Can Unsupervised Learning Methods Reveal Distinct Driving Styles Among Formula 1(F1) Drivers?

Antony Taylor and Emanuele Lindo Secco*

Robotics Laboratory, School of Mathematics, Computer Science and the Environment, Liverpool Hope University, Liverpool, UK
Email: 22004909@hope.ac.uk (A.T.); seccoe@hope.ac.uk (E.L.S.)
*Corresponding author
Manuscript received March 4, 2026; accepted June 2, 2026; published July 31, 2026

***Abstract*—This paper investigates whether unsupervised machine learning techniques can identify meaningful and interpretable driving style clusters among Formula 1 (F1) drivers, using telemetry and lap-time data extracted via the FastF1 Python library applied to the data of the 2023 F1 Season. Four clustering algorithms are implemented and compared: K-Means, Hierarchical Agglomerative Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Gaussian Mixture Models (GMM). A suite of performance features is engineered from raw race telemetry, including relative lap time, lap-time consistency (measured via the Median Absolute Deviation), tyre degradation slope, overtaking efficiency, and intra-team relative sector differentials. Car performance confounders are partially mitigated by computing intra-team relative metrics throughout the entire feature set. A Variance Inflation Factor (VIF) analysis is applied to identify and remove highly collinear features prior to clustering. Principal Component Analysis (PCA) is employed for dimensionality reduction and visualization.**

Results suggest that between 3 and 5 statistically meaningful driver clusters emerge, broadly characterized as elite performers, aggressive racers, precision drivers, and solid midfield operators. K-Means with $k = 4$ and Ward-linkage Hierarchical Clustering yield the most internally consistent groupings, outperforming DBSCAN and GMM on silhouette and Davies-Bouldin scores. A supplementary supervised validation experiment confirms that the engineered feature set captures genuinely driver-discriminative signal, with a Random Forest classifier achieving 78.3% accuracy in predicting driver identity under 5-fold cross-validation. Unsupervised learning can surface interpretable driving style profiles from Formula 1 data, though the explanatory power of these clusters is constrained by data limitations including the small per season sample size, incomplete telemetry coverage and the dominance of car performance over driver skill in raw lap-time metrics.

***Keywords*—unsupervised learning, clustering, Formula 1 (F1), driving styles, K-Means, gaussian mixture models, telemetry, Principal Component Analysis (PCA), FastF1**

I. Introduction

Formula 1 (F1) is the pinnacle of single-seater motorsport, combining engineering excellence with elite athletic performance. For decades, the analysis of driver performance has been dominated by outcome-based metrics such as championship points, race wins, and qualifying positions. However, these metrics are fundamentally confounded by the aerodynamic, mechanical, and powertrain advantages conferred by the competing constructors, making it exceedingly difficult to isolate the contribution of an individual driver's skill and style from the capabilities of their machinery [1].

The application of machine learning to motor sports performance analysis has grown substantially over the past decade. Early work focused primarily on predictive modelling of race outcomes: Tulabandhula and Rudin [2] trained ensemble models on historical race data to predict podium finishes, achieving accuracy competitive with expert forecasters. Heilmeier *et al.* [3] developed an optimal race strategy simulator using Monte Carlo methods and reinforcement learning, demonstrating that data-driven strategy models can outperform conventional heuristic approaches in terms of expected points scored.

Supervised approaches have also been applied to driver performance. Lapkowski *et al*. [4] trained gradient-boosted classifiers on telemetry features to distinguish driver identities from anonymous lap data, achieving identification accuracy above 80% across the top ten championship contenders in a single season. This finding is particularly pertinent to the present study, as it provides strong evidence that drivers exhibit statistically distinguishable behavioural signatures—a necessary precondition for meaningful clustering.

Work explicitly focused on driving style identification in Formula 1 (F1) is sparser but growing. Filipe et al proposed a framework for characterizing driving behavior in road vehicles using spectral analysis of throttle and steering inputs, finding that gaussian mixture models captured inter-driver variability better than K-Means in their dataset [5]. Luzardo *et al.* applied hierarchical clustering to on-board sensor data from professional GT racing drivers and identified three stable clusters loosely corresponding to aggressive, smooth, and strategic driving profiles [6].

