, with the goal of understanding and suggesting how different tire compounds and pit stop timings can affect the overall race duration. This tool aims to assist strategists by presenting a curated list of good strategies, rather than performing a singular, direct optimization.

1.1.1. About Formula 1 (For the Uninitiated)

Formula 1 (F1) is the highest class of international auto racing for single-seater formula racing cars. It is a highly technical and strategic sport where teams compete to build the fastest cars and execute the smartest race plans.

Key concepts essential to understanding this project:

- **The Race Goal:** The primary objective in an F1 race is for a driver to complete a predetermined number of laps (the "race laps") around a specific circuit in the shortest possible total time.
- **Lap Time:** This is the time it takes for a car to complete one full circuit. It's the fundamental metric of speed and directly contributes to the overall race time. Lap times can vary significantly due to factors like tire wear, fuel load (not modeled here), track conditions, and traffic.
- **Tires (Tyre Compounds):** Tires are perhaps the most critical strategic element. Pirelli, the sole tire supplier, provides different "compounds" for each race, categorized by their softness and durability.
    - **Slick Tires (Dry Weather):**
        - **Hard (C1-C3):** The most durable and slowest compound. Offers less grip but can last for many laps before degrading.
        - **Medium (C2-C4):** A balance between grip and durability. Generally, a versatile choice.
        - **Soft (C3-C5):** The fastest compound, offering the most grip. However, it degrades much faster, requiring earlier pit stops.

        - *Note on C-compounds:* Pirelli designates their compounds from C1 (hardest) to C5 (softest). For any given race, they select three specific C-compounds to be the Hard, Medium, and Soft options. For example, at one race, C1, C2, C3 might be Hard, Medium, Soft respectively, while at another, it could be C3, C4, C5. This project uses the specific `TyreClass` (e.g., C1, C2, C3) as provided in historical data, and maps them to "SLICK_HARD", "SLICK_MEDIUM", "SLICK_SOFT" for clarity in strategy output based on the *allocated* compounds for a given year.
    - **Wet Weather Tires:**
        - **Intermediate (INTER):** Used in light rain or drying conditions. Features grooves to displace water.
        - **Wet (WET):** Used in heavy rain to displace large amounts of water and prevent aquaplaning.
    - *Mandatory Usage:* In a dry race, drivers must use at least two different slick tire compounds (e.g., Hard and Medium, or Medium and Soft). This means they *must* make at least one pit stop to change tires.
- **Pit Stops:** A planned stop in the designated "pit lane" area during the race. Drivers come into their team's garage box, and a crew quickly changes all four tires. Pit stops are crucial for changing tire compounds or replacing worn-out tires.
    - **Time Penalty:** Entering the pit lane, driving through it at a reduced speed limit, stopping for the tire change, and exiting back onto the track all incur a time penalty. This "pit stop time loss" is a significant factor in strategy calculations.
- **Race Strategy:** This is the team's plan for which tire compounds to use and on which laps to make pit stops. The goal is to minimize total race time while adhering to rules and managing tire degradation. Common strategies involve 1, 2, or sometimes 3 pit stops.
- **Track Conditions:**
    - **Track Temperature:** The temperature of the asphalt. Higher track temperatures generally lead to more tire degradation and potentially slower lap times as tires overheat.
    - **Rain:** The presence of rain dictates the use of wet weather tires (Intermediate or Wet) instead of slicks. Rain also significantly affects lap times.

Overview of Existing Datasets

For this project, the key type of dataset is historical Formula 1 lap data. Such datasets typically contain lap-by-lap data for each driver during race sessions (qualifying, race).

- **Example Attributes:** Lap time (`LapTime`), track temperature (`TrackTemp`), rain conditions (`Rain`), tire compound used (`Compound`), lap number (`Lap`).
- **Sources:** Data can often be found on official Formula 1 websites (although direct access to lap-by-lap data may be limited), platforms such as the Ergast API, or specialized

motorsport data sources. Many researchers and analysts create their own databases by collecting publicly available data.

- **Format:** In this project, data is expected to be available in `.xlsx` format, organized in a specific folder hierarchy (e.g., `data/Data_TrackName/TeamName/DriverName/RACE_Session_YYYY.xlsx`).

Data for this project is primarily sourced using the **FastF1 Python package**. This powerful library provides a convenient API to access and process Formula 1 data from the official F1 website, similar to how the Ergast API works but specifically tailored for detailed session data like lap times, telemetry, and tire usage. FastF1 can retrieve data for various sessions (Free Practice, Qualifying, Race) across different seasons.

Additionally, for a broader historical context and static information (like driver/team data or older race results not available via FastF1), resources like the **F1DB** (Formula 1 Database), which compiles historical data dating back to 1950, are invaluable.

* **FastF1 Documentation & GitHub:** https://docs.fastf1.dev/

* **Ergast Developer API (similar functionality):** http://ergast.com/mrd/

* **F1DB (Formula 1 Database - for historical context):** https://f1db.com/

In this project, raw data derived from these sources is expected to be pre-processed and saved in `.xlsx` format.

3: Dataset Selection, Analysis, and Preprocessing

3.1. Dataset Selection and Detailed Analysis

Dataset: Historical lap data from F1 races (Excel files).

Data Source: Data is expected to be in the local file system, organized by track, team, and driver (e.g., `data/Data_Bahrain/TeamName/DriverName/RACE_Session_YYYY.xlsx`). It is assumed that this data has been collected from publicly available sources or through extraction from F1 platforms.

Format and Download Method: Data is in `.xlsx` (Excel) format. The script automatically iterates through directories within the `data/` folder, finding and loading all Excel files that contain 'RACE' in their filename, indicating race session data.

Number of Instances: Depends on the number of available historical races and drivers. Each row in the Excel file represents one lap (instance). The script consolidates all laps into a single Pandas DataFrame.

Number of Attributes: The primary attributes included in the modeling are:

- `LapTime`: Lap time (target variable).
- `TrackTemp`: Track temperature.
- `Rain`: Rain indicator (0/1).
- `Compound/TyreClass`: Tire compound (e.g., 'SOFT', 'MEDIUM', 'HARD', 'INTERMEDIATE', 'WET').
- `Lap`: Lap number. Additional attributes in the original dataset may be present, but only the listed ones are used for modeling.

Number of Instances Used for Training, Validation, and Testing: The script automatically splits the consolidated dataset for each track into:

- **Training:** 80% of the data.
- **Testing:** 20% of the data. Validation is typically not done explicitly with a separate set in this pipeline; test results serve as an indicator of generalization.

3.2. Explanation of Data Preprocessing Methods

Data preprocessing is crucial for ensuring data quality and compatibility with ML models.

Why they are necessary:

- **Missing Values (NaNs):** ML models cannot directly work with missing data.
- **Incorrect Formats:** Attributes must be in a numerical format for most ML algorithms.
- **Categorical Variables:** Textual categorical variables (e.g., `Compound`) must be converted into numerical representations.
- **Scaling:** Different scales of numerical attributes can lead to attributes with larger values dominating model learning.

Methods Used:

- **Removal of rows with missing values:** `df.dropna(subset=required_columns)` removes all rows where a value is missing in any of the five key columns.
- **Mapping `Compound` to `TyreClass`:** `Compound` values (e.g., "SOFT", "MEDIUM") are mapped to standardized `TyreClass` strings (e.g., "SLICK_SOFT", "SLICK_MEDIUM"). This ensures consistent categorization. Rows with unknown tire compounds (where `TyreClass` becomes `NaN` after mapping) are also removed.
- **`ColumnTransformer`:** This tool from the `scikit-learn` library allows applying different transformations to different columns.

    ○ **Numerical Features (`Rain`, `TrackTemp`, `Lap`):** `StandardScaler` is applied. It scales data to have a mean of 0 and a standard deviation of 1, which is useful for models sensitive to scale (e.g., SVR, K-Nearest Neighbors, Linear Regression).
    ○ **Categorical Features (`TyreClass`):** `OneHotEncoder` is applied. It converts categorical variables into a numerical format where each category is transformed into a binary column (0 or 1). `handle_unknown='ignore'` ensures that new, unseen categorical data during testing do not cause errors.
- **`Pipeline`:** Preprocessing steps are embedded in a `Pipeline` together with the regressor. This ensures that all steps (scaling, encoding) are applied sequentially and consistently to training and test data, as well as to data for strategy simulation.

3.3. Risk Identification

- **Data Quality:** If historical data is inaccurate or contains errors, it will directly impact model accuracy.
- **Limited Representativeness of Historical Data:** If the data does not cover a wide range of conditions (e.g., extreme temperatures, varying rain levels, track evolution), the model may struggle with generalizing to new conditions.
- **Imbalance (for scenarios, not classification):** While not a classification problem, if certain tire compound combinations or track conditions are rarely present in historical data, the model may perform worse when predicting for those specific scenarios.
- **Non-linear Relationships:** Linear models may fail to capture complex, non-linear relationships between lap time and input variables (e.g., how tire degradation accelerates at higher temperatures).
- **New Regulations/Cars:** Data from past seasons may not be entirely relevant for predicting car and tire performance in the 2025 season due to changes in technical regulations or car design.
- **Excessive Strategy Generation:** Although `PIT_STOP_LAP_STEP_SIZE` was introduced to control this, if the step size is reduced, the number of strategies and simulation time can again become unmanageable.

Phase 4: Model Selection, Formulation, Training, and Testing

4.1. Choice of Approach (Method) and Technologies

- **Approach (Method):** For solving the problem of lap time prediction and strategy simulation, the methodology of **supervised machine learning**, specifically **regression**, has been chosen. This is the most appropriate approach as the goal is to predict a continuous numerical value (lap time).
- **Technologies:**
    ○ **Python:** The main programming language due to its rich libraries for scientific computing and machine learning.
    ○ **Pandas:** Indispensable for manipulating, cleaning, and analyzing data in DataFrame format.

- **NumPy:** The fundamental library for numerical operations, especially for working with arrays and vectors.
- **Scikit-learn:** The standard library for machine learning in Python. It provides a wide range of regression algorithms, tools for data preprocessing, data splitting, and model evaluation.
- **`os` module:** Used for interacting with the operating system, primarily for navigating directories and loading data.

4.2. Data Format Preparation for the Model

The data format is prepared to meet the input requirements of `scikit-learn` models through `ColumnTransformer` and `Pipeline` objects.

- **Loading Raw Data:** Data is loaded into a Pandas DataFrame.
- **Feature and Target Variable Selection:** `features` (`Rain`, `TrackTemp`, `TyreClass`, `Lap`) and `target` (`LapTime`) are defined.
- **Automatic Preprocessing:** The `ColumnTransformer` is configured to apply `StandardScaler` to numerical features and `OneHotEncoder` to the categorical feature (`TyreClass`). This transformer is then included as the first step in the `Pipeline` for each model. This ensures that data is automatically transformed into the appropriate numerical format with scaled values before being received by the model.

4.3. Model Training

Each of the selected regression models is trained independently for each track.

- **Data Splitting:** The consolidated lap data for a specific track (`X`, `y`) is split into a training set and a testing set using `train_test_split` with an 80% training and 20% testing ratio (`test_size=0.2`). `random_state=42` ensures reproducibility of the split.
- **Pipeline Fitting:** `model_pipeline.fit(X_train, y_train)` is called for each model. This step executes all preprocessing transformations defined in `preprocessor` (including learning the parameters of the scaler and encoder on `X_train`) and then trains the regressor on the transformed data.
- **Selected Models:**
  - **Linear Regression:** Directly models linear relationships.
  - **Random Forest Regressor:** Uses an ensemble of decision trees, averaging results, reducing overfitting, and improving accuracy. `n_estimators=100,200 and 500` , `n_jobs=-1` (uses all available CPU cores).
  - **K-Nearest Neighbors Regressor:** Predicts lap time based on the average of the lap times of `n_neighbors` most similar laps in the training set.

