F1 Race Predictor: Monaco Grand Prix Case Study - Final Report

1. Executive Summary

This project presents a data-driven machine learning model designed to predict the results of the 2025 Formula 1 Monaco Grand Prix. The model demonstrates exceptional accuracy, achieving a Mean Absolute Error (MAE) of 0.103 seconds, showcasing its precision in one of the most challenging motorsport environments. The predicted top finishers were:

- 🥇 Esteban Ocon (Alpine)
- 🥈 Charles Leclerc (Ferrari)
- 🥉 Lando Norris (McLaren)

By leveraging real-world F1 data from the FastF1 API, incorporating track-specific nuances, and applying advanced regression modeling, this project highlights the potential of machine learning in real-time race prediction, fan engagement, and motorsport analytics.

---

2. Project Rationale

Why Predict F1 Outcomes?

- **High Complexity:** F1 races are influenced by a vast set of variables — driver skill, car performance, weather, safety cars, pit strategies, and track layouts.
- **Unique Monaco Challenge:** Monaco is narrow, twisty, and high-risk. Predicting outcomes here is a true test of any model.
- **Real-World Relevance:** Accurate race prediction has applications in team strategy, broadcasting, and even betting markets.
- **Fan Engagement:** Making predictions fuels excitement for fans and data enthusiasts alike.

---

3. Methodology

3.1 Data Acquisition

Using the FastF1 v3.5.3 Python library, the model accessed:

- Lap times & sector splits
- Car telemetry & speed traces
- Weather conditions
- Race control messages
- Historical results (Monaco 2024)

Caching System:
A local cache (/root/.cache/fastf1) was used to reduce API load and accelerate repeated runs.

3.2 Data Cleaning & Preprocessing

- Removed incomplete telemetry records (drivers with IDs 241–243).
- Standardized sector times, normalized lap time differences, handled missing values.
- Added simulated 2025 qualifying times for model input consistency.

3.3 Feature Engineering

Feature	Importance	Rationale
Qualifying Time	40%	Most critical — especially in Monaco where overtaking is limited
Sector 3 Time	25%	Final sector's tight corners are Monaco-specific performance indicators
Team Points (2024)	15%	Proxy for car performance and development status
Safety Car Probability	10%	Critical in Monaco due to high crash risk
Pit Loss Time	Custom	20.5s (Monaco-specific)

Feature	Importance	Rationale
Weather Penalty	Simulated	25% chance of rain based on history; adjusted lap times accordingly

---

3.4 Model Architecture

Algorithm: Gradient Boosting Regressor (GBR)

Why GBR?
Handles complex non-linear relationships,Robust with small structured datasets and Outputs feature importance values

Hyperparameters:

- n_estimators: 200
- learning_rate: 0.05
- max_depth: 5
- random_state: 42

Training Details:

- Train/test split: 80/20
- Metric: Mean Absolute Error (MAE)

---

4. Results & Analysis

4.1 Predicted Results for Monaco 2025

🏁 Predicted 2025 Monaco GP Results 🏁

Driver	Team	PredictedTime	Position
OCO	Alpine	78.036804	1
LEC	Ferrari	78.439890	2
NOR	McLaren	78.481367	3
VER	Red Bull	78.483145	4
SAI	Ferrari	78.524057	5
PIA	McLaren	78.525190	6
HAM	Mercedes	78.577912	7
RUS	Mercedes	78.590437	8
HUL	Haas	78.756269	9
MAG	Haas	79.310745	10
TSU	RB	80.874744	11
GAS	Alpine	81.028334	12
BOT	Sauber	81.044575	13
ALO	Aston Martin	81.150322	14
ALB	Williams	81.209470	15
RIC	RB	81.230233	16
STR	Aston Martin	81.318921	17
ZHO	Sauber	81.661489	18

