FACTORIAL PERFORMANCE ANALYSIS WITH ML-ENHANCED CATEGORIZATION: K-MEANS CLUSTERING OF F1 CORNER DYNAMICS (LEC vs. VER, 2025)

Emanuele Russo

December 2025

1 Aim of the Document

This research aims to conduct a rigorous, factorial performance analysis comparing the technical capabilities of the cars driven by Charles Leclerc and Max Verstappen throughout the 2025 Formula 1 Qualifying sessions.
The notebook with the analysis can be found at: https://github.com/emarussoo/F1_telemetry_analysis.
The core methodological objective is to employ Unsupervised Machine Learning, specifically K-Means Clustering, to objectively segment the diverse set of track corners into four distinct dynamic categories (Slow, Medium, High-Speed, Power). This ML-enhanced categorization allows for a scientifically sound comparison of the two machines' respective strengths and weaknesses across critical performance factors (Speed, Traction, Braking), ensuring that conclusions are derived from statistically pure, category-specific data.

2 Used tools

The data acquisition, processing, and statistical modeling utilized a specialized Python-based environment:

- **Data Acquisition:** The `fastf1` library was employed to download official Formula 1 telemetry and timing data for all Qualifying sessions of the 2025 season.
- **Data Processing and Engineering:** The `pandas` and `numpy` libraries were used for data manipulation, cleaning, and calculating custom metrics within defined spatial windows.
- **Apex Detection:** The `scipy.signal.find_peaks` function was instrumental in identifying objective minimum speed points, which define the empirical corner apex.
- **Statistical Modeling:** The `sklearn` library, specifically `KMeans`, was used for unsupervised machine learning to classify corners into distinct physical categories (clusters).

3 Introduction to Factorial Analysis

3.1 Rationale for Corner Selection

In Formula 1, performance is predominantly won or lost during the cornering phase. Analyzing the total lap time can mask specific car characteristics. Therefore, this study isolates and aggregates data points taken exclusively at the true performance limit of the car. The dataset comprises over 280 specific corner instances across the season. This granular focus ensures that the conclusions directly relate to the car's dynamic behavior, not external factors like straight-line speed (which is heavily influenced by engine mode and drag reduction systems).

3.2 Detailed Analysis Parameters

The comparison is based on the delta between LEC and VER for three standardized engineering parameters, measured in a spatial window of $\pm\mathbf{50}$ meters around the minimum speed point:

1. **Delta Minimum Speed ($\mathbf{V_{min}}$) [km/h]:** Measures the car's ability to maintain energy through the corner apex. This is the primary indicator of chassis balance, mechanical grip, and aerodynamic load (Handling/Carico Aerodinamico).
2. **Delta Traction Score (% Throttle) [%]:** Represents the average throttle pedal application percentage on corner exit. This is the direct measure of mechanical grip, differential setup, and rear-axle stability under power deployment (Traction).
3. **Delta Braking ($\mathbf{\Delta V}$)** [km/h]: Represents the difference in the velocity drop (deceleration intensity) between the start of the braking zone ($\approx$ 50m before V_{min}) and the apex. This quantifies the driver's braking aggression and the car's stability during corner entry (Braking/Staccata).

4 Design Choices and Methodological Integrity

4.1 Rejection of `FastF1` Geometric Corners

A deliberate design choice was made to **not** utilize pre-defined corner sections or geometric markers provided by systems like `FastF1`.

- **Objective Definition of Apex:** The true performance apex, where the car reaches its minimum velocity and transitions to acceleration, is an *empirical* point determined by the driver's inputs and the car's limits, not by the track's fixed geometry.
- **Filtering Trivial Events:** Geometric markers include minor kinks or curves taken flat out. A custom `scipy.signal.find_peaks` algorithm was implemented on the inverted speed trace ($-V$) to ensure only velocity drops with a **prominence of 15** and a minimum separation distance of **50** meters were logged. This ensures the dataset only contains performance-critical corners that demand significant deceleration or steering input.

