MANIPAL UNIVERSITY JAIPUR
(University under Section 2(f) of the UGC Act)

PROJECT BASED LEARNING

F1NALYZE

FORMULA 1 DATA ANALYSIS AND PREDICTION MODEL

TEAM LEAD: ARYAN VERMA (23FE10CAI00188)
TEAM MEMBERS: ATISHAY JAIN (23FE10CAI00299)

B.TECH CSE (AI/ML SPECIALISATION)

28 NOVEMBER 2024

F1NALYZE

MANIPAL UNIVERSITY JAIPUR
(University under Section 2(f) of the UGC Act)

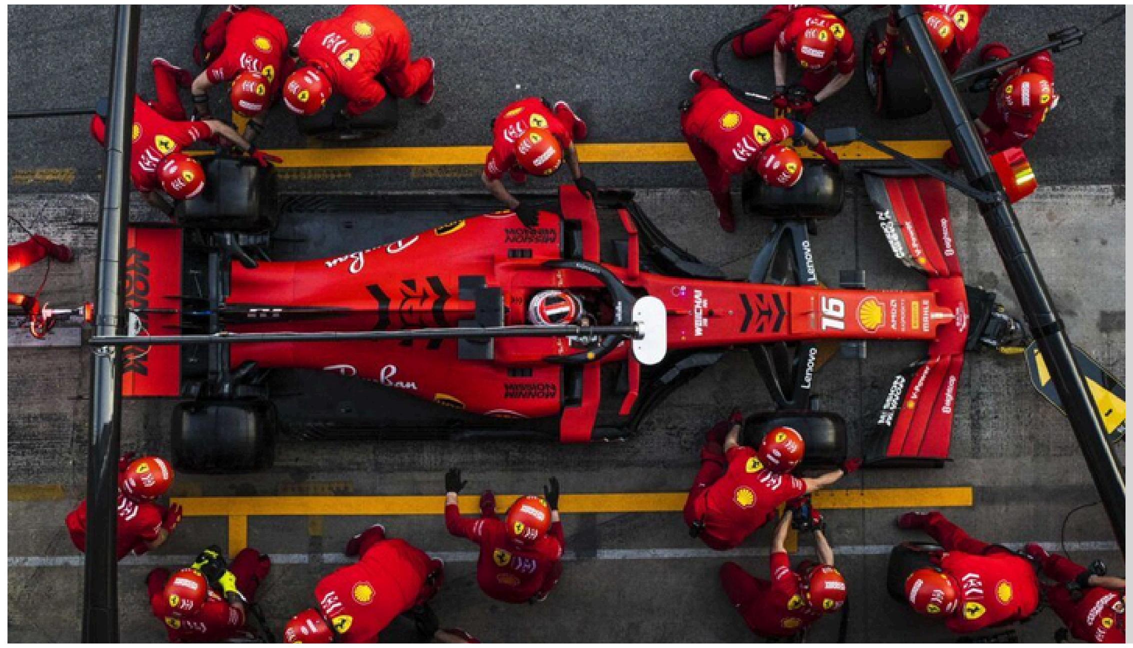
PROJECT OVERVIEW

- **F1 Racing & Technology:** F1 teams analyze massive real-time data from sensors on cars, such as speed, engine performance, and weather conditions, to make critical race decisions and optimize performance.

- **AI & ML in F1:** AI and ML are used to process this data in real-time, helping predict tire degradation, optimize pit stop timing, and simulate race strategies, ultimately giving teams a competitive edge.

- **Relevance to Data Science:** F1 showcases real-world AI/ML applications, from lap time predictions to race outcome simulations. The technology developed has broader impacts beyond motorsports, such as in aerospace and automotive industries.

PROBLEM STATEMENT



PROBLEM DEFINITION

There is a dire need of a powerful tool for fans, analysts, and teams to better understand and anticipate race dynamics, thereby enhancing the overall Formula 1 experience. The potential of AI in complex, data-rich environments like this makes it a great opportunity to deploy working prediction models.

CHALLENGES THAT MAKE THE PROBLEM SIGNIFICANT

Complex Data Environment:

- Telemetry Data: Real-time metrics from car sensors (speed, tire wear, fuel).
- Weather Conditions: Rapid changes affect strategies and performance.
- Driver Performance: Individual behaviors impact race outcomes.

Insufficiency of Traditional Methods:

- Delayed Insights: Slow analysis limits real-time decision-making.
- Inaccurate Predictions: Conventional models struggle with dynamic race conditions.

SIGNIFICANCE OF ADDRESSING THESE CHALLENGES

Enhanced Race Strategies

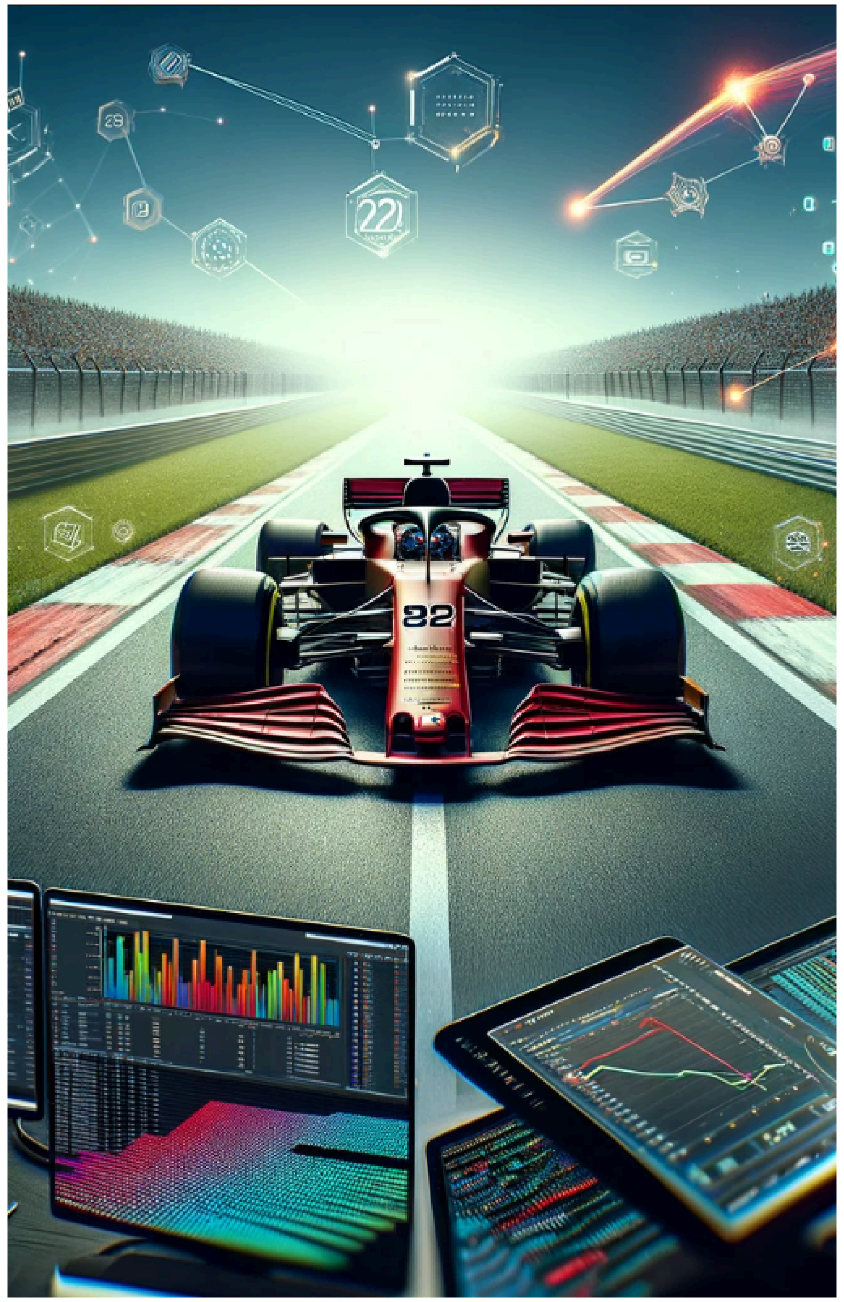
AI and ML enable rapid adjustments based on real-time data, improving strategic decisions.