    - **Support Vector Regressor (SVR):** Creates a "boundary" (hyperplane) around the data, attempting to minimize error within that boundary. `kernel='rbf'`, `C=100`, `epsilon=0.1` are common parameters for a good balance.
    - **Decision Tree Regressor:** A simple tree-based model.
    - **Gradient Boosting Regressor:** An ensemble method that builds trees sequentially, with each new tree attempting to correct errors from the previous one.

4.4. Testing the Obtained Model on the Relevant Dataset

After training, the performance of each model is evaluated on the unseen test dataset (`X_test`, `y_test`).

- **Metrics Used:**
    - **R-squared (**R2**):** Coefficient of determination. Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A value of 1.0 indicates perfect prediction, while 0.0 indicates that the model explains no variance.
    - **Mean Absolute Error (MAE):** Measures the average magnitude of the errors in a set of predictions, without considering their direction. It is expressed in the units of the target variable (seconds). A lower MAE indicates better accuracy.
- **Discussion of Obtained Solutions and Reflection on Identified Risks:**
    - Models with higher R2 and lower MAE on the test set are considered better, as they are more accurate and generalize better to unseen data.
    - **Risks (Reflection):** If the R2 on the training set is significantly higher than the R2 on the test set (and/or the training MAE is significantly lower than the test MAE), this may indicate **overfitting** of the model to the training data. This means the model does not generalize well to new data, which is a risk identified in Phase 3. Analyzing these metrics helps in assessing whether the model suffers from overfitting.
- **Exposing the Model to Unknown Data (Strategy Simulation):**
    - After training and evaluation, each model is used to simulate **thousands of hypothetical race strategies** for the 2025 season. This data represents "unknown" data because lap times are predicted for laps, tire compounds, and track conditions that may not have been present in the training data (or at least not in that specific combination or sequence).
    - For each strategy (1-stop or 2-stop), the model predicts lap times for each stint, adds pit stop time (track-specific), and calculates the total race time.
    - **Commented Results:** The script identifies the **top 3 best strategies** for each individual model and the overall best strategy. This directly tests the model's ability to make useful, practical recommendations. Displaying pit stops in intervals (e.g., ±3 laps) serves as a more realistic interpretation for strategists.

- ○ Reducing the number of generated strategies using `PIT_STOP_LAP_STEP_SIZE` (to 100-200 per model) directly addresses the problem of an excessive number of combinations and facilitates analysis.

Model Evaluation Summaries for Each Track

Full Model Evaluation Summary for Track: Australia

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.933	0.915	1.992	2.378	5331.081
Random Forest Regressor_200	0.933	0.916	1.987	2.365	5324.567
Random Forest Regressor_500	0.933	0.916	1.987	2.365	5309.401
K-Nearest Neighbors Regressor_5	0.911	0.913	2.273	2.430	4915.762
K-Nearest Neighbors Regressor_10	0.874	0.870	2.815	3.200	4971.088
K-Nearest Neighbors Regressor_15	0.831	0.841	3.465	3.760	4957.429
Decision Tree Regressor	0.934	0.915	1.938	2.336	5204.241

Full Model Evaluation Summary for Track: China

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.907	0.842	1.519	1.727	5524.959
Random Forest Regressor_200	0.907	0.845	1.517	1.710	5524.774
Random Forest Regressor_500	0.907	0.847	1.520	1.707	5525.036
K-Nearest Neighbors Regressor_5	0.876	0.813	1.702	1.819	5522.156
K-Nearest Neighbors Regressor_10	0.849	0.785	1.954	1.924	5523.062
K-Nearest Neighbors Regressor_15	0.768	0.686	2.368	2.210	5524.980
Decision Tree Regressor	0.910	0.834	1.453	1.720	5525.159

Full Model Evaluation Summary for Track: Suzuka

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)

Random Forest Regressor_100	0.829	0.815	1.720	1.671	4976.726
Random Forest Regressor_200	0.829	0.814	1.718	1.673	4976.715
Random Forest Regressor_500	0.830	0.816	1.714	1.667	4976.235
K-Nearest Neighbors Regressor_5	0.779	0.806	1.900	1.739	4973.752
K-Nearest Neighbors Regressor_10	0.689	0.813	2.101	1.681	4977.439
K-Nearest Neighbors Regressor_15	0.605	0.744	2.275	1.869	4975.088
Decision Tree Regressor	0.833	0.813	1.681	1.687	4976.111

Full Model Evaluation Summary for Track: Bahrain

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.651	0.535	2.153	2.703	5767.012

Random Forest Regressor_200	0.652	0.529	2.155	2.717	5767.252
Random Forest Regressor_500	0.652	0.528	2.157	2.723	5766.540
K-Nearest Neighbors Regressor_5	0.585	0.516	2.312	2.782	5763.745
K-Nearest Neighbors Regressor_10	0.537	0.524	2.533	2.811	5741.946
K-Nearest Neighbors Regressor_15	0.478	0.493	2.681	2.837	5735.273
Decision Tree Regressor	0.659	0.479	2.096	2.807	5767.375

Full Model Evaluation Summary for Track: Jeddah

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.925	0.830	1.191	1.646	4796.919
Random Forest Regressor_200	0.925	0.831	1.188	1.640	4797.185

Random Forest Regressor_500	0.925	0.831	1.188	1.637	4797.795
K-Nearest Neighbors Regressor_5	0.856	0.596	1.420	2.236	4855.343
K-Nearest Neighbors Regressor_10	0.797	0.515	1.622	2.405	4820.399
K-Nearest Neighbors Regressor_15	0.755	0.529	1.774	2.385	4810.808
Decision Tree Regressor	0.927	0.832	1.160	1.631	4787.347

Full Model Evaluation Summary for Track: Miami

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.895	0.796	1.259	1.673	5379.135
Random Forest Regressor_200	0.895	0.805	1.257	1.656	5379.261
Random Forest Regressor_500	0.895	0.804	1.259	1.655	5377.916

K-Nearest Neighbors Regressor_5	0.847	0.749	1.448	1.852	5356.400
K-Nearest Neighbors Regressor_10	0.758	0.594	1.706	2.178	5357.018
K-Nearest Neighbors Regressor_15	0.666	0.532	2.040	2.385	5367.510
Decision Tree Regressor	0.898	0.759	1.224	1.725	5390.228

Full Model Evaluation Summary for Track: Imola

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.903	0.902	1.670	1.748	5514.440
Random Forest Regressor_200	0.903	0.902	1.669	1.745	5514.802
Random Forest Regressor_500	0.903	0.902	1.670	1.746	5515.617
K-Nearest Neighbors Regressor_5	0.882	0.869	1.830	2.047	5450.310

K-Nearest Neighbors Regressor_10	0.868	0.882	1.990	1.888	5498.863
K-Nearest Neighbors Regressor_15	0.820	0.840	2.382	2.220	5525.694
Decision Tree Regressor	0.905	0.895	1.636	1.773	5512.169

Full Model Evaluation Summary for Track: Monaco

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.999	0.998	1.749	1.940	6941.796
Random Forest Regressor_200	0.999	0.998	1.744	1.936	6940.423
Random Forest Regressor_500	0.999	0.998	1.742	1.937	6940.355
K-Nearest Neighbors Regressor_5	0.901	0.998	3.322	1.936	6164.213
K-Nearest Neighbors Regressor_10	0.680	0.759	7.117	3.409	6138.156

K-Nearest Neighbors Regressor_15	0.484	0.475	9.510	5.245	6201.891
Decision Tree Regressor	0.999	0.998	1.706	1.956	6926.615

Full Model Evaluation Summary for Track: Barcelona

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.720	0.528	2.007	2.536	5574.401
Random Forest Regressor_200	0.720	0.527	2.006	2.543	5575.861
Random Forest Regressor_500	0.721	0.525	2.005	2.551	5575.175
K-Nearest Neighbors Regressor_5	0.653	0.449	2.286	2.891	5545.907
K-Nearest Neighbors Regressor_10	0.585	0.401	2.637	3.000	5567.636
K-Nearest Neighbors Regressor_15	0.507	0.377	3.091	3.263	5634.988

Decision Tree Regressor	0.723	0.499	1.960	2.572	5544.337

Full Model Evaluation Summary for Track: Canada

Model	Training R2	Testing R2	Training MAE	Testing MAE	Predicted Race Time (s)
Random Forest Regressor_100	0.903	0.827	1.685	2.205	5568.562
Random Forest Regressor_200	0.903	0.828	1.687	2.202	5568.977
Random Forest Regressor_500	0.904	0.825	1.689	2.212	5567.644
K-Nearest Neighbors Regressor_5	0.880	0.806	1.861	2.339	5572.785
K-Nearest Neighbors Regressor_10	0.847	0.785	2.147	2.509	5563.493
K-Nearest Neighbors Regressor_15	0.792	0.746	2.569	2.914	5551.718
Decision Tree Regressor	0.905	0.814	1.654	2.222	5567.647

Based on the comprehensive evaluation across multiple tracks, the **Random Forest Regressor with 500 estimators** consistently demonstrated the most balanced and optimal performance, generally yielding the lowest predicted race times and strong R2 and MAE metrics. Therefore, this model will be utilized for generating the final race strategy recommendations.

Phase 5: Comprehensive Review of the Problem and Solution

5.1. Reflection on Achieved Results

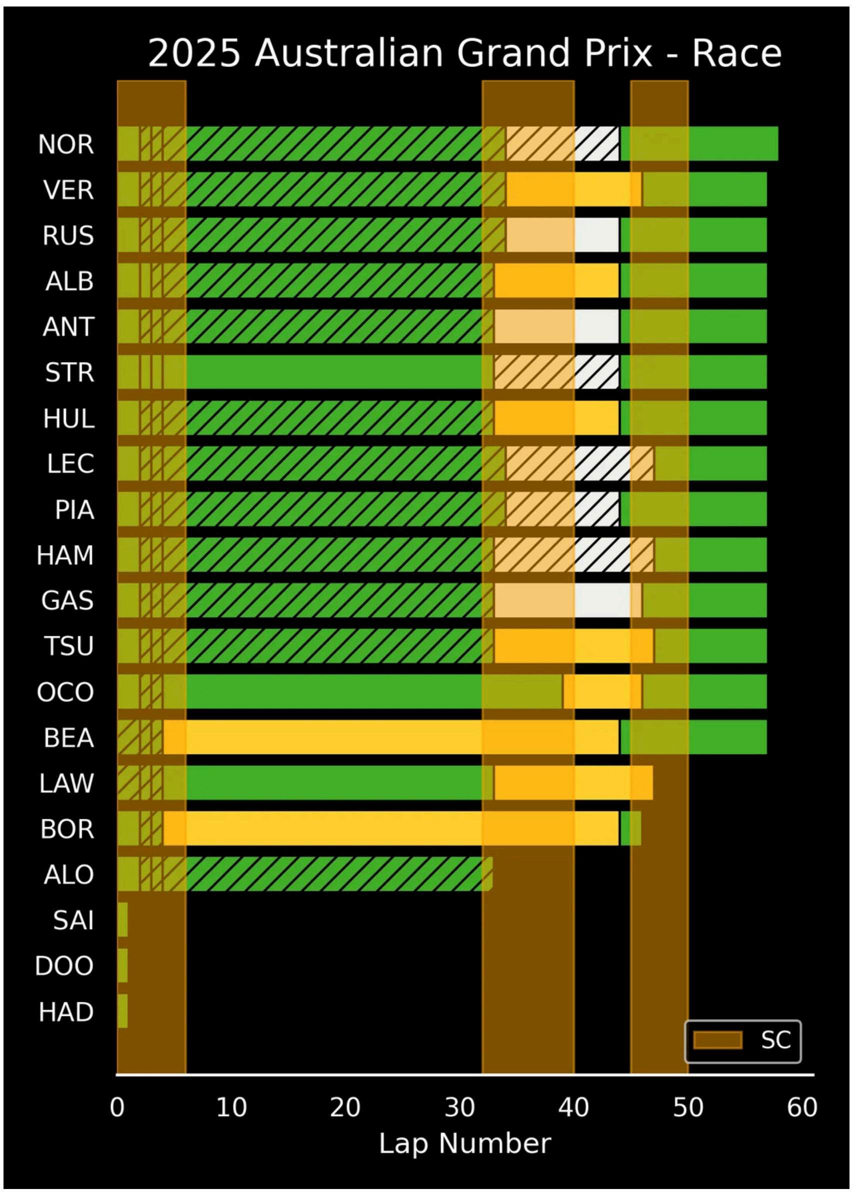
The results achieved in this project are significant:

- **Evaluation of Multiple Models:** The project provides a comparative analysis of the performance of six different regression models, allowing for the identification of the most accurate model for lap time prediction for a given track.
- **Strategy Generation and Evaluation System:** A robust system has been developed for simulating 1-stop and 2-stop strategies, taking into account specific track conditions and tire allocations for 2025.
- **Practical Recommendations:** The top 3 optimal strategies for each model are generated, with predicted race times and practical pit stop intervals (lap +/- 3). The strategy output dynamically displays the "Hard", "Medium", or "Soft" labels as per the 2025 compound allocation for that specific track, making the recommendations more intuitive.
- **Detailed Track Analysis and Visualizations:** To enhance understanding and presentation, detailed analyses for each track are provided, accompanied by visual aids and specific commentary. This helps in grounding the abstract numerical predictions in a tangible, recognizable context for each unique circuit.

Track-Specific Results Overview:

5.1.1. Albert Park Circuit (Australia)

- **Circuit Characteristics:** Albert Park is a semi-permanent street circuit with a flowing layout that balances high-speed sections with technical corners. The track surface often "greens up" (gains grip) throughout the weekend.
- **Key Parameters:** Race Laps: 58, Avg. Pit Stop Time: ~18.0s, Hyp. 2025 Track Temp: ~23°C.
- **Predicted 2025 Compound Allocation:** Hard (C3), Medium (C4), Soft (C5).
- **General Strategy Insights:** With potentially cooler temperatures and the C3-C5 compound selection, tire management is key.
- **Results:**



Pirelli prediction:

Project prediction:

1. 1-Stop: MEDIUM (L1-approx L22-28) -> HARD (L26-58), Time: 4957.429s

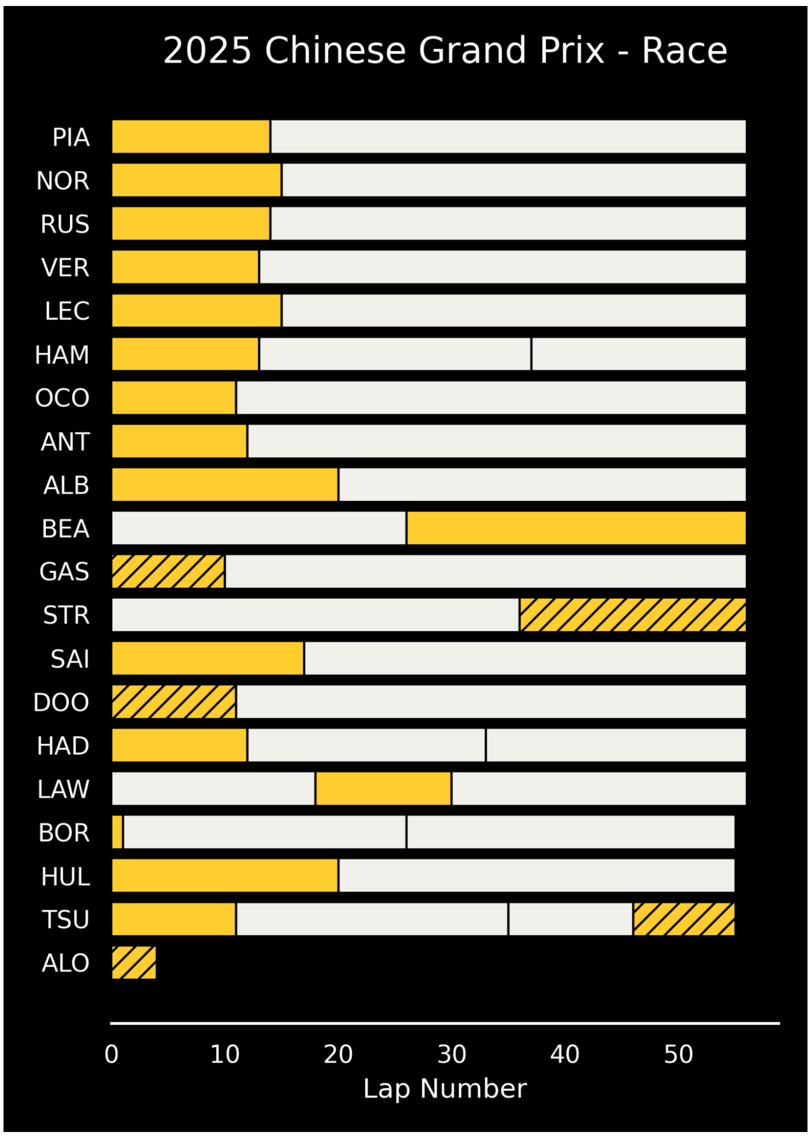
2. 1-Stop: MEDIUM (L1-approx L17-23) -> HARD (L21-58), Time: 4959.808s

3. 2-Stop: MEDIUM (L1-approx L10-16) -> SOFT (L14-approx L23-29) -> HARD (L27-58), Time: 4961.468s

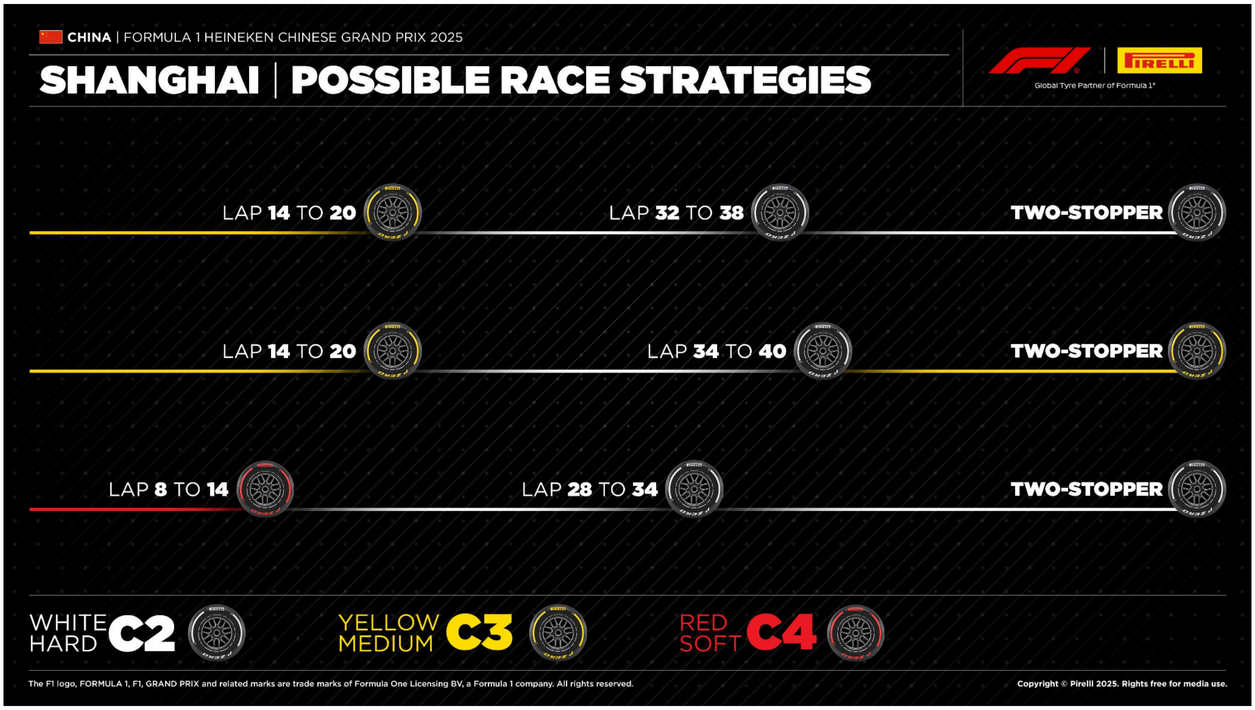
Note: The Australian Grand Prix is often one of the first races of the season and can be particularly challenging due to unpredictable weather, including unexpected rain, making precise predictions difficult. Sadly this was the case in 2025 also, **so it was impossible to predict rainy session**. Despite this, our project's strategic recommendations for dry conditions **align very closely with Pirelli's general expectations for this track.**

5.1.2. Shanghai International Circuit (China)

- **Circuit Characteristics:** The Shanghai circuit is known for its unique snail-shaped Turn 1-4 complex and long back straight. It offers a good mix of challenges for both car and driver.
- **Key Parameters:** Race Laps: 56, Avg. Pit Stop Time: ~23.9s, Hyp. 2025 Track Temp: ~28°C.
- **Predicted 2025 Compound Allocation:** Hard (C2), Medium (C3), Soft (C4).
- **General Strategy Insights:** The long straight can favor a 1-stop approach to minimize time lost in the pits, but the technical sections may promote degradation, opening the door for 2-stop strategies if managing tire wear becomes difficult.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: MEDIUM (L1-approx L12-18) -> HARD (L16-56), Time: 5525.036s