4.2 The Role of Corner Clustering (K-Means) and K Selection

The $K = 4$ clustering method was applied to the combined dataset of V_{min}, Traction, and Braking Delta. The selection of $K = 4$ was a deliberate engineering choice, overriding simpler heuristics.

- **Rejection of Low K Values (e.g., K = 2):** A low value of $K = 2$ (e.g., often suggested by the Elbow Method) is statistically insufficient for a nuanced performance breakdown, as it only distinguishes broadly between 'Slow' and 'Fast' regimes. This fails to separate the distinct physical mechanisms of performance.
- **Engineering Necessity of K = 4:** The choice of $K = 4$ is methodologically required to isolate the three distinct and performance-critical regimes necessary for factorial analysis, plus the straight-line segment:
  1. Mechanical Grip (Cluster 1: Slow Corners - Low V_{min}, high ΔV)
  2. Chassis Handling (Cluster 2: Medium Corners - Mid-range V_{min})
  3. Aerodynamic Stability (Cluster 3: High-Speed Corners - High V_{min})
- **Normalization and Labeling:** This clustering provides a robust basis for comparing physically similar cornering events across different circuits, labeled coherently based on their mean minimum speed profile.

5 Dataset Cleaning: Outlier Rejection

5.1 Justification for Discarding Data

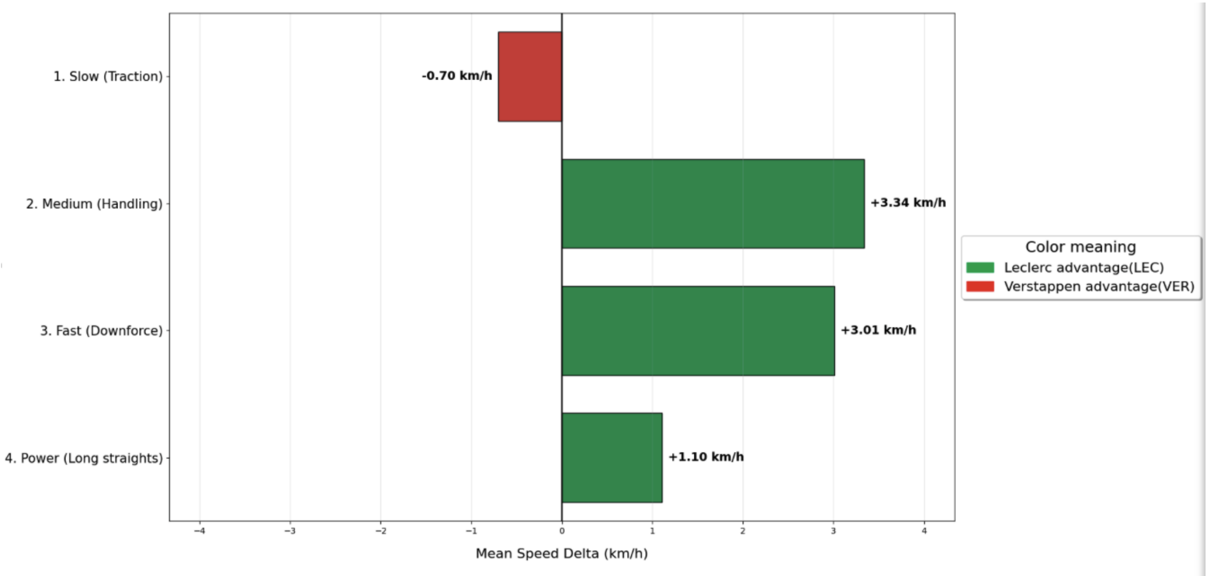
Before calculating the final means, a critical data cleaning phase was executed. The analysis of raw telemetry often includes data points where one driver was not pushing to the limit (e.g., in-lap, warm-up lap, yellow flags, traffic, or major mistakes). These "outliers" artificially inflate the mean delta, leading to misleading conclusions.

5.2 Implementation and Impact of the Filter

- **Filter Definition:** We used the granular, pre-aggregated data (Circuit $\times$ Cluster) and calculated the mean $\Delta V_{\min}$ for each combination. Any combination exceeding a difference of $\pm$**15** km/h was flagged as an outlier and **physically removed** from the final aggregation.
- **Quantitative Impact (LEC vs VER):** The unfiltered dataset contained 62 valid (LEC $\cap$ VER) combinations. The outlier filter removed 17 combinations (approx. 21%) that were statistically anomalous. The final Heatmap is based on **45 clean comparisons**.