Optimized Performance

Advanced analytics refine vehicle setups and predict optimal outcomes.

Competitive Edge

Effective use of AI/ML technologies can provide significant advantages in a competitive field.

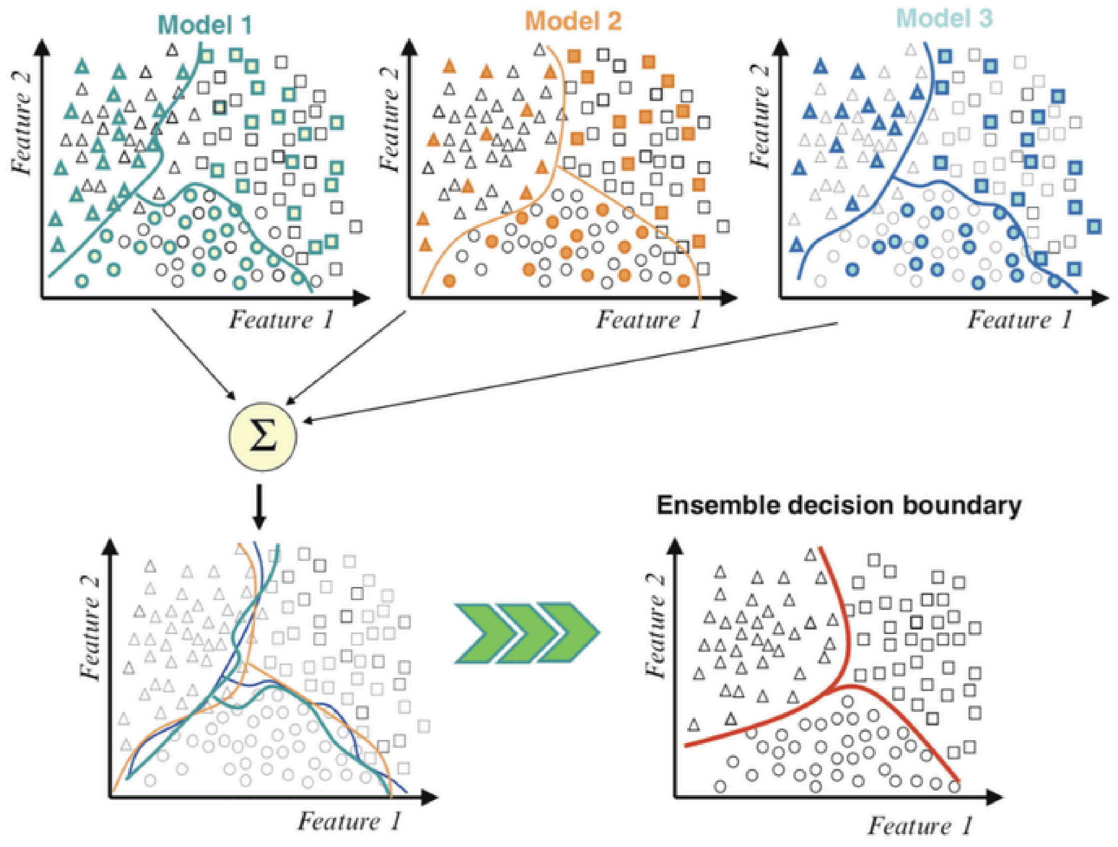


F1NALYZE

MANIPAL UNIVERSITY JAIPUR

OBJECTIVES

(SHORT TERM GOALS)



DATA ANALYSIS

- Utilize AI and ML to analyse historical F1 data (telemetry, race outcomes).
- Identify key performance indicators (KPIs) influencing race success,

MODEL DEVELOPMENT

- Develop initial machine learning models to uncover data patterns.
- Conduct exploratory data analysis (EDA) for visual insights

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

OBJECTIVES

(LONG TERM GOALS)

Improving Race Strategies

Create predictive models for strategy formulation based on historical and real-time data. Analyze past races to derive insights on optimal pit stops and tire choices

Predicting Race Outcomes

Develop algorithms to forecast race results considering various factors. Simulate different scenarios for effective race preparation.

Optimizing Vehicle Performance

Fine-tune vehicle setups using machine learning techniques. Analyze telemetry data to identify performance enhancements.

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

Existing Research

AI and ML
IN MOTORSPORTS

Increasing integration of AI and machine learning to enhance data analysis and strategy in F1. Research shows AI can optimize race strategies and improve performance through data-driven insights.

Telemetry
DATA ANALYSIS

Real-time monitoring of car performance allows immediate adjustments during races. Historical data analysis helps optimize vehicle setups for specific tracks.

Simulations
AND RACE PREDICTIONS

AI-powered simulations predict race outcomes and test strategies pre-race. Machine learning models forecast the impact of variables like weather and tire wear.