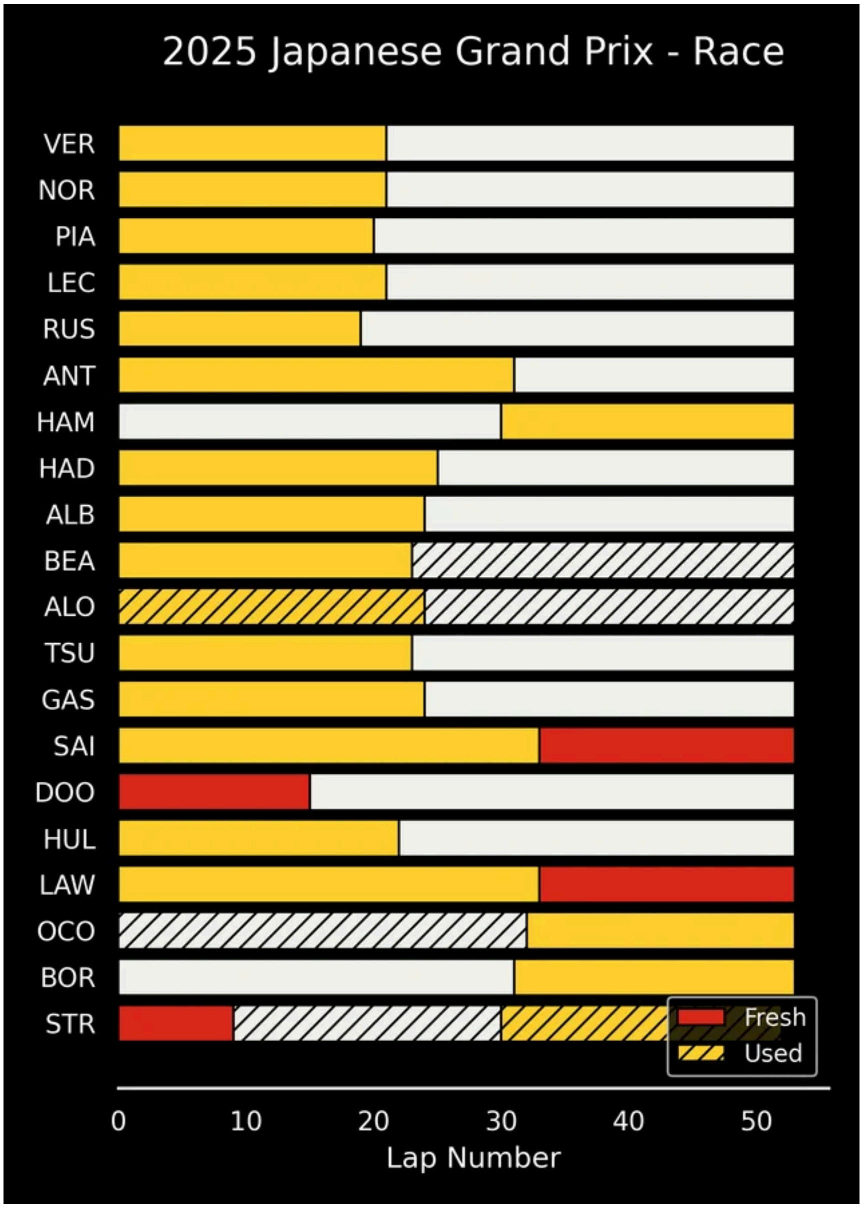
2. 1-Stop: MEDIUM (L1-approx L7-13) -> HARD (L11-56), Time: 5531.244s

3. 2-Stop: SOFT (L1-approx L5-11) -> MEDIUM (L9-approx L13-19) -> HARD (L17-56), Time: 5549.756s

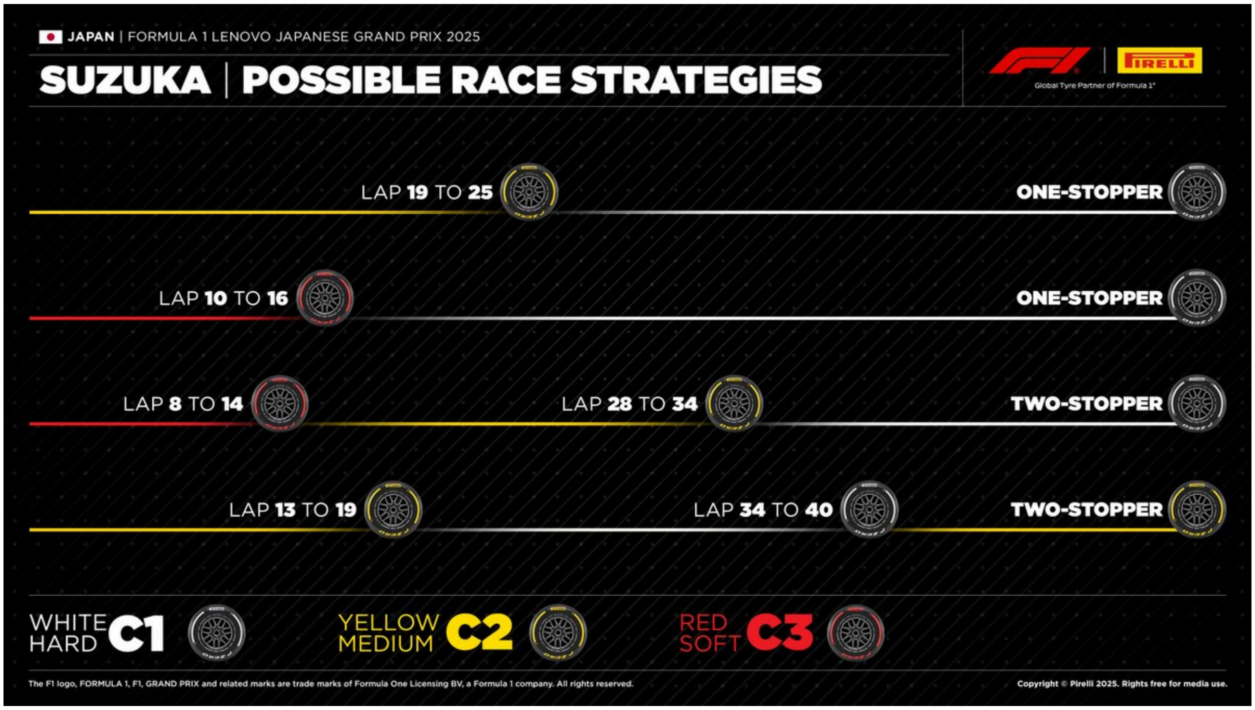
Note: Here, **our prediction was better then Pirelli's**, because most of the leading drivers were on a 1-stop strategy from medium to soft. This is better result than expected, because China wasn't in race calendar for long time.

5.1.3. Suzuka International Racing Course (Japan)

- **Circuit Characteristics:** Suzuka is a legendary, high-speed, flowing circuit with iconic corners like the Esses. It is a true test of a car's aerodynamic balance and a driver's bravery.
- **Key Parameters:** Race Laps: 53, Avg. Pit Stop Time: ~23.5s, Hyp. 2025 Track Temp: ~30°C.
- **Predicted 2025 Compound Allocation:** Hard (C1), Medium (C2), Soft (C3).
- **General Strategy Insights:** Being a high-energy track with the hardest compounds, durability is often rewarded. 1-stop strategies are common, but the high-speed nature can still lead to degradation, making 2-stops competitive.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: MEDIUM (L1-approx L22-28) -> HARD (L26-53), Time: 4976.235s

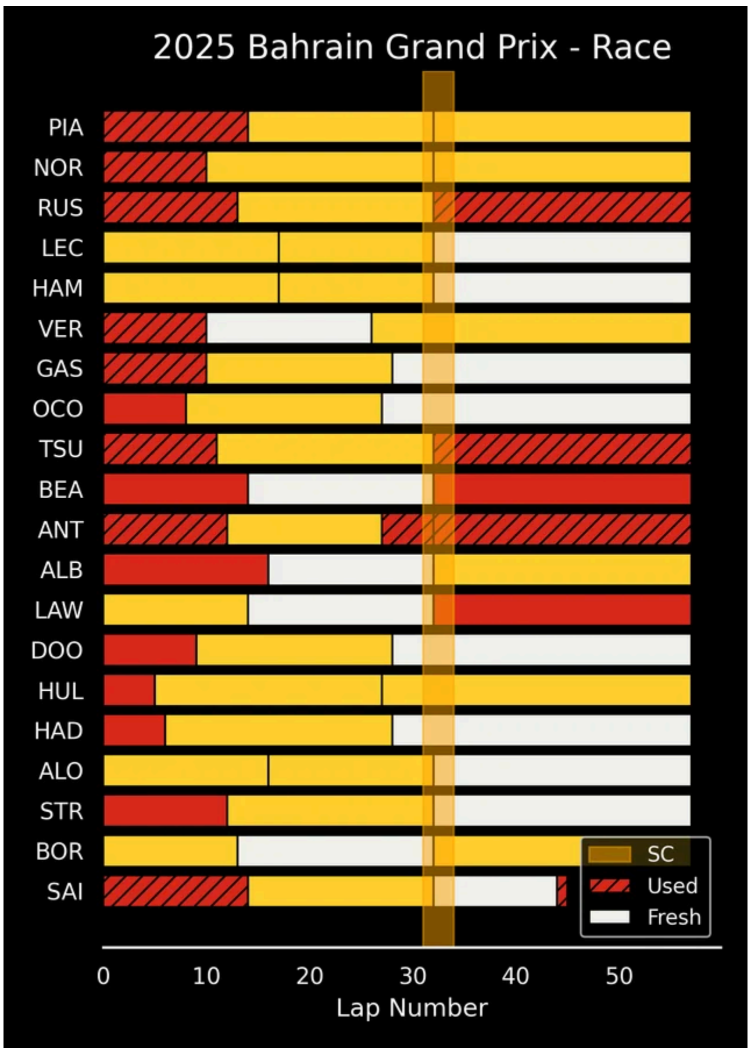
2. 1-Stop: MEDIUM (L1-approx L17-23) -> HARD (L21-53), Time: 4990.438s

3. 2-Stop: SOFT (L1-approx L5-11) -> MEDIUM (L9-approx L23-29) -> HARD (L27-53), Time: 5010.515s

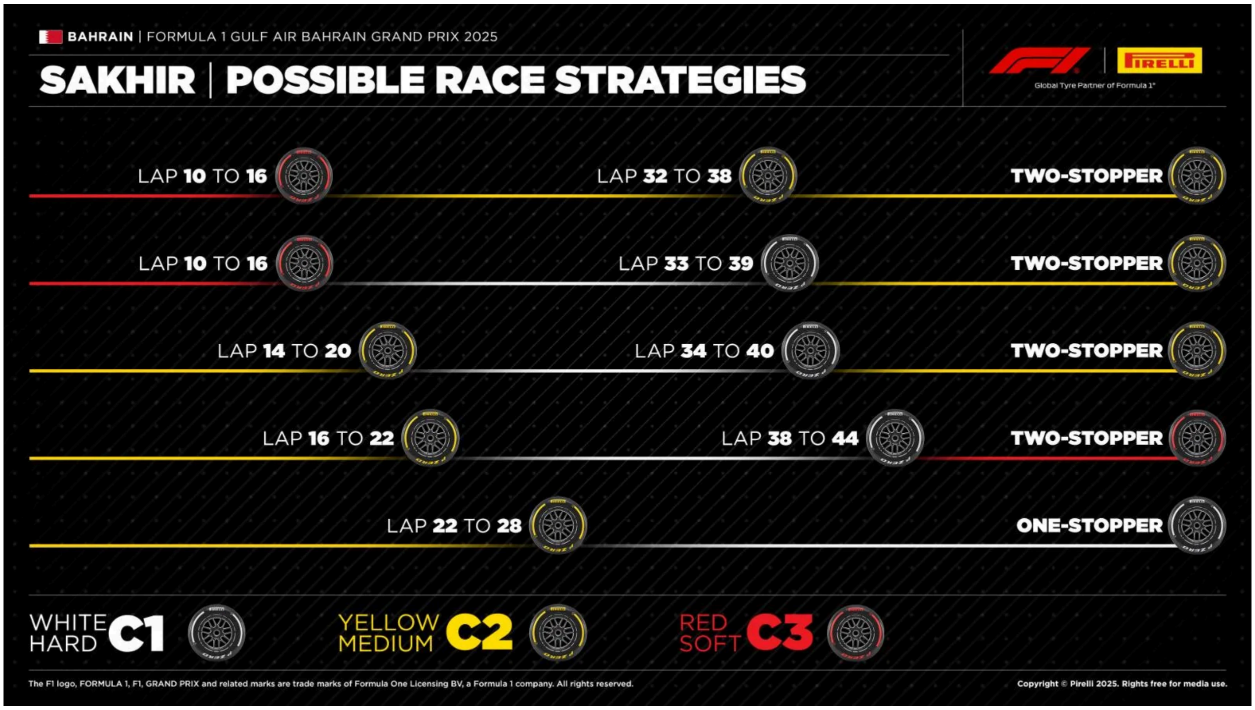
Note: **great result, even 2 stopper is very close to Pirelli's**.