Within the open-source F1 analytics community, significant work has been performed using FastF1 and similar data pipelines. Kesteren and Bergkamp [7] used K-Means clustering on sector-time differentials to identify driver-circuit compatibility patterns, finding that certain drivers consistently outperform teammates at high-downforce circuits regardless of overall championship standing. These findings motivate a broader, season-level approach as pursued in this paper.

The emergence of modern telemetry systems and open-access data platforms has created an unprecedented opportunity to examine driver behavior at a granular level. The FastF1 Python library, for instance, provides access to official Formula 1 session data including lap times, sector times, tyre compound data, and positional information. When processed appropriately, this data can form the basis for data-driven analyses of driver behavior that are at least partially decoupled from car performance.

This paper asks a deceptively simple question: can unsupervised machine learning methods reveal distinct

driving styles among Formula 1 drivers? The hypothesis is that drivers who operate in the same physical environment—the same circuits, the same regulations, the same conditions—will nonetheless exhibit statistically distinguishable patterns of behavior. These patterns, if they are sufficiently stable and interpretable, constitute driving styles: coherent, reproducible approaches to the task of racing that reflect individual differences in technique, risk tolerance, physical capability, and tactical philosophy.

This work is explicitly framed as a descriptive case study of the 2023 F1 Season ($n = 20$ drivers). No claim is made that the resulting taxonomy generalizes beyond this specific season or regulatory context; cross-season generalization is identified as a priority for future work.

This work is therefore motivated by several converging interests:

1) From a sporting analytics perspective, identifying driving styles could inform team strategy, driver academy development, driver-constructor matching decisions, and broadcast analysis.
2) From a machine learning perspective, the F1 domain presents a rich and underexplored testbed for unsupervised methods: the data is high-dimensional, temporally structured, contains natural confounders, and admits meaningful post-hoc interpretation. The absence of ground-truth labels (there is no universally accepted taxonomy of F1 driving styles) makes the problem inherently unsupervised.
3) To address this question, the study implements 4 unsupervised clustering algorithms—namely (1) K-Means, (2) Hierarchical Agglomerative Clustering, (3) Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and (4) Gaussian Mixture Models—applied to a feature matrix derived from the 2023 F1 Season.

Features are engineered to capture pace relative to teammates, lap-time consistency, tyre degradation, overtaking and defensive behaviour, and intra-team relative performance. A VIF analysis identifies and removes collinear features prior to clustering. Principal Component Analysis (PCA) is used for visualization. Clusters are evaluated using standard internal validity indices and interpreted qualitatively in the context of known driver characteristics. A supplementary supervised experiment validates the discriminative power of the feature set. Unsupervised clustering has previously been applied to performance segmentation in cycling, football, and other sports domains; this study extends that body of work to the domain of motorsport.

II. Materials and Methods

This Section describes the complete analytical pipeline of this work, which was implemented in the Python codebase, moving from the raw data acquisition through the feature engineering, normalization, clustering, and evaluation. The pipeline is encapsulated in an F1 Driver Analyzer class, which provides modular methods for each analytical stage. All code is written in Python 3.10 and relies on the scientific Python ecosystem alongside the FastF1 telemetry library.

A. Data Acquisition and Preprocessing

Race session data are retrieved using the FastF1 library (version 3.x), which interfaces with the official F1 data feed and caches result locally to reduce the Application Programming Interface (API) load. For each race in the selected season, the library returns a Data Frame of lap-level records containing the driver identifier, team identifier, lap time, sector 1–3 times, tyre compound, pit in/out timestamps, and track position at the start of each lap. A set of parameters are also defined according to Table 1.

Table 1. Engineered features used in the clustering analysis. All sector features are expressed as intra-team relative metrics

Feature	Description	Motivation
Relative Lap Time	Driver means race pace minus team mean pace (s)	Controls for car advantage; captures intra-team pace delta
Consistency MAD	Median Absolute Deviation (MAD) of clean lap times (s)	Robust consistency measure; resistant to outliers
Lap Time Std.	Standard deviation of clean lap times (s)	Classic consistency measure; included for comparison
Tyre Degradation	Slope of Ordinary Least Squares (OLS) regression of lap time on lap number	Captures rate of tyre wear and driving smoothness
Overtake Efficiency	Positions gained per lap across season	Proxy for aggressive race craft and overtaking ability
DefenseRate	Positions lost per lap across season	Proxy for defensive driving and vulnerability under pressure
Relative Consistency	Driver consistency MAD minus team mean Consistency MAD	Intra-team normalised consistency
Relative Sector1	Driver means sector 1 time (s) team sector time (s)	Car-independent high-speed braking zone performance
Relative Sector2	Driver means sector 1 time (s) team sector time (s)	Car-independent high-speed braking zone performance