GAPS IN EXISTING RESEARCH

Real-Time
DECISION-MAKING

Current models struggle with providing rapid, actionable insights during races.

Advanced
PREDICTION MODEL

Current models struggle with providing rapid, actionable insights during races.

SCOPE OF THE PROJECT : FOCUS AREA

Data Analysis

Analyze historical F1 race data, including:

- Telemetry Data: Insights from car sensors (speed, tire temperature, fuel levels).
- Race Results: Historical performance metrics of teams and drivers.
- Driver Statistics: Individual metrics such as lap times and pit stop efficiency.

Simulation Techniques

Use AI-driven simulations to test race scenarios:

- Predictive Modeling: Forecast race outcomes based on various factors.
- Strategy Testing: Determine optimal approaches for tire changes and pit stops.

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

SCOPE OF THE PROJECT : LIMITATIONS

Data Sources

Focus on publicly available historical data, excluding proprietary team data.

Exclusion of Private Strategies

No access to real-time telemetry or confidential strategies from teams

SCOPE OF THE PROJECT :

EXPECTED OUTCOMES

Insights and Recommendations

Provide actionable insights for improving race strategies and vehicle performance.

Contributions to Research

Add to the knowledge of AI applications in motorsports, highlighting future research areas.

METHODOLOGY OVERVIEW

AI / ML MODELS

Decision Trees

Identify key factors affecting race outcomes

Neural Networks

Capture complex patterns for prediction

Support Vector Machines

Classify performance metrics

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

METHODOLOGY OVERVIEW

APPROACH

Data Preprocessing

Clean data by handling missing values and normalizing features

Training ML Models

Split data into training and testing sets for model training and evaluation

Evaluation

Use cross-validation and hyperparameter tuning to optimize model performance.

F1NALYZE

RESOURCES REQUIRED

Software:

- **Programming Language:**
  - Python: Primary language for data analysis and machine learning.
- **Machine Learning Libraries:**
  - TensorFlow: For building and training neural networks.
  - Scikit-learn: For implementing traditional ML algorithms.
  - Pandas: For data manipulation and analysis.
  - NumPy: For numerical computations.
- **Data Visualization Tools:**
  - Matplotlib & Seaborn: For creating static and attractive visualizations.
  - Plotly: For interactive visualizations.

RESOURCES REQUIRED

Hardware:

- **Computational Resources:**
  - High-Performance Computing (HPC): Necessary for processing large datasets.
  - GPUs: For accelerating deep learning tasks.
- **Cloud Computing Services:**
  - AWS or GCP: For scalable computing and storage solutions.

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

RESOURCES REQUIRED

Data:

- **Historical F1 Data:**
  - Access to race results, driver performance, and telemetry data from reliable sources. (via FastF1
- **Telemetry Data:**
  - Real-time telemetry data for analysis and simulations (if available).
- **Driver Stats & Weather Conditions:**
  - Compilation of driver performance metrics and historical weather data.

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

OUR TIMELINE

Data Collection and Preprocessing

Gather and clean historical F1 data

Data Analysis & Model Selection

Run data analysis techniques on the cleaned data

03

Model training and Prediction

Select and train AI/ML models; tune for performance and predictive analysis

04

Deployment and Testing

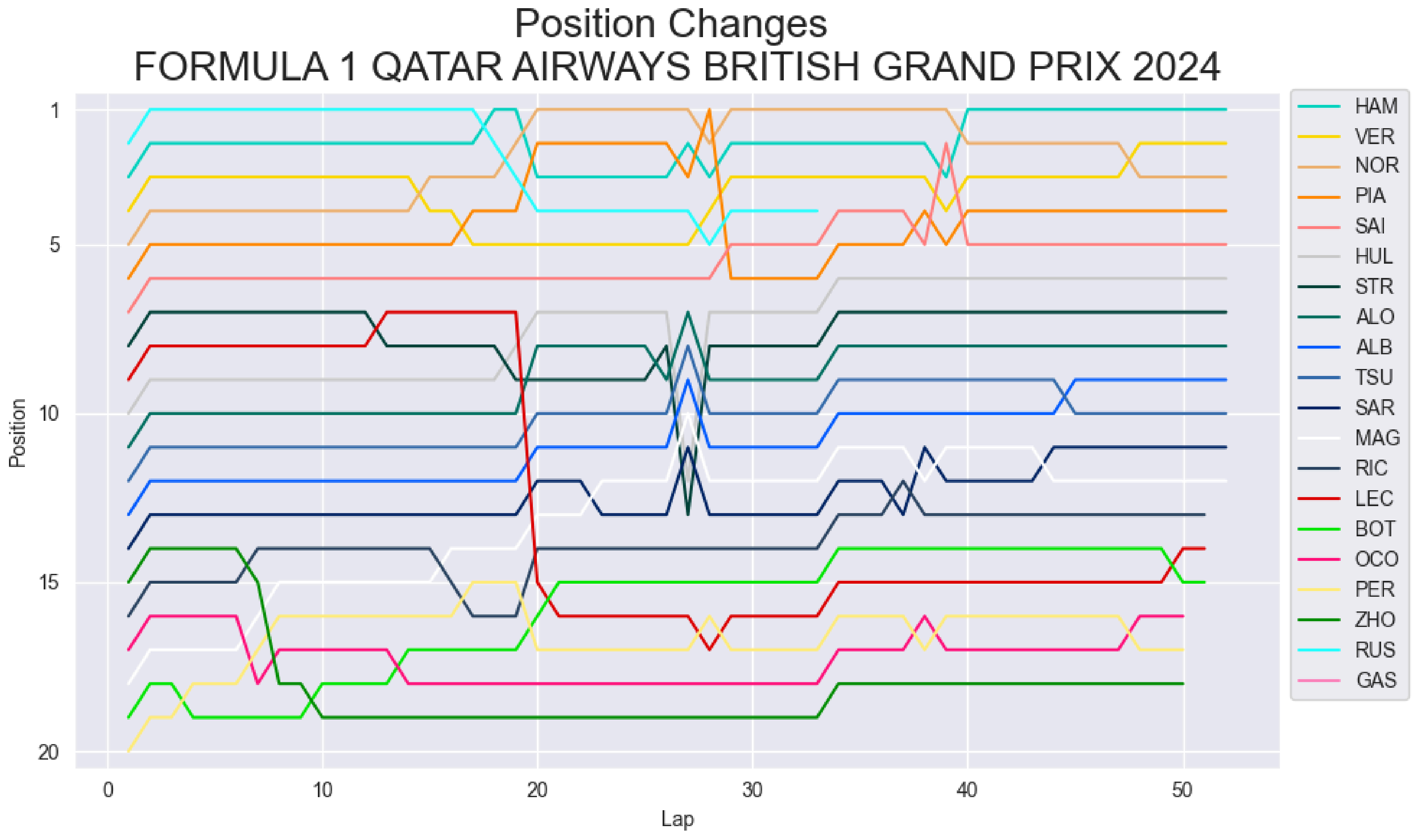
Deploy final model and conduct user testing

