



F1 Pitstop Performance Analysis (2018–2024)

Project Documentation



Table of Contents

1. Project Planning

- Objective
- Key Questions
- Project Phases

2. Stakeholder Analysis

3. Data & Methodology

- Data Source
- Data Dictionary
- Phase 1 Summary: Data Cleaning

4. Modeling, Analysis & Visualization Plan

- Phase 2: Modeling Plan
- Phase 3: Analysis Plan
- Phase 4: Predictive Analysis Plan
- Phase 5: Dashboard Plan

5. Implementation (Source Code)

6. Project Presentation

1. Project Planning

Project Objective

The goal of this project is to analyze Formula 1 pitstop data from the **2018–2024** seasons.

We aim to:

- Identify trends and performance changes over time
- Evaluate team and driver efficiency
- Predict future pitstop performance
- Present insights in an **interactive dashboard**

Key Questions

1. How have average pitstop times evolved from 2018 to 2024?
2. Which constructors and drivers are consistently the fastest or slowest in the pits?
3. Is there a correlation between pitstop performance (time, consistency) and final race position?
4. What impact does tire compound choice have on pit strategy and stint length?
5. Can we predict a driver's average pitstop time using factors such as constructor, circuit, and weather conditions?

Project Phases

Phase	Description	Status
1	Data Cleaning – Acquiring, loading, assessing, and cleaning the raw dataset	Completed
2	Data Modeling – Structuring cleaned data into relational schema for efficient querying	Pending
3	Exploratory & In-Depth Analysis (EDA) – Visualizing and answering key questions	Pending
4	Predictive Analysis – Building ML models for pitstop performance forecasting	Pending
5	Dashboard – Creating an interactive dashboard to present results	Pending

2. Stakeholder Analysis

Stakeholder Type	Description
Primary	Project Instructors / Assessors

3. Data & Methodology

Data Source

- **File:** f1_pitstops_2018_2024.xlsx
- **Description:** Contains detailed records of pitstops, race conditions, and results for each driver in every race between 2018 and 2024.

Data Dictionary

Column Name	Data Type	Description
Season	Integer	Championship year (e.g., 2018)
Round	Integer	Race round number
Circuit	String	Name of the race circuit
Driver	String	Full driver name
Constructor	String	Team name (e.g., Ferrari, Sauber)
Position	Integer	Driver's final finishing position
AvgPitStopTime	Float	Average pitstop time (seconds)
Race Name	String	Official Grand Prix name
Date	Datetime	Date of the race
Air_Temp_C	Float	Air temperature (°C)
Track_Temp_C	Float	Track surface temperature (°C)
Humidity_%	Float	Relative humidity (%)
Wind_Speed_KMH	Float	Wind speed (km/h)
Tire Compound	String	Tire compound (HARD, MEDIUM, SOFT)
Stint	Integer	Stint number (1st, 2nd, etc.)
Pit_Lap	Integer	Lap number of the pitstop
Pit_Time	Float	Duration of the specific pitstop (seconds)

Phase 1 Summary: Data Cleaning

Notebook: F1_Cleaning.ipynb

Key Actions Performed:

- **Handled Missing Values:** Addressed NaN values in columns like Air_Temp_C, Pit_Time, etc.
- **Corrected Data Types:** Converted Date to datetime, Pit_Time to float.
- **Removed Duplicates:** Dropped all fully duplicate rows.
- **Standardized Categories:** Unified categorical data (e.g., merged Ferrari and Scuderia Ferrari → Ferrari).
- **Removed Irrelevant Columns:** Dropped unused or redundant fields.

4. Modeling, Analysis & Visualization Plan

Phase 2: Modeling Plan (to-do)

Objective: Structure the cleaned dataset into a relational schema for analysis.

Plan:

- Design a **Star Schema**:
 - **Fact Table:** Fact_Pitstops (metrics like Pit_Time, Stint Length, Position)
 - **Dimension Tables:**
 - Dim_Race
 - Dim_Driver
 - Dim_Constructor
 - Dim_Conditions
 - Dim_Date
- Establish foreign key relationships for efficient BI integration and querying.

Phase 3: Analysis Plan (to-do)

Objective: Conduct Exploratory and In-Depth Data Analysis (EDA).

Plan:

- **Univariate Analysis:**
 - Distributions of Pit_Time, AvgPitStopTime, and Stint Length.
- **Bivariate / Multivariate Analysis:**
 - AvgPitStopTime by Constructor (Boxplot/Lineplot)
 - AvgPitStopTime by Driver (Bar chart)
 - Correlation heatmap of numeric features (Pit_Time, Air_Temp_C, Position, etc.)
 - Stint Length by Tire Compound (Boxplot)
- **Hypothesis Testing:**
 - Apply **t-tests** or **ANOVA** to check for significant differences in mean pit times between teams.

Phase 4: Predictive Analysis Plan (to-do)

Objective: Build ML models to predict pitstop outcomes.

Plan:

- **Feature Engineering:**

Add features like is_first_lap, weather_condition_category, and team_avg_last_3_races.
- **Model Selection:**
 - **Regression Models:** Linear Regression, Random Forest Regressor
 - **(Optional) Classification Models:** Label pitstops as *fast*, *average*, or *slow*
- **Evaluation Metrics:**
 - Regression → R², RMSE
 - Classification → Accuracy, F1-Score

Phase 5: Dashboard Plan (to-do)

Objective: Present all findings in an interactive dashboard.

Tools: Tableau / Power BI / Streamlit / Dash

Planned Views:

1. **Overview Page:** KPIs (fastest team, avg pit time trends, etc.)
2. **Team & Driver Deep Dive:** Filtered insights per constructor/driver
3. **Race-Level Analysis:** Pitstop effects on race outcomes
4. **Predictive Insights:** Visualize ML predictions (e.g., *Predicted Ferrari pit time for Monza GP*)

5. Implementation (Source Code)

Phase	File / Notebook
Phase 1 (Cleaning)	[Github]
Phase 2 (Modeling)	[Link to be added]
Phase 3 (Analysis)	[Link to be added]
Phase 4 (Predictive Analysis)	[Link to be added]
Phase 5 (Dashboard)	[Link to be added]
GitHub Repository	[Github]

6. Project Presentation

Deliverable	Link
Final Report (PDF)	[To be added]
Video Demonstration (Optional)	[To be added]