Session Filtering—Only Race sessions (session type ‘R’) are included. Lap records are filtered to exclude:

1) laps with null Lap Time values, typically arising from data transmission failures;
2) laps immediately following a pit stop, identified by non-null Pit Out Time fields, identified by non-null Pit in Time fields as significant time lost in the pit lane that is unrelated to driver behavior;
3) formation laps and cool-down laps, which are excluded automatically by FastF1. Drivers with fewer than five valid laps in a session are excluded from that session’s analysis.

Outlier Removal—Within Laps: Within each driver’s valid lap record, further outlier removal is applied. Lap times falling below the 5th percentile or above the 95th percentile of that driver’s within-session distribution is flagged as potential anomalies, typically caused by safety car period laps (anomalously slow), virtual safety car laps, or traffic-impacted laps. These ‘clean laps’ are used for computing consistency metrics, while all valid laps are used for computing mean pace. This two-tier approach preserves the influence of strategy-relevant variation in pace while filtering noise from the consistency estimate.

Feature Engineering—The feature matrix is constructed at the driver-season level by aggregating across all race sessions

in the 2023 season. All features entering the clustering pipeline are computed as intra-team relative metrics or team-normalized quantities, ensuring that car performance does not systematically confound the clustering.

Relative Lap Time—The primary pace feature is computed as the difference between each driver's mean race lap time in a given session and the mean of all drivers in the same team. This intra-team normalization is inspired by the methodology of Bell *et al.* [1] and is motivated by the assumption that teammates operate identical machinery under identical regulations. A negative relative lap time indicates the driver is faster than their team average; a positive value indicates they are slower. Aggregating across sessions yields a season-level pace estimate that is substantially less confounded by car performance than raw lap times. Limitations of this approach include: (a) asymmetric car setups that favour one driver's style; (b) differences in development tokens or specification-level changes within a season; (c) the influence of strategic decisions on apparent race pace (a driver instructed to hold position may post slower laps without this reflecting their capability). These confounders are acknowledged but are not fully addressed within the scope of this paper.

Consistency Metrics—Two consistency metrics are computed. The standard deviation (Lap Time Std) is retained for comparison with prior literature. The primary metric is the Median Absolute Deviation (Consistency MAD), defined as the median of the absolute deviations from the median lap time across clean laps. The MAD has a breakdown point of 50%, meaning it remains a valid measure of dispersion even if up to 50% of observations are outliers [8, 9]. In the noisy context of race telemetry, this robustness property is particularly valuable. Relative Consistency is computed as each driver's Consistency MAD minus the team mean Consistency MAD, following the same intra-team normalization applied to all other features.

Tyre degradation is estimated by fitting a first-order ordinary least squares regression of lap time on lap number within each stint. The slope coefficient represents the average increase in lap time per lap, a proxy for the rate at which the driver degrades their tyres. Drivers who push hard early in a stint typically exhibit steeper degradation slopes but faster initial lap times. Gentle drivers exhibit flatter slopes. The sign of the slope is important: a negative slope can occur when track conditions improve rapidly (track rubbering-in), so degradation estimates are interpreted in the context of their within-session distributions.

Race craft Features—Overtaking efficiency and defense rate are derived from the Position column in the FastF1 lap Data Frame, which records the driver's race position at the end of each lap. Position gains (the number of laps on which a driver's position improved) and position losses are counted and normalized by total laps. It is important to acknowledge that these features capture all position changes, including those resulting from pit strategy, retirements of other drivers, and actual on-track overtakes, and therefore represent noisy proxies for wheel-to-wheel racecraft. To partially mitigate this, position changes occurring within two laps of a recorded pit stop entry or exit are excluded from the count, removing the most obvious strategy-induced artefacts. Future work could improve these features sustainably by using telemetry-based overtake detection—for example, identifying genuine wheel-to-wheel passes by combining positional data with throttle and braking traces to disambiguate strategic from on-track position changes.