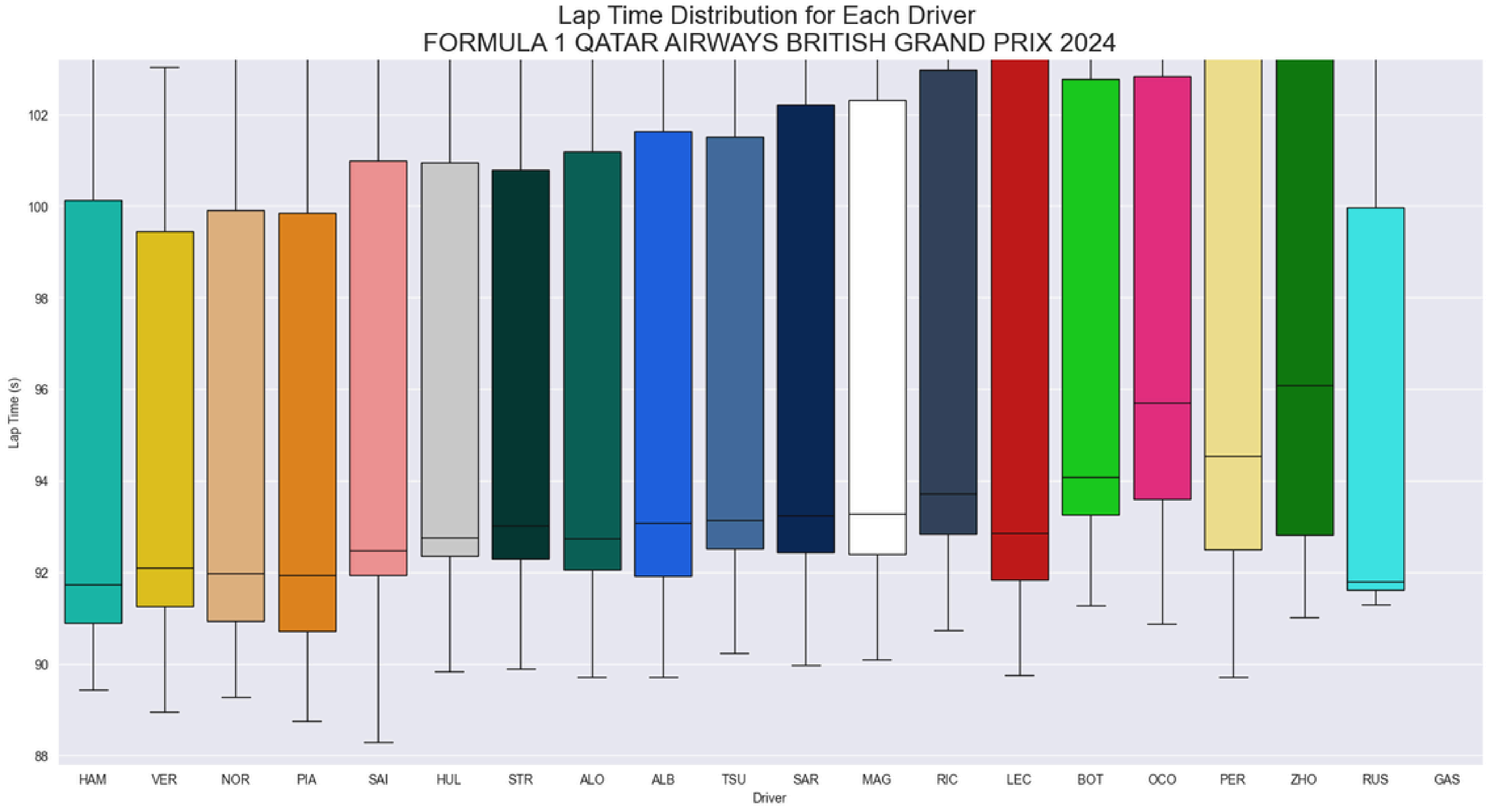
F1NALYZE

MANIPAL UNIVERSITY JAIPUR

ANALYSED DATA

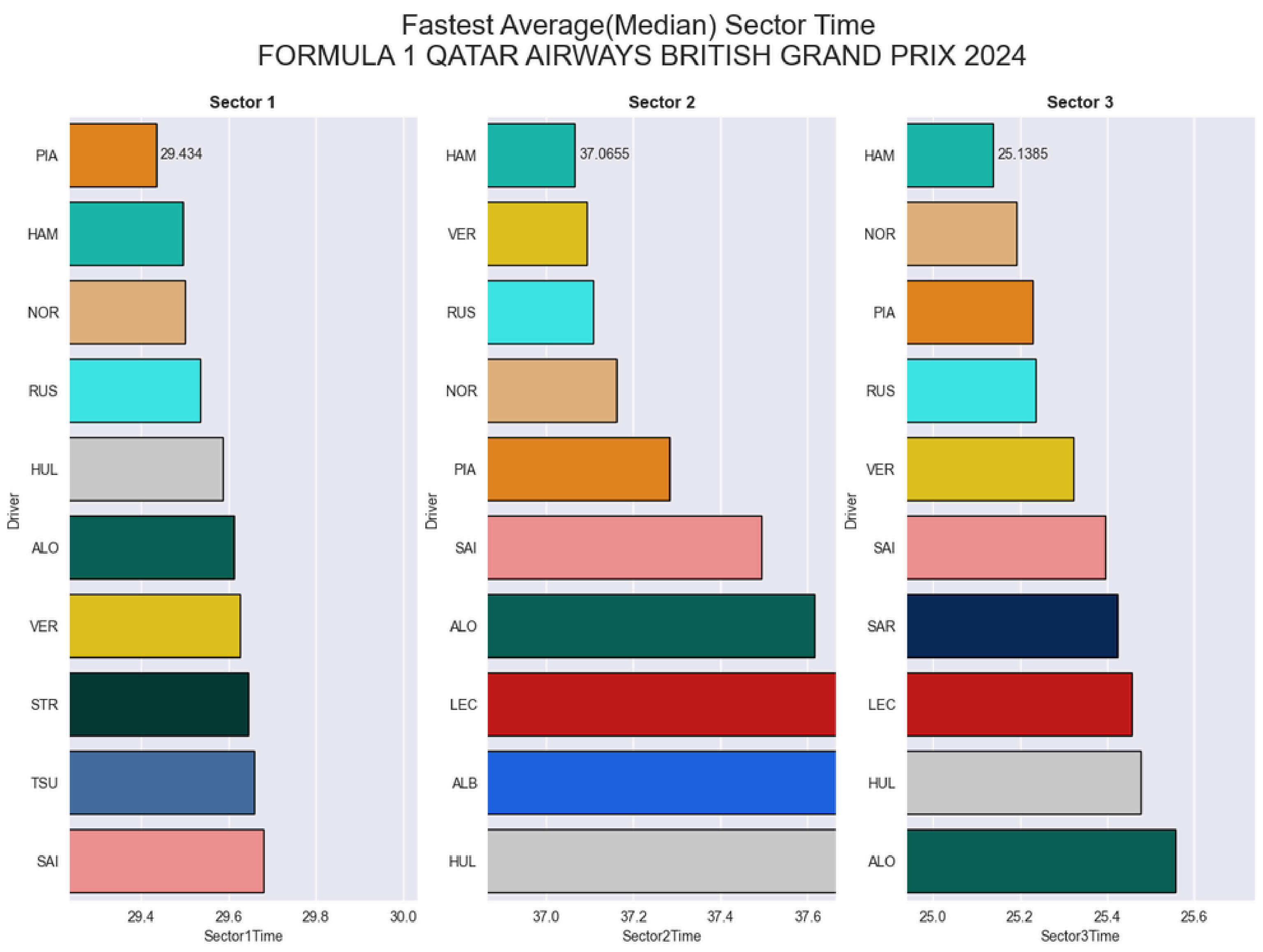
Based on 2024 British Grand Prix Dataset from Kaggle