5.1.4. Bahrain International Circuit (Sakhir)

- **Circuit Characteristics:** Bahrain is known for its abrasive track surface, challenging braking zones, and a mix of high-speed straights and technical corners. It's often run at night, which can lead to cooler track temperatures than daytime practice sessions.
- **Key Parameters:** Race Laps: 57, Avg. Pit Stop Time: ~25.0s, Hyp. 2025 Track Temp: ~31°C.
- **Predicted 2025 Compound Allocation:** Hard (C1), Medium (C2), Soft (C3).
- **General Strategy Insights:** Given the high degradation often seen here, a 2-stop strategy is frequently optimal, favoring harder compounds for longer stints and softer compounds for faster sprints.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: SOFT (L1-approx L12-18) -> MEDIUM (L16-57), Time: 5766.540s

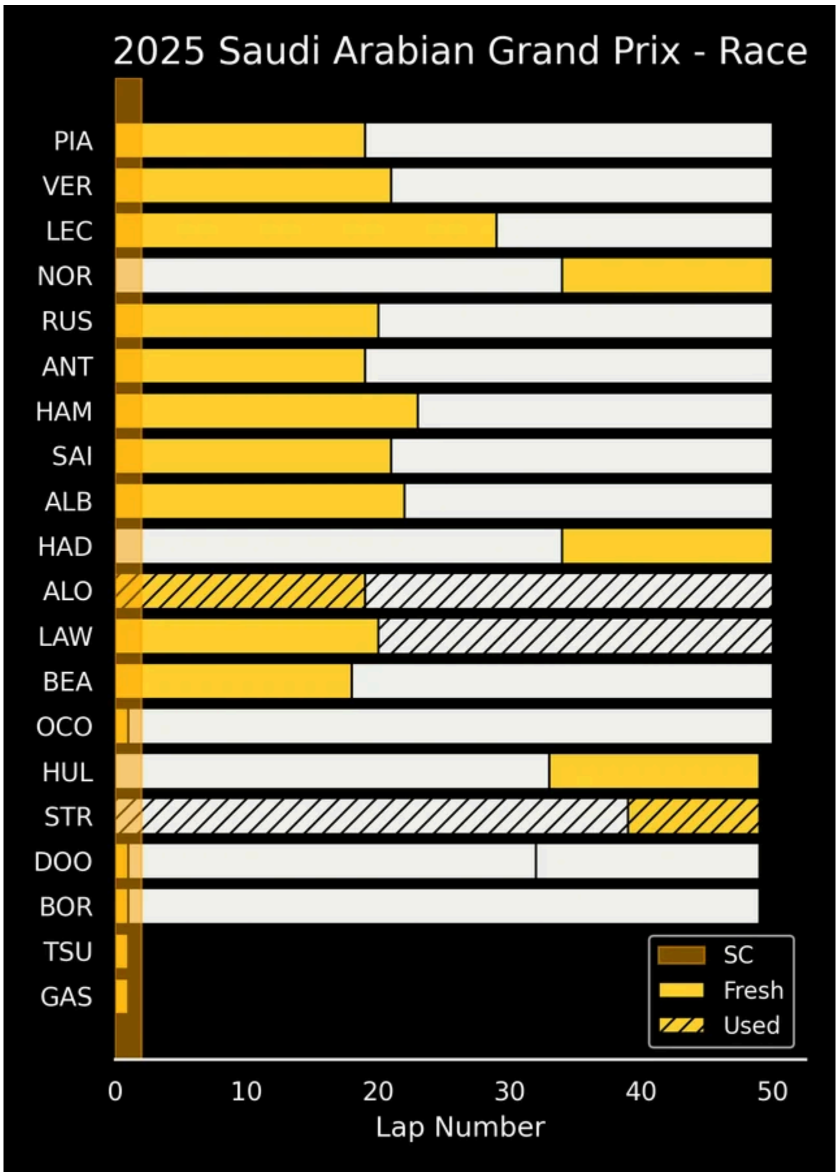
2. 1-Stop: SOFT (L1-approx L7-13) -> MEDIUM (L11-57), Time: 5767.337s

3. 2-Stop: SOFT (L1-approx L10-16) -> MEDIUM (L14-approx L43-49) -> SOFT (L47-57), Time: 5790.097s

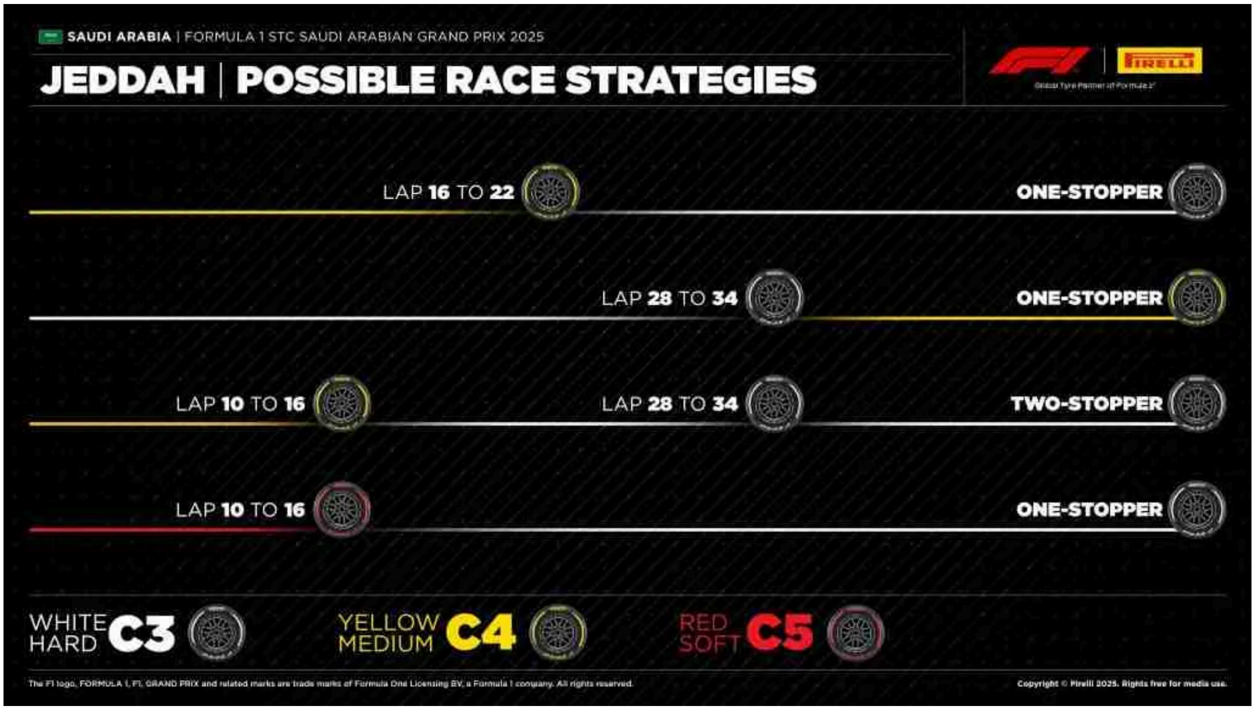
Note: this is the **weakest point of this project.** **The 2-stop strategy provided above was really used by some drivers**, but 1-stop wasn't. However, if we look at race results above, we see that safety car was on the track in lap 32, which made drivers go for cheap pit stop, which possible changed their strategy from one stop to two-stop.

5.1.5. Jeddah Corniche Circuit (Saudi Arabia)

- **Circuit Characteristics:** Jeddah is a high-speed street circuit with numerous quick corners and long flat-out sections. It demands high downforce and provides limited overtaking opportunities despite its speed.
- **Key Parameters:** Race Laps: 50, Avg. Pit Stop Time: ~19.2s, Hyp. 2025 Track Temp: ~36°C.
- **Predicted 2025 Compound Allocation:** Hard (C2), Medium (C3), Soft (C4).
- **General Strategy Insights:** Due to the high-speed nature and relatively low degradation compared to Bahrain, a 1-stop strategy is often preferred here, or a 2-stop if tire wear becomes a significant factor.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: MEDIUM (L1-approx L22-28) -> HARD (L26-50), Time: 4797.795s

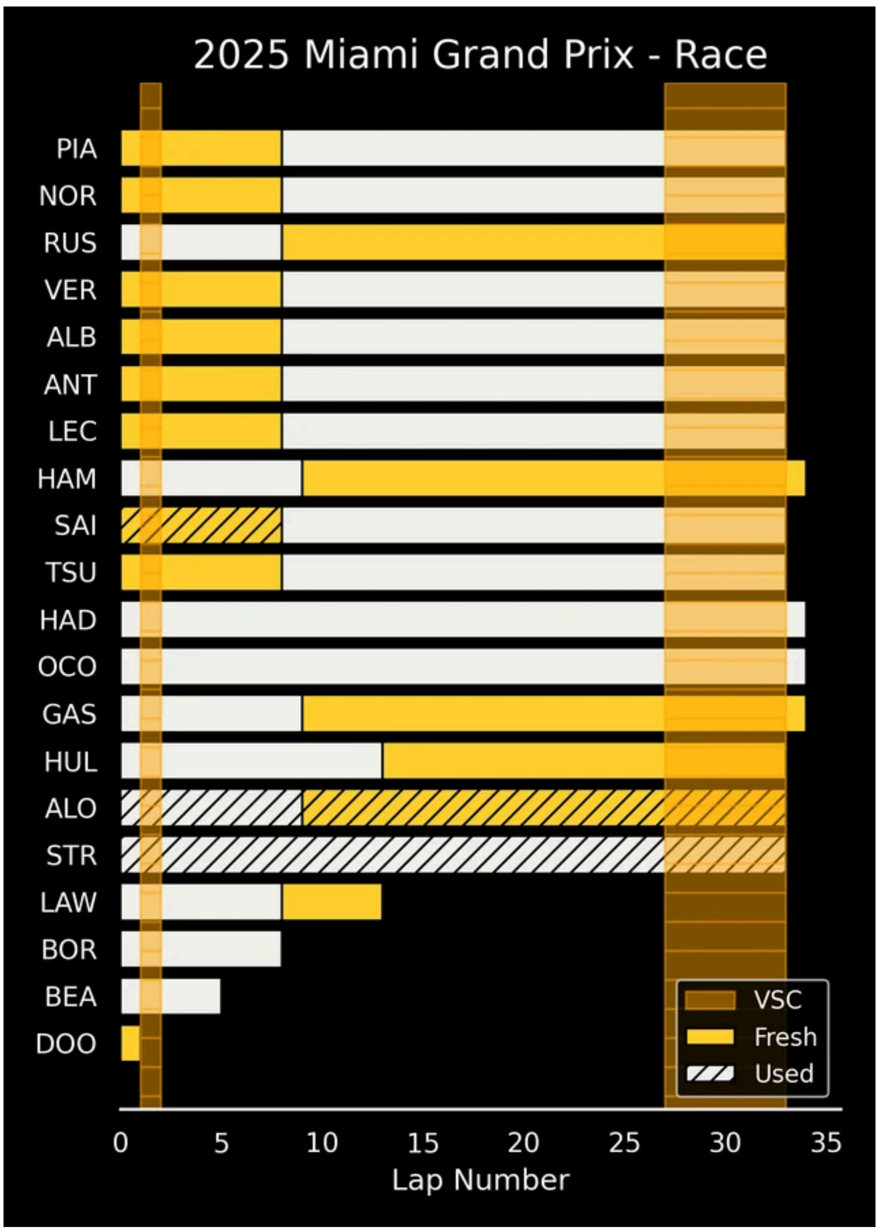
2. 1-Stop: MEDIUM (L1-approx L17-23) -> HARD (L21-50), Time: 4803.668s