6 Cleaned Mean Speed Delta by Performance Cluster

This graph shows the difference of the apex min speed between Charles Leclerc and Max Verstappen.



- **X-Axis (Mean Speed Delta in km/h):** Represents the velocity difference at the corner apex ($V_{\min}$). Positive values indicate a Ferrari (Leclerc) advantage, while negative values indicate a Red Bull (Verstappen) advantage.
- **Y-Axis (Corner Clusters):** Corners are classified into four tiers: Slow (Traction), Medium (Handling), Fast (Downforce), and Power (Long straights).

6.1 Data Interpretation by Cluster

Slow (Traction): $-0.70 km/h$ in low-speed sections where mechanical grip is dominant, the Red Bull holds a slight advantage. Verstappen is able to navigate these tight corners 0.70 km/h faster at the apex than Leclerc, suggesting a more stable front-end or better rotation in high-steering-angle scenarios.

Medium (Handling): $+3.34$ km/h This is the Ferrari's strongest performance regime. In technical sections requiring mid-speed handling and chassis balance, Leclerc is significantly faster ($+3.34$ km/h). This indicates that the SF-24 chassis is exceptionally well-balanced in the 100–180 km/h speed range.

Fast (Downforce): +3.01 km/h In high-speed corners where aerodynamic load (downforce) is the limiting factor, Ferrari maintains a dominant lead. Leclerc carries over 3 km/h more speed through these sections, proving the high aerodynamic efficiency and peak downforce levels of the car under extreme load.

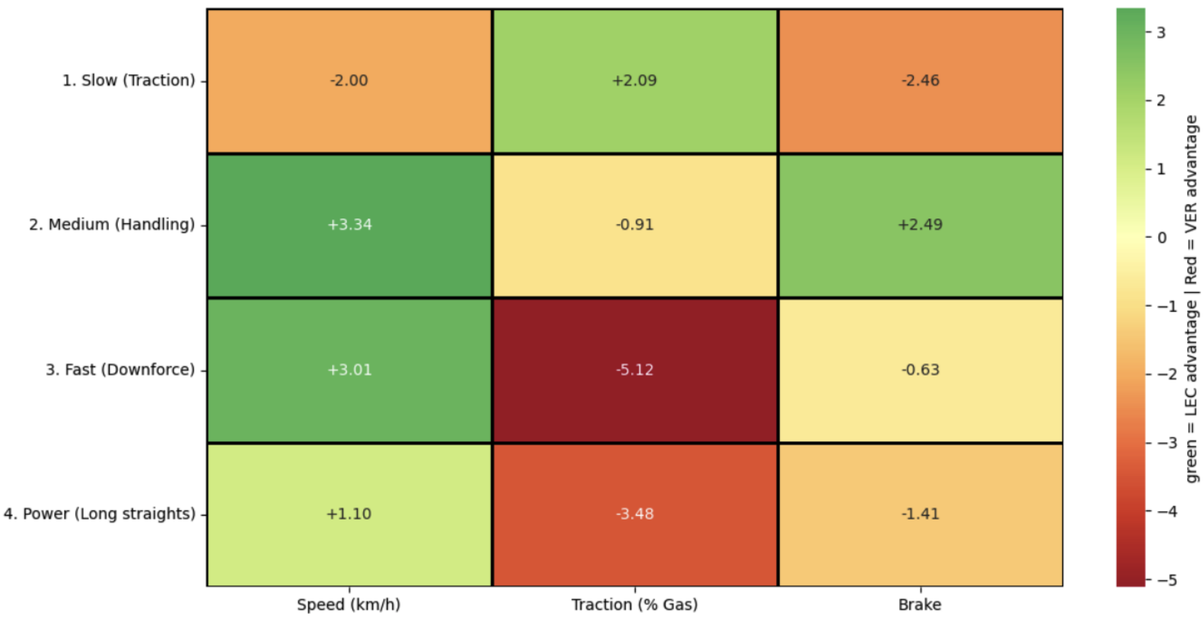
Power (Long straights): +1.10 km/h In "Power" sections—which include high-speed kinks and exits onto long straights—Leclerc remains faster at the minimum speed point by 1.10 km/h. This suggests that the Ferrari engine/ERS deployment allows for slightly higher speeds even in sectors traditionally dominated by drag efficiency.