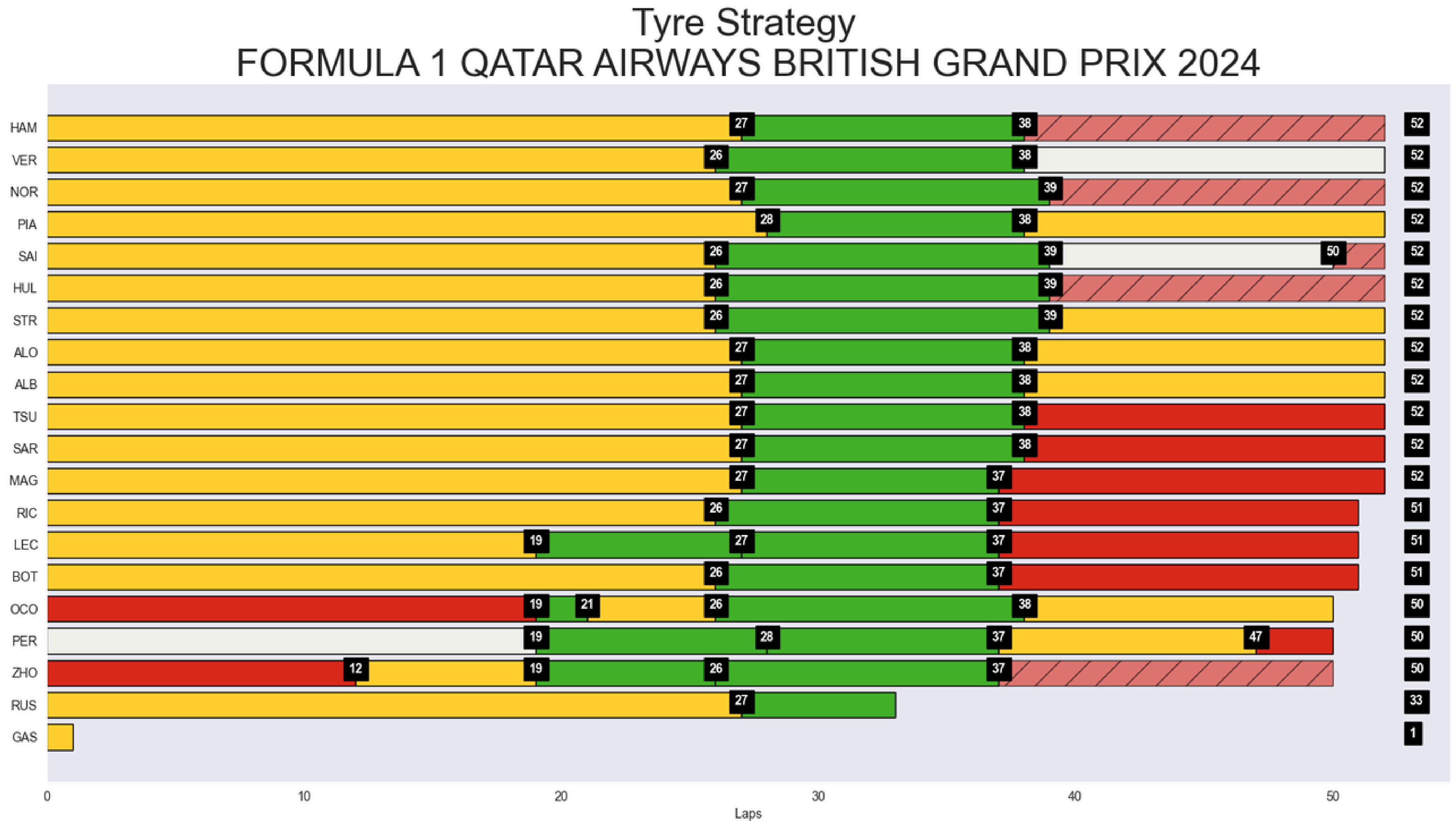




ANALYSED DATA

Based on 2024 British Grand Prix Dataset from Kaggle



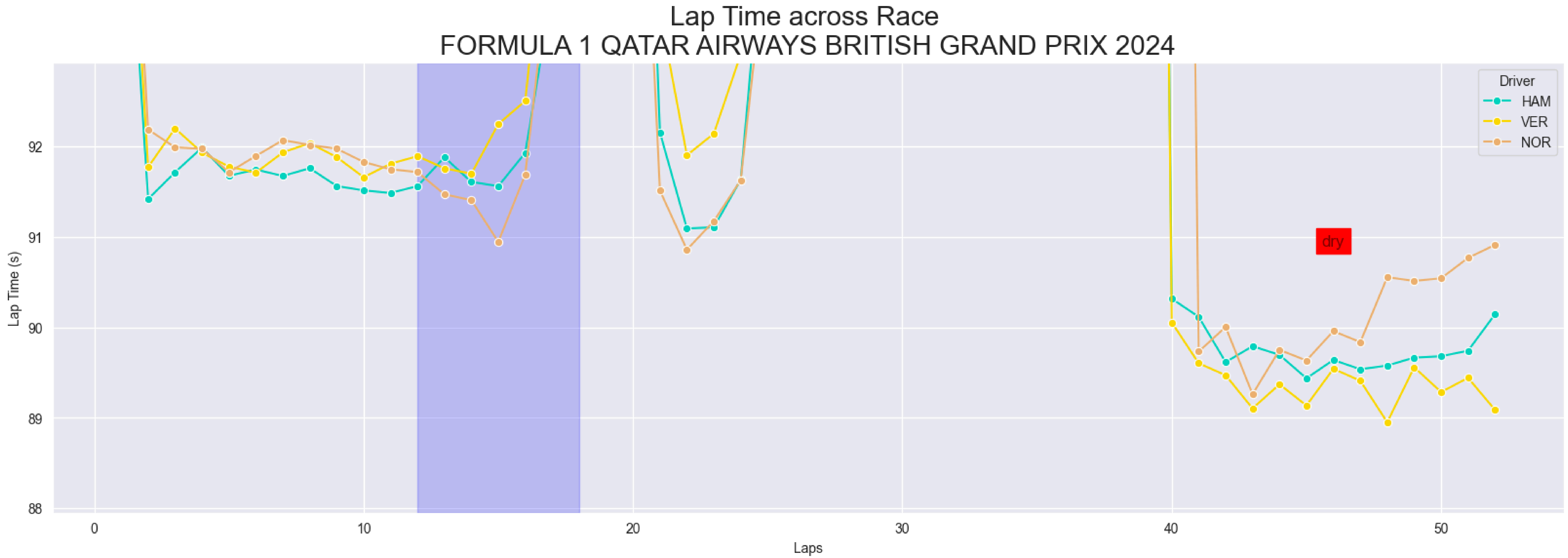


F1NALYZE

MANIPAL UNIVERSITY JAIPUR

ANALYSED DATA

Based on 2024 British Grand Prix Dataset from Kaggle



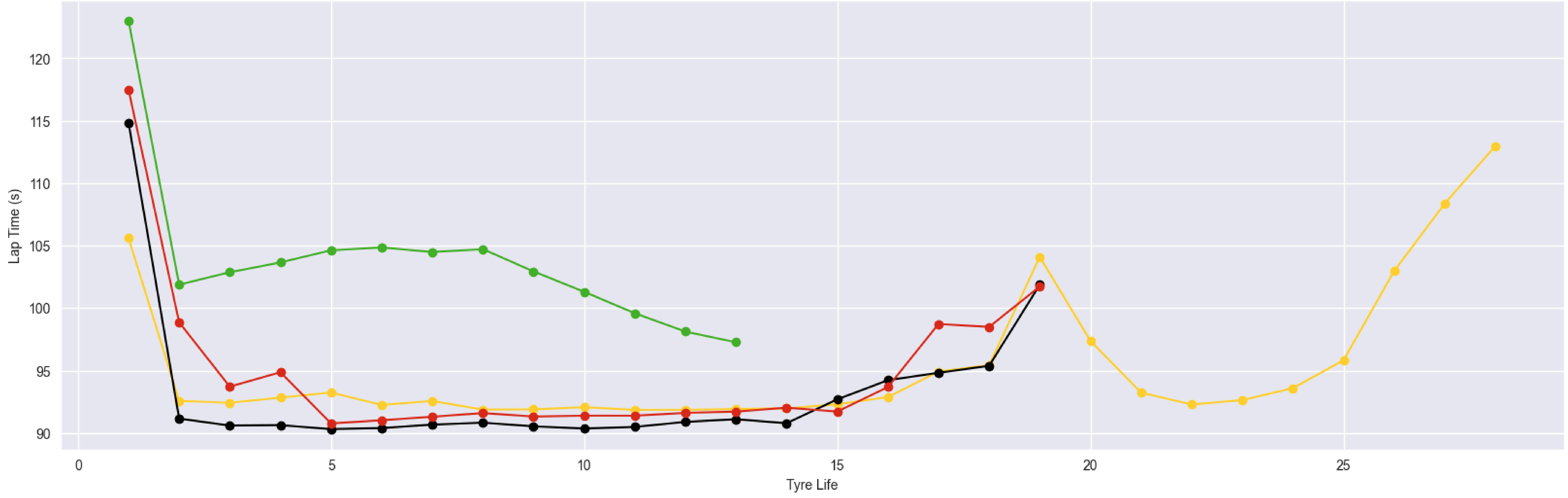
NALYZE

MANIPAL UNIVERSITY JAIPUR

ANALYSED DATA

Based on 2024 British Grand Prix Dataset from Kaggle

Lap Time across Tyre Life
FORMULA 1 QATAR AIRWAYS BRITISH GRAND PRIX 2024

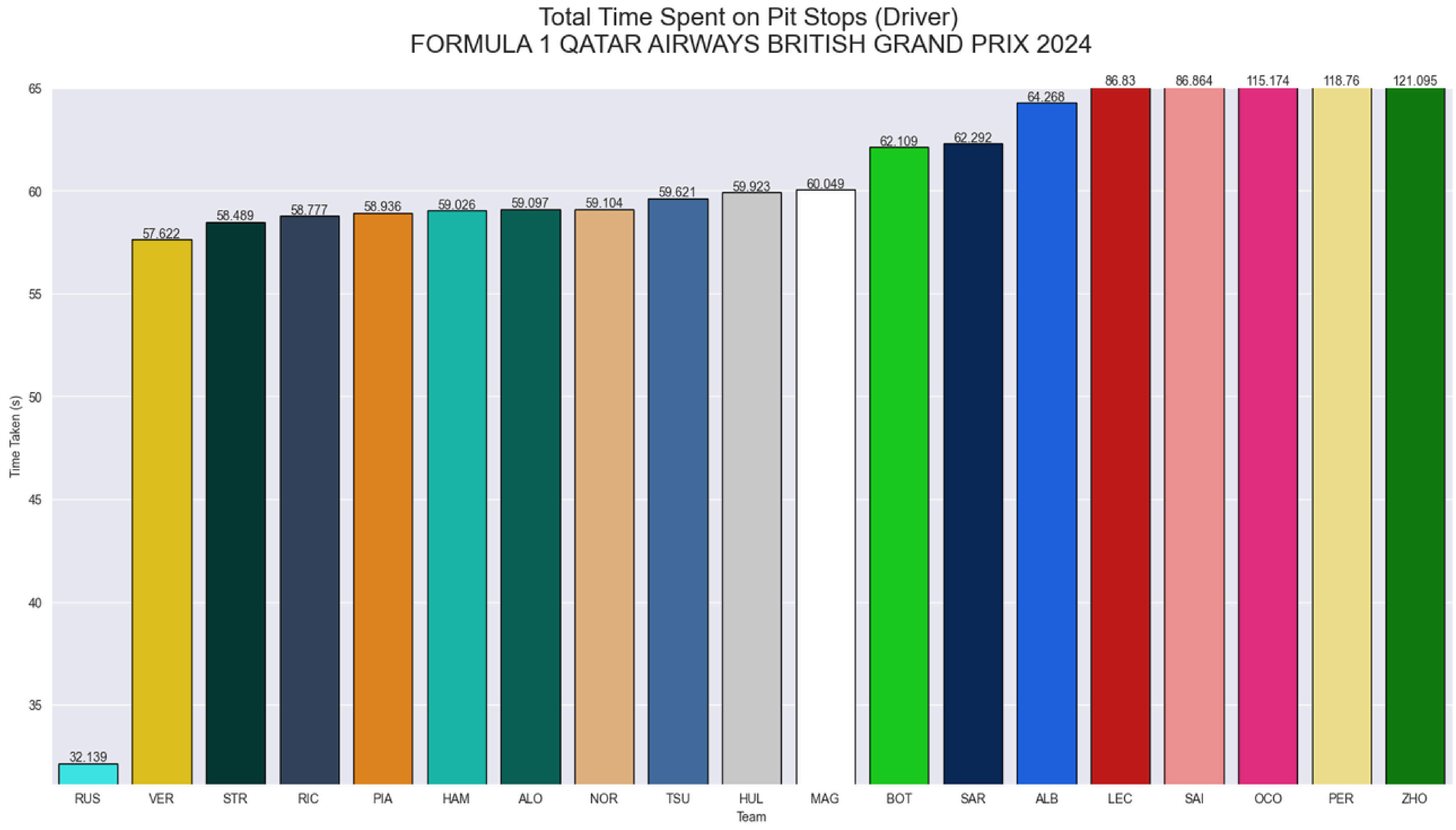


F1NALYZE

MANIPAL UNIVERSITY JAIPUR

ANALYSED DATA

Based on 2024 British Grand Prix Dataset from Kaggle



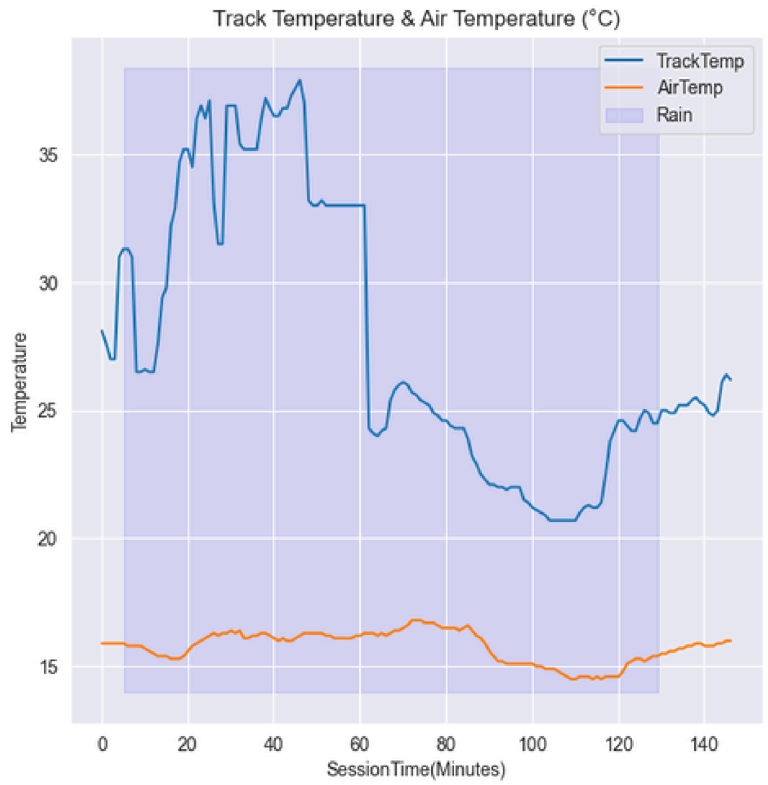