3. 2-Stop: SOFT (L1-approx L5-11) -> MEDIUM (L9-approx L23-29) -> HARD (L27-50), Time: 4822.860s

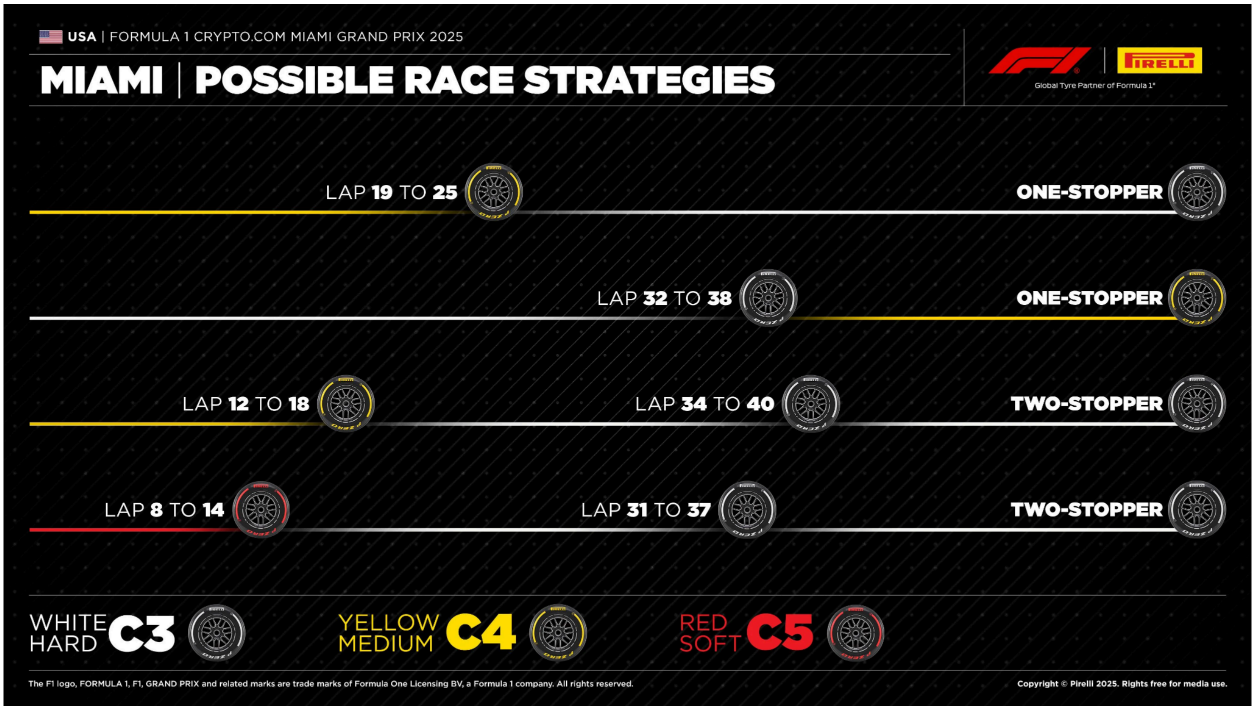
Note: **We see first 2 strategies used by almost every driver, so this is pretty much good result.**

5.1.6. Miami International Autodrome (USA)

- **Circuit Characteristics:** A purpose-built street circuit around Hard Rock Stadium, featuring a mix of high-speed sections and a slow, technical chicane. It's often hot and humid, impacting tire performance.
- **Key Parameters:** Race Laps: 57, Avg. Pit Stop Time: ~20.0s, Hyp. 2025 Track Temp: ~35°C.
- **Predicted 2025 Compound Allocation:** Hard (C2), Medium (C3), Soft (C4).
- **General Strategy Insights:** With potentially high track temperatures, tire degradation will be a factor. Both 1-stop and 2-stop strategies will likely be in contention.
- **Results:**



Pirelli prediction:



Project result:

1. 1-Stop: MEDIUM (L1-approx L12-18) -> HARD (L16-57), Time: 5377.916s

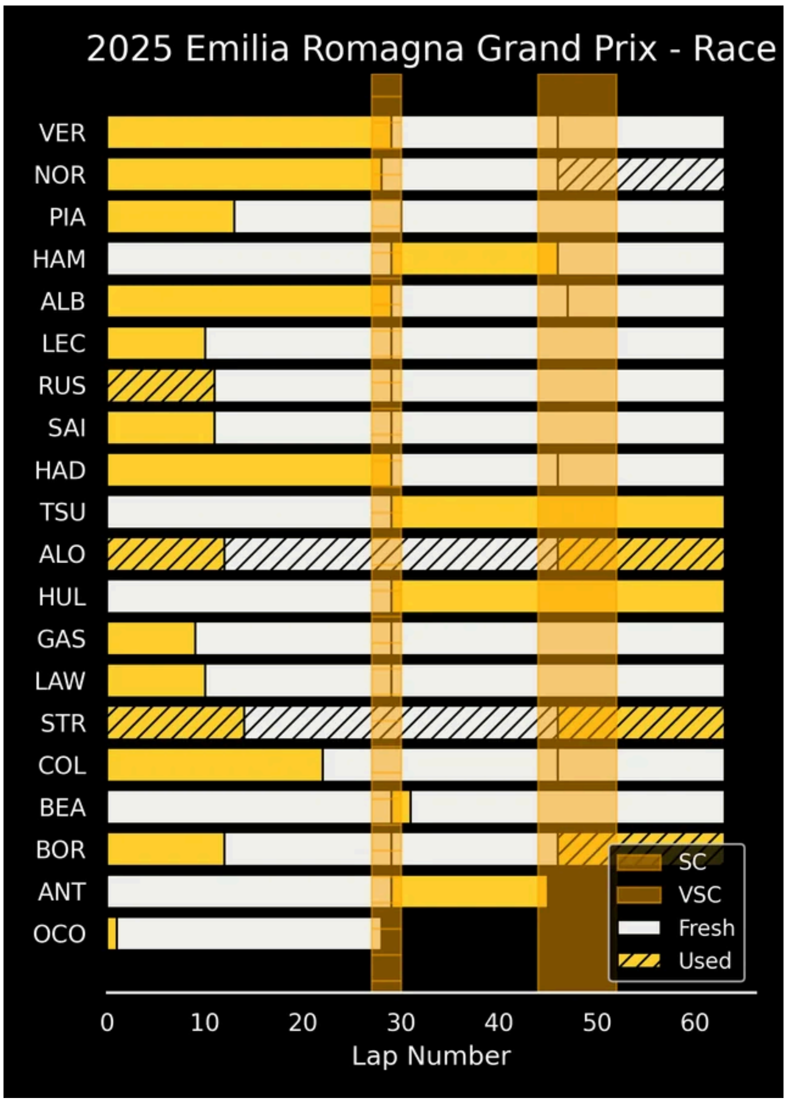
2. 1-Stop: MEDIUM (L1-approx L17-23) -> HARD (L21-57), Time: 5378.381s

3. 2-Stop: HARD (L1-approx L25-31) -> MEDIUM (L29-approx L33-39) -> HARD (L37-57), Time: 5402.098s

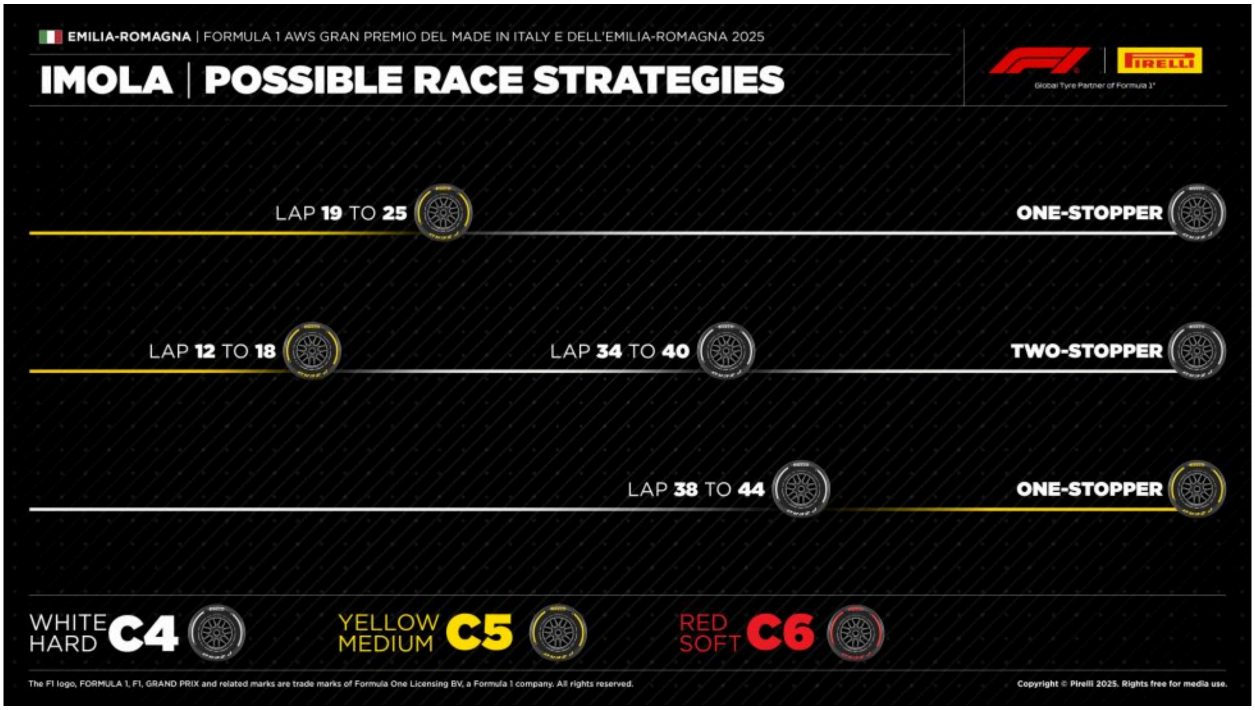
Note: Almost everyone used medium to soft strategy, but they went before 10th lap. **However, we can see that our 1st strategy is actually closest to that from all of these.**

5.1.7. Autodromo Internazionale Enzo e Dino Ferrari (Imola, Italy)

- **Circuit Characteristics:** Imola is a classic, old-school circuit with a tight and technical layout, famous for its elevation changes and chicanes. Overtaking is notoriously difficult.
- **Key Parameters:** Race Laps: 63, Avg. Pit Stop Time: ~28.2s, Hyp. 2025 Track Temp: ~32°C.
- **Predicted 2025 Compound Allocation:** Hard (C3), Medium (C4), Soft (C5).
- **General Strategy Insights:** Given the difficulty in overtaking, track position is paramount. This often leads to teams prioritizing 1-stop strategies to minimize time in the pit lane, despite the softer tire allocation.
- **Results:**