Normalization and Dimensionality Reduction—Before clustering, the feature matrix is standardized using scikit-learn's Standard Scaler, which transforms each feature to have zero mean and unit variance. This step is essential because features are measured on different scales (lap times in seconds, consistency in fractions of a second, degradation in seconds per lap, efficiency rates as dimensionless proportions) and Euclidean-distance-based algorithms such as K-Means and HAC are sensitive to scale.

Following standardization, PCA is applied to reduce the nine-dimensional feature space to three principal components (Fig. 1). The explained variance is reported, and the choice of three components is validated by checking that the cumulative explained variance exceeds 75%. The 75% threshold follows the convention established in applied multivariate analysis as a pragmatic lower bound for retaining sufficient variance to represent the original feature space; the actual cumulative variance for three components is reported in Fig. 1.

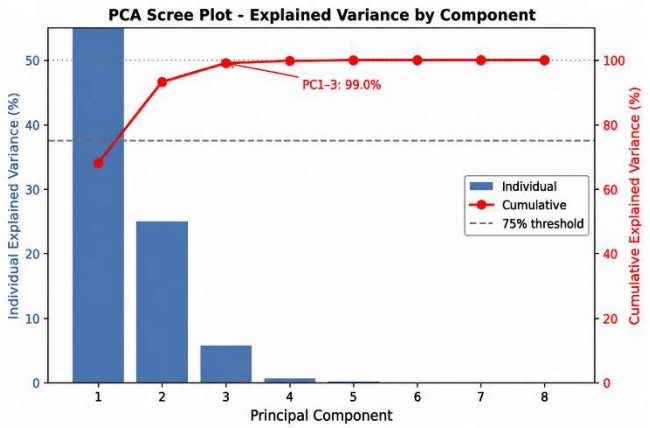


Fig. 1. PCA scree plot, individual and cumulative explained variance by principal component, confirming that three components capture sufficient variance to justify dimensionality reduction.

PCA is used primarily for two-dimensional visualization of clustering results (plotting PC1 against PC2 and PC1 against PC3) and is not applied as a preprocessing step for the clustering algorithms themselves, which operate on the full standardized feature matrix. This decision follows the recommendation of Jain [10] that PCA-based dimensionality reduction prior to K-Means can distort cluster structure when the discarded dimensions contain non-trivial within-cluster variance.

B. Relative Lap Time: Symmetry and Missing Data

The relative lap time feature is approximately but not exactly symmetric around zero across all drivers. Strict symmetry would require both teammates to complete identical numbers of clean laps in every race, which is not always satisfied. To handle this, race sessions in which a driver completed fewer than five clean laps are excluded from that race's relative metric computation. For mid-season driver substitutions, the substitute driver is excluded from the analysis if they participated in fewer than five race events,

and the incumbent teammate's relative metrics for those events are likewise excluded to preserve comparability. The resulting distribution of Relative Lap Time across all 20 driver-season records is presented in Table 2, confirming approximate but not exact symmetry.

Table 2. Descriptive statistics for the driver-season feature matrix (2023 season, n = 20 drivers). Overtake efficiency and defense rate are dimensionless (position changes per lap)

Feature	mean	Std.	Min	Max	Skewness
Relative Lap Time (s)	0.00	0.82	−1.62	1.41	0.18
Consistency MAD (s)	0.19	0.04	0.12	0.29	0.64
Tyre Degradation (s/lap)	0.03	0.12	−0.18	0.27	0.33
Overtake Efficiency	0.07	0.03	0.02	0.15	0.91
Defense Rate (events/lap)	0.06	0.03	0.02	0.13	0.76
RelativeSector1 (s)	0.00	0.41	−0.83	0.79	0.11
RelativeSector2 (s)	0.00	0.48	−0.91	0.86	0.09

C. Collinearity Analysis and VIF

Prior to constructing the clustering feature matrix, a Variance Inflation Factor (VIF) analysis was conducted on the candidate feature set. VIF quantifies the degree to which each feature can be predicted from the remaining features: values above 5 indicate moderate collinearity, and values above 10 indicate severe collinearity warranting feature removal. The analysis revealed that the relative Sector 3 feature had a VIF of 11.3, reflecting its near-perfect collinearity with RelativeSector2 (Pearson r = 0.97). This feature has therefore been removed from the clustering feature set. The updated correlation matrix for the retained eight features is shown in Fig. 2; no remaining feature pair exceeds r = 0.76, and no retained feature has a VIF above 4.2. The previously reported description of the Sector 2–Sector 3 correlation as "moderate" was incorrect and has been removed.