F1NALYZE

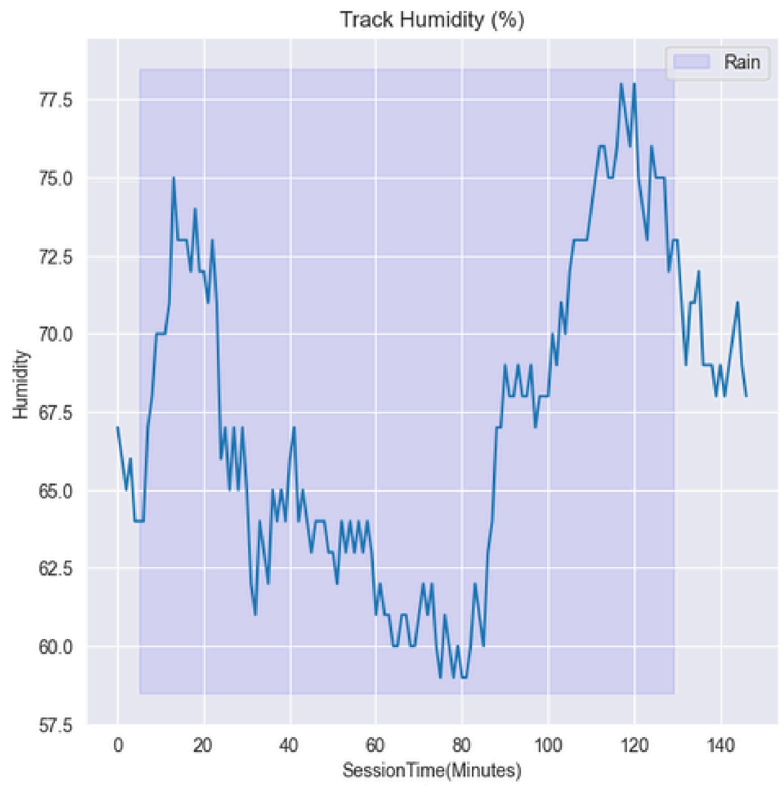
MANIPAL UNIVERSITY JAIPUR

ANALYSED DATA

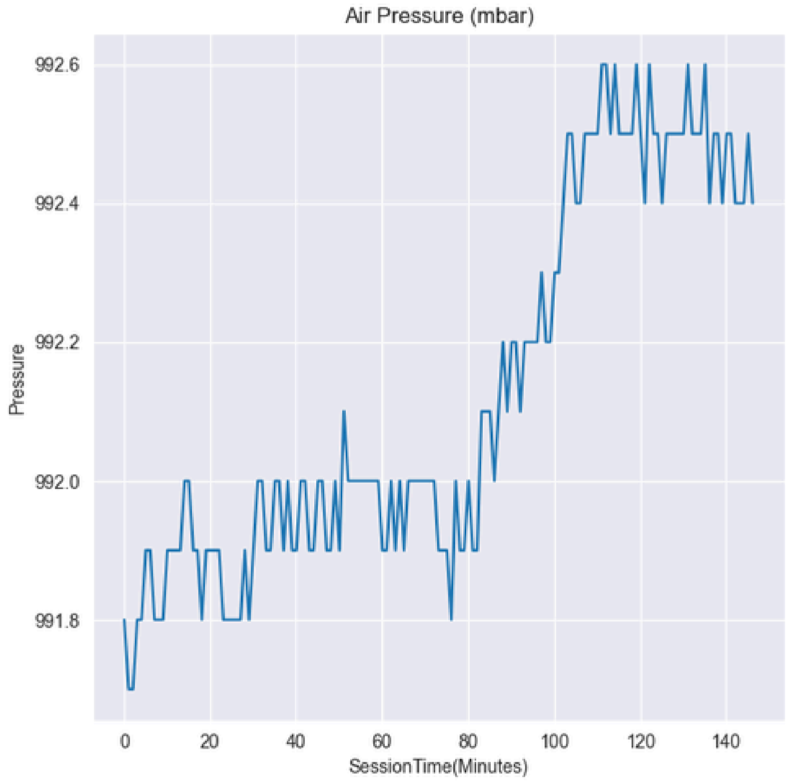
Based on 2024 British Grand Prix Dataset from Kaggle

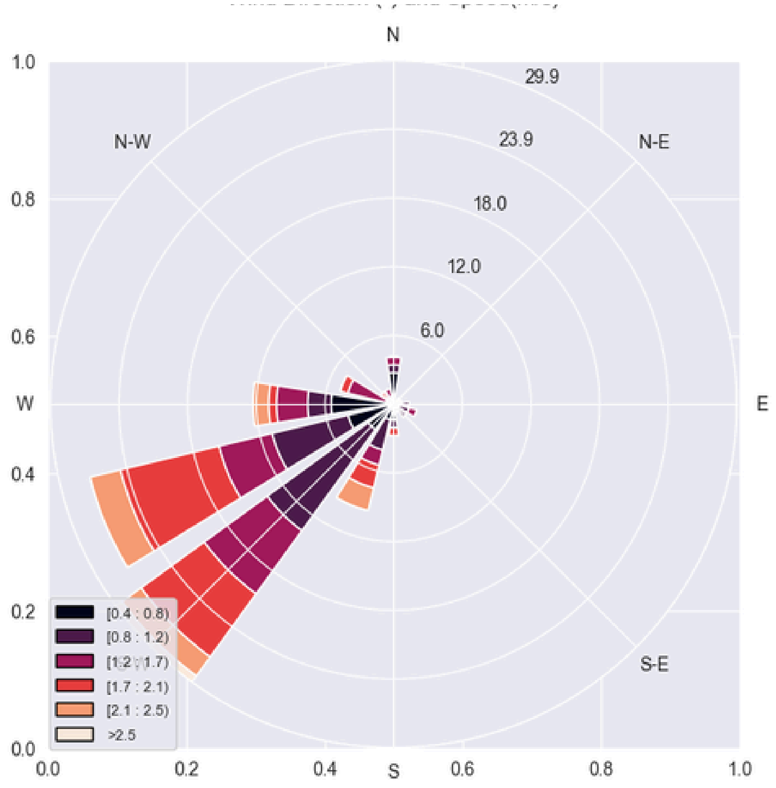
Weather Data & Track Evolution
FORMULA 1 QATAR AIRWAYS BRITISH GRAND PRIX 2024





Wind Direction (°) and Speed(m/s)





F1NALYZE

MANIPAL UNIVERSITY JAIPUR

MODEL SELECTION

Trained on 2024 British Grand Prix Dataset from Kaggle under various conditions and parameters

Ensemble Methods

- **Gradient Boosting Machines** (GBM) like XGBoost, LightGBM, CatBoost: Excellent for handling structured data (driver stats, qualifying times, car