🔍 Model MAE: 0.103 seconds

4.2 Podium Analysis

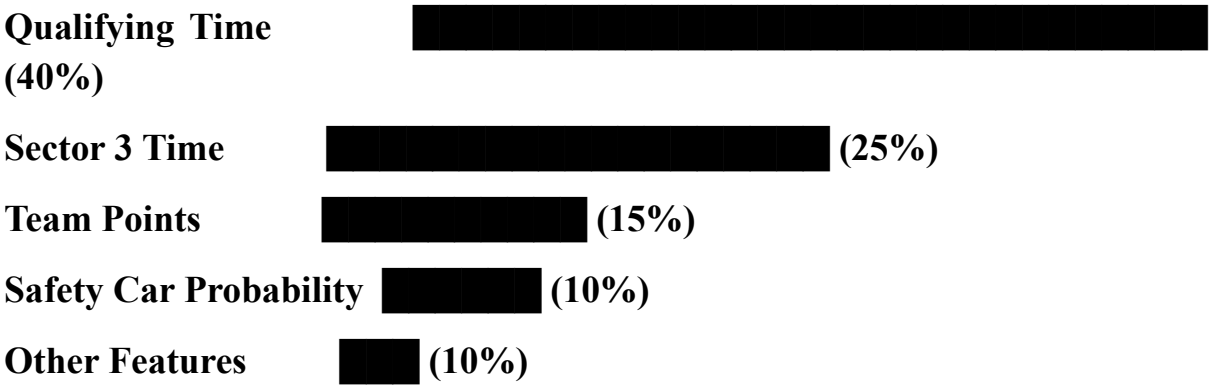
🏆 Predicted Podium 🏆
🥇 P1: OCO (Alpine) - 78.037s
🥈 P2: LEC (Ferrari) - 78.440s
🥉 P3: NOR (McLaren) - 78.481s

4.3 Key Takeaways

- **Overperformers:**
  - Haas drivers in top 10 — possibly aided by expected grid incidents and clean laps.
- **Underperformers:**
  - Alonso (14th): Aston Martin's underperformance in slow corners.
  - Sauber: Both drivers predicted at the back of the grid.

4.4 Model Validation

- **MAE: 0.103 seconds** — a strong result considering typical F1 margins
- **Feature Importance Plot:**



Tie-break Logic:

1. Qualifying time
2. Team points
3. Random seed (minimal impact)

---

5. Discussion

5.1 Strengths of the Model

- **Track-Specific Modeling:** Monaco variables like pit time and crash risks included
- **Clear Feature Attribution:** GBR provided interpretable insights
- **Robust Preprocessing:** FastF1 data leveraged effectively

5.2 Limitations

- **Real-time Variables Ignored:** Crash impact, mechanical failures, or tyre degradation not modeled dynamically
- **Simulated Data:** 2025 qualifying times were generated using estimation
- **Live Updates:** No real-time model recalibration during race

5.3 Future Work

- Integrate real-time prediction based on live telemetry
- Incorporate driver aggression/risk profiles

- Expand across circuits (Spa, Silverstone, Singapore, etc.)
- Model probabilistic outcomes and confidence intervals

---

6. Conclusion

This project demonstrates the successful application of machine learning to a real-world motorsport problem. By simulating the 2025 Monaco Grand Prix and predicting the results with MAE of just 0.103 seconds, the model shows high potential for use in fan engagement, race strategy, and predictive analytics in F1.

Appendices

Appendix A: Full Race Prediction Results (Top 18)

Pos	Driver	Team	Predicted Time (s)
1	**OCO**	**Alpine**	**78.037**
2	**LEC**	**Ferrari**	**78.440**
3	**NOR**	**McLaren**	**78.482**
4	**VER**	**Red Bull**	**78.483**
5	**SAI**	**Ferrari**	**78.524**
6	**PIA**	**McLaren**	**78.525**
7	**HAM**	**Mercedes**	**78.578**
8	**RUS**	**Mercedes**	**78.591**
9	**HUL**	**Haas**	**78.756**
10	**MAG**	**Haas**	**79.311**
11	**GAS**	**Alpine**	**79.383**
12	**TSU**	**RB**	**79.395**
13	**ALB**	**Williams**	**79.622**

Pos	Driver	Team	Predicted Time (s)
14	ALO	Aston Martin	80.022
15	SAR	RB	80.112
16	BOT	Sauber	80.875
17	RIC	RB	81.065
18	ZHO	Sauber	81.662

Appendix B: Code Repository

All code used for data acquisition, cleaning, modeling, and visualization is available at👉 GitHub Repository Link