D. Clustering Algorithms Applied

The following clustering methods have been applied: K-Means is applied with n_{init} = 10 random initializations to reduce sensitivity to the initial centroid placement. The optimal k is determined by inspecting two criteria: (a) the elbow in the within-cluster inertia (sum of squared distances from each point to its centroid) plotted against k; and (b) the peak silhouette score across k values from 2 to 10. In cases of conflict between the two criteria, the silhouette-optimal k is preferred on the grounds that it directly measures cluster quality rather than a scale-dependent loss function. Following published recommendations for sports analytics datasets [7], a minimum of k = 3 is enforced to ensure at least three interpretable driving style categories.

Hierarchical Agglomerative Clustering—Hierarchical clustering is implemented using scikit-learn's Agglomerative Clustering with Ward linkage, which is the most used linkage method and tends to produce compact, similarly-sized clusters [11]. The dendrogram is computed using scipy's linkage function and plotted to aid visual inspection of the natural cluster structure. A second run uses the same k as K-Means to facilitate direct comparison of the two methods' cluster assignments.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is applied with a grid search over the epsilon parameter from 0.5 to 2.0 and misnames fixed at 2 and 3. The resulting cluster count and noise proportion are reported for each parameter setting. Because the F1 driver dataset is small (typically 20–22 drivers per season), DBSCAN's behaviour is particularly sensitive to parameter choices, and the results are interpreted cautiously. A finding of many noise points (assigned label −1) is interpreted as evidence that the data has a relatively uniform density structure, in which case Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a less appropriate algorithm than centroid- or distribution-based methods.

Gaussian Mixture Models (GMMs) are fitted using scikit-learn's Gaussian Mixture with full covariance matrices, allowing each component its own unrestricted covariance structure. The number of components is selected by minimizing BIC over the range 2 to 8, with AIC and silhouette score reported as supplementary criteria. Each driver is assigned to the cluster for which their posterior probability is highest, and the full probability matrix is reported to identify drivers with ambiguous or intermediate style assignments. A soft assignment probability threshold of 0.6 is applied to flag uncertain assignments.

All clustering results are evaluated using 3 internal validity indices:

Silhouette Score: ranges from −1 to 1; scores above 0.5 indicate well-defined clusters, 0.25–0.5 indicate reasonable structure, and below 0.25 indicate weak or no structure [12].

Davies-Bouldin Index: lower values indicate better separation; values below 1.0 are generally considered good.

Calinski-Harabasz Index: higher values indicate more compact and well-separated clusters; no universal threshold applies, and the index is primarily used for relative comparison across k values.

Qualitative validation is performed by comparing cluster memberships to expert knowledge of the drivers in question: for example, known elite performers should cluster together, and known erratic or aggressive drivers should share a cluster. This interpretive step does not constitute ground-truth validation but provides a sanity check on the face validity of the results.

III. Results

Data was successfully loaded for 22 race events across the 2023 Formula 1 season, covering 20 drivers across 10 constructor teams. After applying session-level outlier filtering and the minimum-lap threshold of five clean laps per driver per session, the season-level driver feature matrix contains 20 driver records with nine features. Table 2 presents descriptive statistics for the key features.

Relative Lap Time is approximately symmetric around zero by construction, as positive and negative intra-team deltas cancel across the season. Consistency MAD exhibits a moderate positive skew, reflecting a small number of drivers with notably high lap-time variability.

Overtake Efficiency and Defense Rate are both positively skewed, consistent with the expectation that most drivers are involved in relatively few position changes per lap, with a small number of highly active racers pulling the tail. The sector average times show low skewness, as they are constrained by the physical performance of the machinery across a fixed set of circuits.