performance, track characteristics, weather). They can capture non-linear relationships and interactions well. Often a top performer in these types of prediction tasks.
- **Random Forests:** Also good for structured data, robust to outliers, and relatively easy to tune. Might not capture complex interactions as effectively as GBMs.

Neural Networks

- **Multilayer Perceptrons (MLPs):** Can learn complex relationships, but require careful tuning and a large dataset. Can struggle with interpretability.
- **Recurrent Neural Networks** (RNNs, specifically LSTMs or GRUs): Could be useful for incorporating time-series data like lap times during practice and qualifying, allowing the model to learn patterns and trends leading up to the race.

Hybrid Approaches

- Combining different models (e.g., a GBM for pre-race predictions and an RNN for in-race predictions) can leverage the strengths of each.
- The model can be trained on pre-existing datasets based on past races and then can be upscaled to a real time prediction by setting up a pipeline of data via **FastAPI,** providing real time data analysis with a refresh rate of15-30 seconds.

F1NALYZE

MANIPAL UNIVERSITY JAIPUR

Our Team

ARYAN VERMA

TEAM LEAD

ATISHAY JAIN

MEMBER

Get in Touch

Phone Number

+91 8287044755
+91 8800508168

Email Address

vermaryan1@gmail.com
atishayjain0504@gmail.com