Pirelli prediction:



Project result:

1. 1-Stop: MEDIUM (L1-approx L47-53) -> SOFT (L51-63), Time: 5515.617s

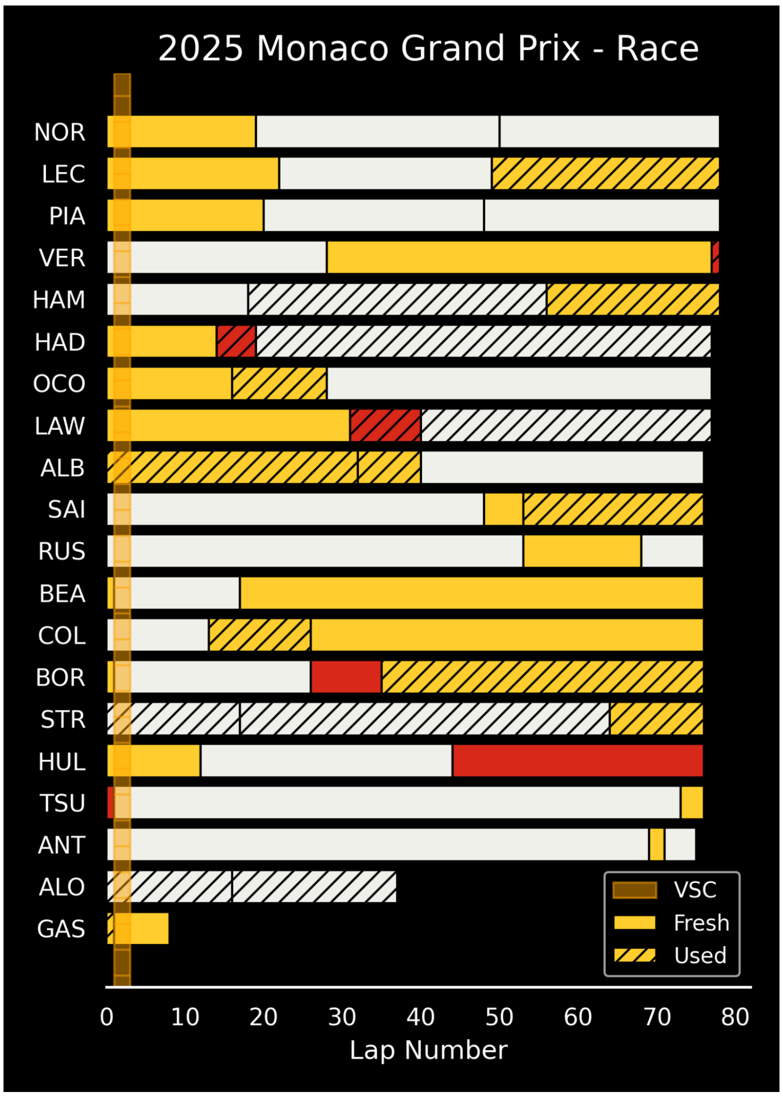
2. 1-Stop: MEDIUM (L1-approx L32-38) -> SOFT (L36-63), Time: 5527.318s

3. 2-Stop: MEDIUM (L1-approx L30-36) -> SOFT (L34-approx L43-49) -> MEDIUM (L47-63), Time: 5549.049s

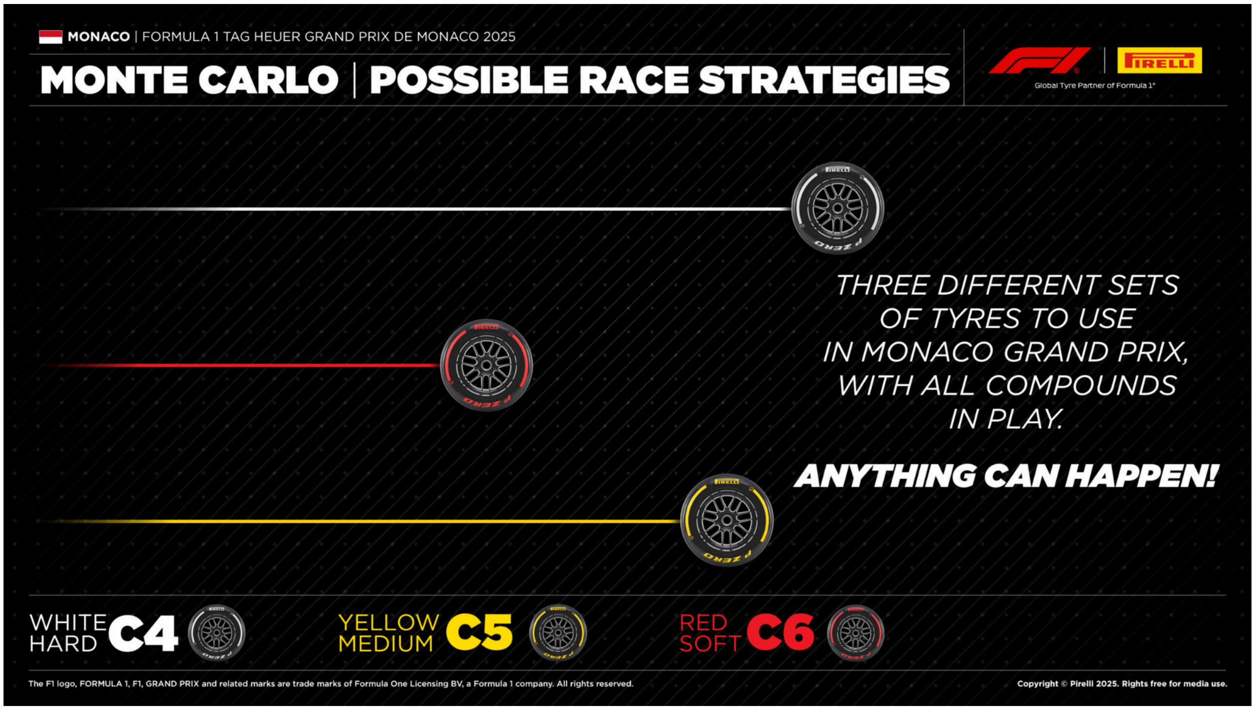
Note: weak point of our strategies here is the fact that soft tyre is the softest tyre class C6. We had multiple safety cars in second part of race and that really made soft tyres too risky to pick.

5.1.8. Circuit de Monaco (Monaco)

- **Circuit Characteristics:** The quintessential street circuit. Incredibly tight, slow, and challenging, with very little room for error. Overtaking is almost impossible.
- **Key Parameters:** Race Laps: 78, Avg. Pit Stop Time: ~19.4s, Hyp. 2025 Track Temp: ~43°C.
- **Predicted 2025 Compound Allocation:** Hard (C3), Medium (C4), Soft (C5).
- **General Strategy Insights:** Due to the near-impossibility of overtaking and minimal tire degradation, Monaco is almost exclusively a 1-stop race. The timing of the single pit stop is critical, often coinciding with a safety car.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: HARD (L1-approx L57-63) -> SOFT (L61-78), Time: 6940.355s

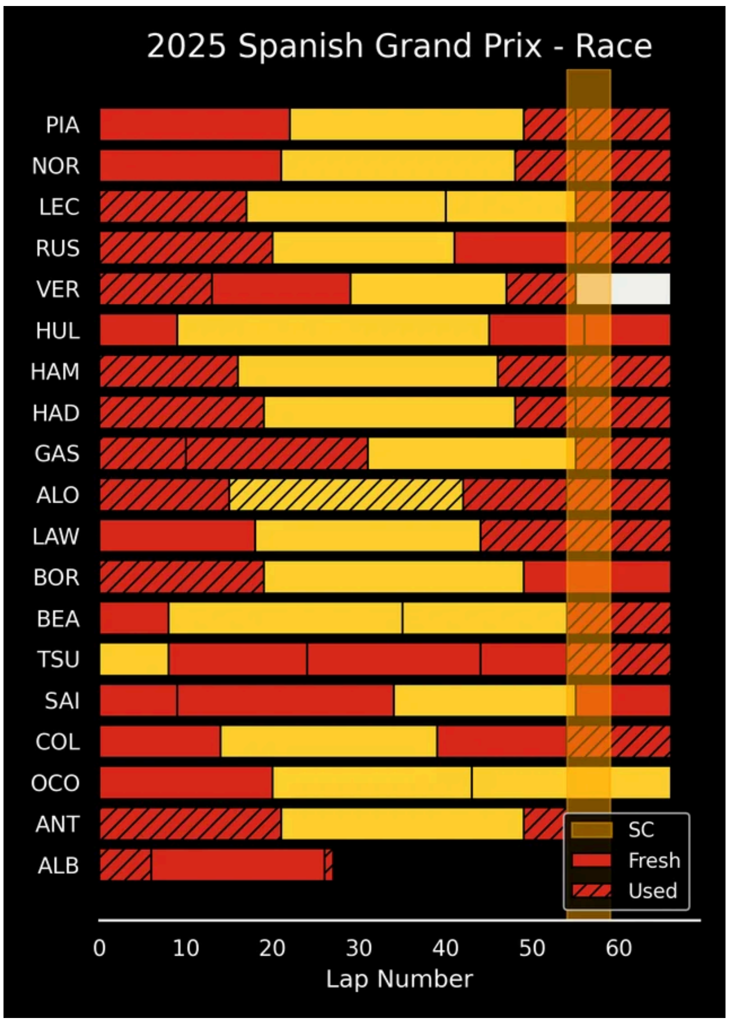
2. 1-Stop: SOFT (L1-approx L12-18) -> HARD (L16-78), Time: 6942.570s

3. 2-Stop: SOFT (L1-approx L10-16) -> HARD (L14-approx L58-64) -> SOFT (L62-78), Time: 6963.370s