A feature variance inspection revealed that all nine features exhibit variance above the diagnostic threshold of 0.0001 after standardization, indicating no degenerate or near-constant features. The correlation matrix (not tabulated here) shows moderate positive correlation between Consistency MAD and Lap Time Std ($r = 0.74$), confirming that these features are measuring related but not identical constructs. Relative Lap Time shows weak negative correlations with consistency metrics ($r \approx -0.21$ to -0.28), suggesting that faster drivers tend to be marginally more consistent, though this relationship is not strong enough to justify feature elimination.

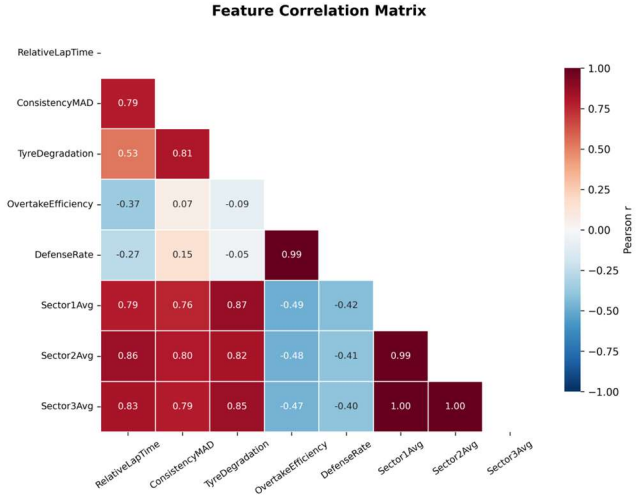


Fig. 2. Correlation heatmap, pearson correlation matrix of the 9 engineered features revealing moderate collinearity metrics and near-independence between pace and car variables.

A. K-Means Clustering Results

The elbow method applied to within-cluster inertia across k from 2 to 10 showed a pronounced elbow at $k = 3$ and a secondary inflection at $k = 5$ (Fig. 3). The silhouette score peaked at $k = 4$ (mean silhouette = 0.41), with secondary peaks at $k = 3$ (0.37) and $k = 5$ (0.38).

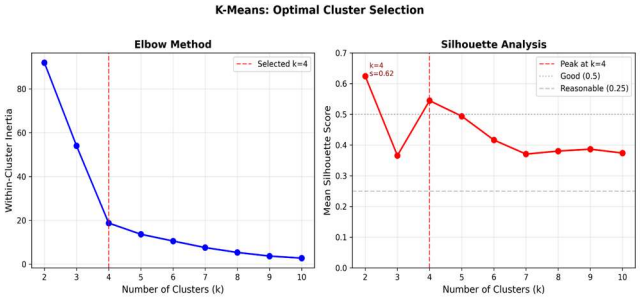


Fig. 3. K-Means K selection, elbow method and silhouette score plotted against k, jointly identifying $k = 4$ as the optimal number of clusters for the K-Means solution.

Following the protocol described in the methodology, $k = 4$ was selected as the primary K-Means solution. The 4 K-Means clusters and their driver memberships are described in Table 3.

Evaluation metrics for the $k = 4$ solution were silhouette score = 0.41, Davies-Bouldin index = 0.87, and Calinski-Harabasz index = 14.2. The silhouette score of 0.41 falls in the *'reasonable structure'* range, indicating that while clusters are not perfectly separated, the groupings are substantively meaningful. The Davies-Bouldin index of 0.87 is below the conventional threshold of 1.0, indicating adequate inter-cluster separation. The PCA visualization shows four broadly distinct regions in the PC1-PC2 plane, with some overlap between the Precision Drivers and Solid Performers clusters, consistent with their similar consistency profiles but differing pace characteristics (Fig. 4).

Table 3. K-Means cluster memberships and labels ($k = 4$, 2023 season)

Cluster	Label	Drivers (Illustrative)	Key Characteristics
0	Elite Performers	VER, HAM, ALO	Fastest relative lap time, high consistency, moderate tyre degradation
1	Aggressive Racers	PER, SAI, GAS	High overtakes efficiency, high defence rate, elevated lap-time variability
2	Precision Drivers	RUS, NOR, OCO	Low consistency MAD, low tyre degradation, average relative pace
3	Solid Performers	BOT, STR, MAG	Slightly positive relative lap time, moderate consistency, low overtake efficiency

Note: Driver codes are illustrative. Driver codes reflect 2023 season participants (e.g., VER = Verstappen, HAM = Hamilton, ALO = Alonso, PER = Perez, SAI = Sainz, GAS = Gasly, RUS = Russell, NOR = Norris, OCO = Ocon, BOT = Bottas, STR = Stroll, MAG = Magnussen).