6.2 Methodological Note (Outlier Removal)

It is important to note that this is a "Real Analysis". This means the dataset has been filtered to remove "noisy" data (outliers). Specifically, any corner where the delta exceeded 15 km/h was discarded to ensure the results reflect pure qualifying performance rather than external factors like traffic, yellow flags, or driver errors.

7 Comparative telemetry heatmap analysis and driving style footprint

This figure presents the final Heatmap, derived exclusively from the **45** validated combinations. This result accurately reflects the pure performance characteristics of the two cars.



7.1 Cluster-by-Cluster Analysis

7.1.1 Cluster 1: Slow (Traction)

Focuses on mechanical grip and low-speed rotation.

- **Speed (-2.00):** Verstappen carries higher V_{min}, suggesting a superior front-end bite during low-speed rotation.
- **Traction (+2.09):** Leclerc excels in power delivery on exit, utilizing rear stability to find grip earlier than the Red Bull.
- **Braking (-2.46):** A significant advantage for Verstappen, who utilizes aggressive trail-braking to rotate the car.

7.1.2 Cluster 2: Medium (Handling)

Technical sectors requiring chassis balance and high lateral G transitions.

- **Speed (+3.34):** Leclerc's strongest area. He maintains a much higher average speed through technical sequences.
- **Braking (+2.49):** Leclerc is more efficient in modulating the brakes to maintain platform stability in mid-speed entries.
- **Traction (-0.91):** Marginal advantage for Verstappen, though largely offset by Leclerc's mid-corner speed.

7.1.3 Cluster 3: Fast (Downforce)

High-speed corners where aerodynamic efficiency and stability are paramount.

- **Speed (+3.01):** The Ferrari platform maintains superior momentum under high aero-load.
- **Traction (-5.12): Critical Delta.** Verstappen is able to reach 100% throttle significantly earlier, suggesting a more stable aero-platform during corner exit.
- **Braking (-0.63):** Minimal advantage for Verstappen on high-speed entries.

7.1.4 Cluster 4: Power (Long Straights)

Evaluation of top speed and ERS deployment efficiency.

- **Speed (+1.10):** Leclerc holds a slight advantage in V-max, indicating a lower drag coefficient or better ERS clipping management.
- **Traction (-3.48):** Verstappen recovers time through superior initial acceleration out of the preceding corners.
- **Braking (-1.41):** Verstappen is more effective in heavy braking zones at the end of long straights.

8 Final Technical Synthesis

The telemetry comparison highlights a clear divergence in vehicle utilization.
Leclerc operates as a *Momentum Driver*, extracting maximum performance during the steady-state phase of the corner (high lateral G). In contrast, **Verstappen** functions as a *Transition Specialist*, gaining his lap time during the longitudinal shifts—braking and rapid throttle application.

The **-5.12 delta in Fast Traction** is the defining metric of this analysis: it suggests that Verstappen's car platform achieves aerodynamic re-attachment or stability significantly faster than Leclerc's, allowing for a more violent and early return to full throttle.

9 Formula1 in 2026: Previsions

The 2026 regulations move away from "aero-efficiency" toward "energy-management efficiency." While Leclerc will likely remain the king of Qualifying and high-speed flow due to his unmatched mid-corner speed, the 2026 race dynamics will favor Verstappen's profile. The increased reliance on traction and braking-based recovery plays directly into the areas where Max currently holds his biggest telemetry leads.

However, if Ferrari provides a Power Unit with superior thermal efficiency to compensate for Leclerc's "Speed Carrying" style, the LEC/VER rivalry will shift from an aerodynamic battle to a software and energy-deployment war.