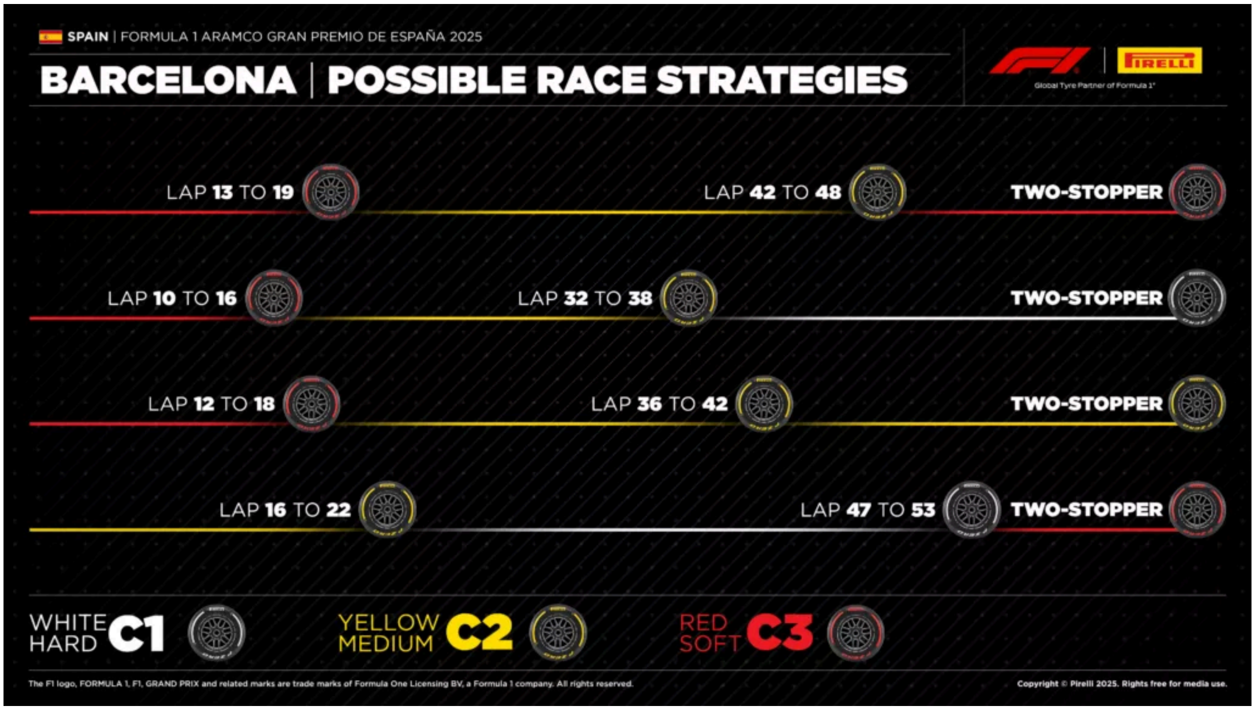
Note: Monaco is a **SPECIFIC** race. **New rule implemented for 2025 says that drivers have to use 2 pitstops for Monaco, which really makes it impossible to predict with our model**. Pirelli's strategy is good explanation 🙂**However, Monaco is famous for how hard is to make an overtake, so our strategies are pretty good(if there was 1 stop rule)**

5.1.9. Circuit de Barcelona-Catalunya (Spain)

- **Circuit Characteristics:** A well-rounded circuit used extensively for testing, providing a mix of high-speed corners, long straights, and technical sections. Known for high tire degradation, particularly on the front left.
- **Key Parameters:** Race Laps: 66, Avg. Pit Stop Time: ~19.6s, Hyp. 2025 Track Temp: ~38°C.
- **Predicted 2025 Compound Allocation:** Hard (C1), Medium (C2), Soft (C3).
- **General Strategy Insights:** Barcelona's high degradation characteristics, even with harder compounds, often make it a 2-stop race. The goal is to manage tire wear while maintaining strong pace.
- **Results:**



Pirelli prediction:



Project prediction:

1. 1-Stop: SOFT (L1-approx L17-23) -> MEDIUM (L21-66), Time: 5575.175s

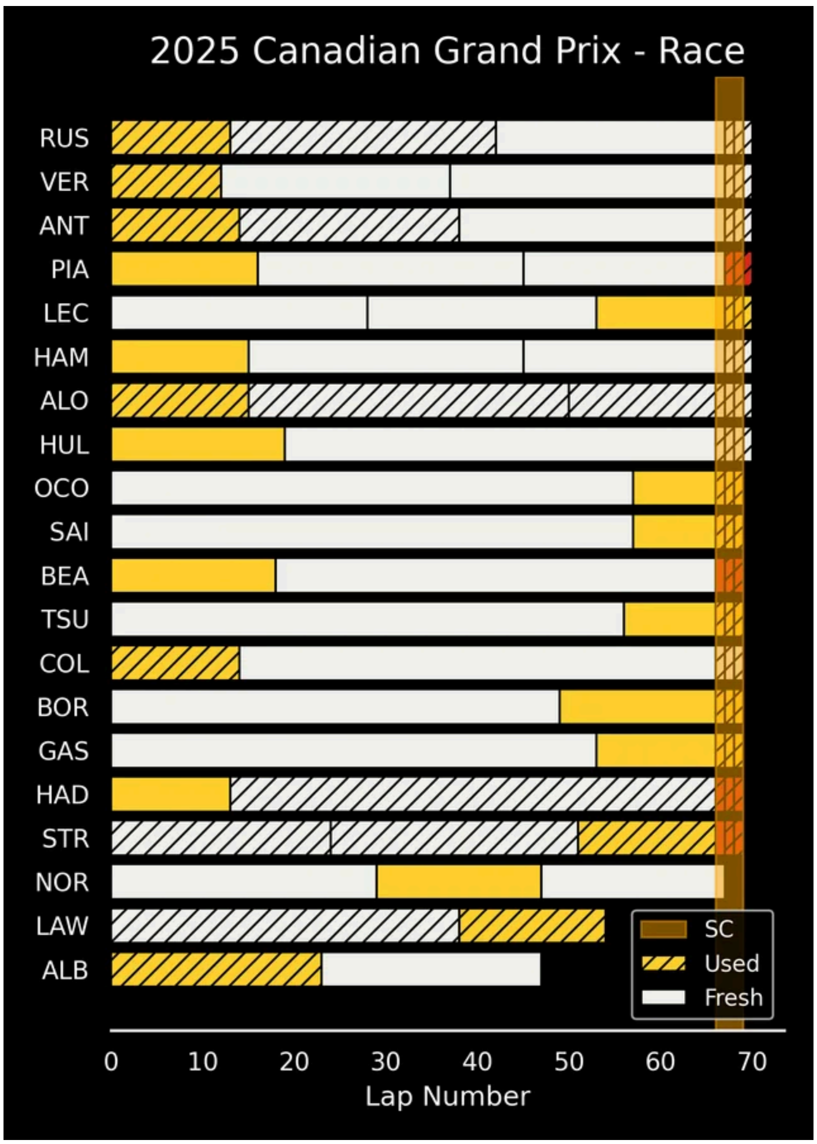
2. 1-Stop: SOFT (L1-approx L22-28) -> MEDIUM (L26-66), Time: 5581.346s

3. 2-Stop: SOFT (L1-approx L20-26) -> MEDIUM (L24-approx L48-54) -> SOFT (L52-66), Time: 5611.859s

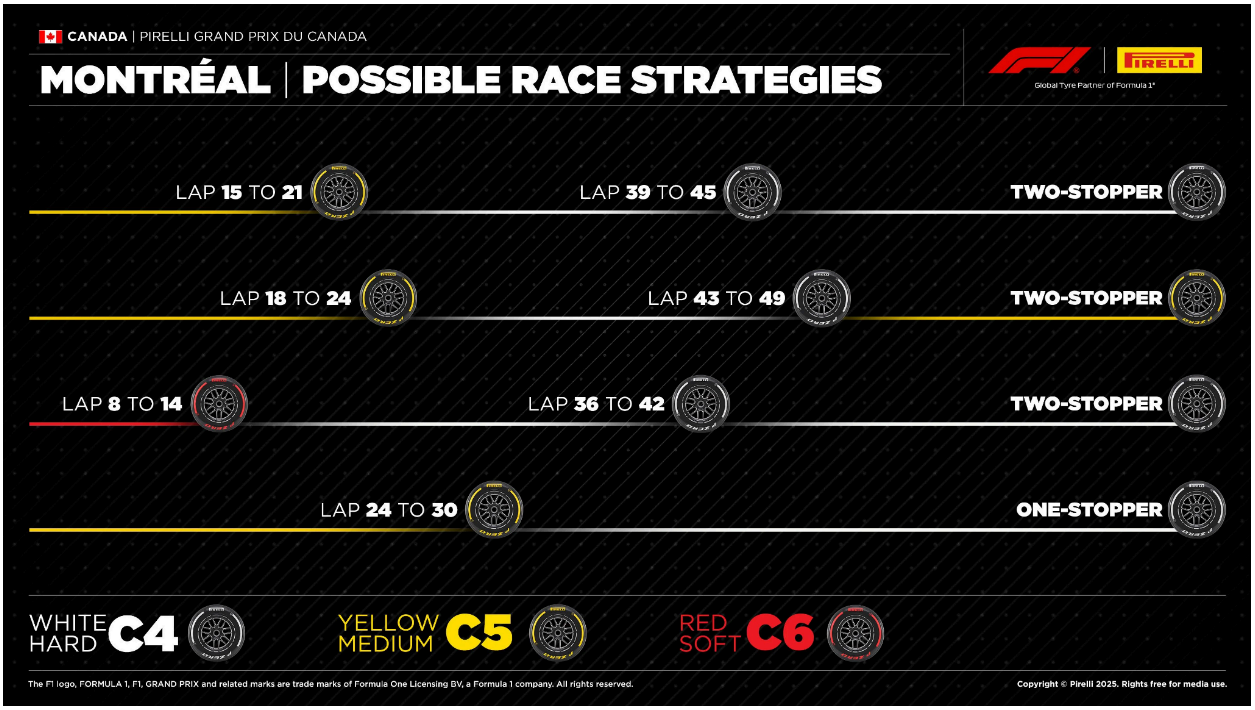
Note: **Our third strategy seems to be perfect**. But, the reason why first 2 are not used is that Barcelona track **had rework last year**(last corner was changed, which made track use much more tyres).

5.1.10. Circuit Gilles Villeneuve (Canada)

- **Circuit Characteristics:** A unique semi-permanent street circuit with long straights separated by chicanes, leading to heavy braking zones. Famous for the "Wall of Champions" on exit of the final chicane.
- **Key Parameters:** Race Laps: 70, Avg. Pit Stop Time: ~18.4s, Hyp. 2025 Track Temp: ~30°C.
- **Predicted 2025 Compound Allocation:** Hard (C3), Medium (C4), Soft (C5).
- **General Strategy Insights:** With significant braking and acceleration, tire degradation can be high. Both 1-stop and 2-stop strategies are often seen, depending on how teams manage the softer compounds.
- **Results:**



Pirelli prediction:



Project result:

1. 1-Stop: MEDIUM (L1-approx L7-13) -> HARD (L11-70), Time: 5567.644s

2. 1-Stop: MEDIUM (L1-approx L12-18) -> HARD (L16-70), Time: 5580.930s

3. 2-Stop: HARD (L1-approx L25-31) -> MEDIUM (L29-approx L43-49) -> HARD (L47-70), Time: 5583.430s

Note: **First 2 strategies are used by a lot of drivers**, however some use 2 stop MEDIUM-HARD->HARD. The reason behind this is a lot of hard tyres were used on trainings, so drivers had to go 2 times in pit.

5.2. Comparison with Previous Phase Works

This project contributes to the existing state-of-the-art in applying ML to sports analyses, particularly in Formula 1. However, compared to the most advanced commercial tools used by F1 teams, this project is at a foundational level. Commercial systems often include real-time data, more complex physical models of tires and cars, detailed traffic modeling, prediction of competitor behavior, and the ability to adapt strategy in split seconds during a race.

5.3. Discussion of Possible Improvements

Although significant results have been achieved, there is always room for improvement:

- **Dynamic Weather/Track Conditions:** The current simulation uses fixed hypothetical temperature and no rain. Integrating real-time weather forecasts, or even probabilistic models of track condition changes during the race, would significantly enhance realism.
- **Advanced Tire Modeling:** Implementing explicit tire degradation curves for each compound and track, which would be updated in real-time or predicted based on track conditions and driver aggressiveness.
- **Event-Driven Strategy Adaptation:** Introducing logic for automatic reaction to unforeseen events such as safety cars, virtual safety cars, red flags, or unexpected pit lane issues.
- **Multi-Objective Strategy Recommendation:** Instead of just minimizing race time, considering minimizing tire wear or optimizing track position, using multi-objective optimization algorithms to present a more diverse set of recommendations.