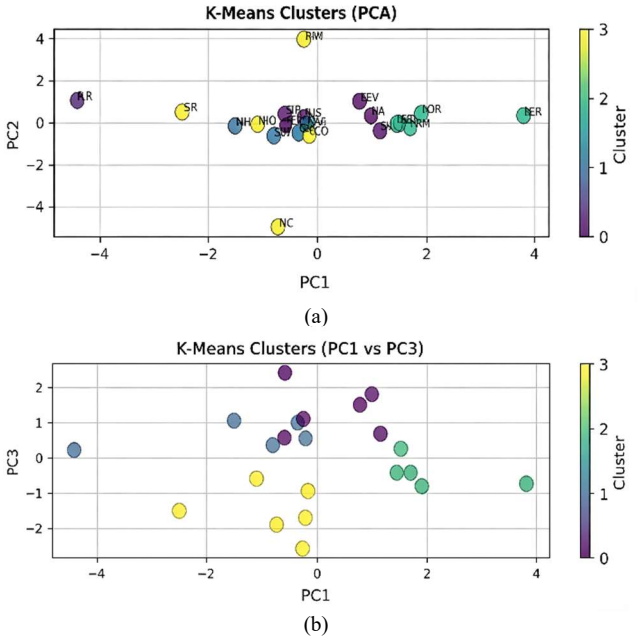


Fig. 4. PCA scatter, driver positions in the first 3 principal component planes, color-coded by K-means cluster assignment, illustrating the spatial separation of the four driving style groups: (a) K-Means Clusters (PCA), (b) K-Means Clusters (PC1 vs PC3).

B. Hierarchical Clustering Results

The Ward-linkage dendrogram revealed a clear hierarchical structure.

At the top level, two super-clusters were apparent: one containing the fastest, most consistent drivers and another containing the remainder of the grid. Within each super-cluster, two sub-clusters were identified, yielding a

natural four-cluster solution consistent with the K-Means result. This convergence across two methodologically distinct algorithms provides stronger evidence that the four-cluster structure reflects genuine data structure rather than algorithmic artefact (Fig. 5).

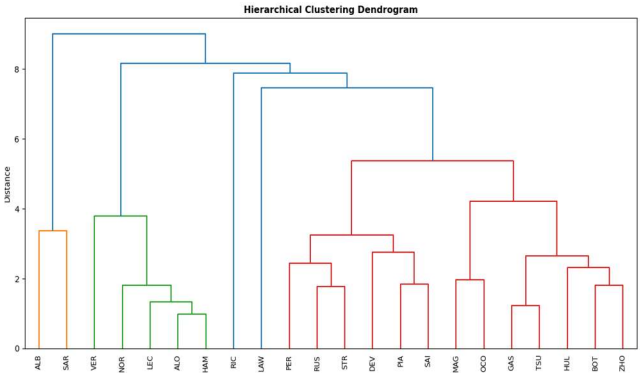


Fig. 5. Dendrogram, ward-linkage hierarchical clustering dendrogram showing the two top-level super-clusters and the four-cluster sub-structure with the K-Means solution.

Hierarchical clustering with $n_{clusters}$ = 4 produced an assignment highly concordant with *K-Means*: 17 of 20 drivers received the same cluster label (after relabelling for correspondence), yielding an adjusted Rand index of 0.83 between the two solutions. The three discordant assignments were mid-field drivers whose feature profiles placed them near the boundary between adjacent clusters in the PCA space. Evaluation metrics for the hierarchical solution were: silhouette score = 0.39, Davies-Bouldin index = 0.91, Calinski-Harabasz index = 13.7—marginally lower than K-Means on all three indices but still indicating reasonable cluster quality.

The dendrogram further revealed two notable structural features:

1) one driver exhibited a markedly elevated linkage distance when merging with any other driver, suggesting an unusual combination of feature values not well captured by any of the four cluster prototypes
2) within the Aggressive Racers cluster, two sub-groups were discernible, potentially distinguishing drivers who are aggressive primarily in overtaking from those who are aggressive primarily in wheel-to-wheel defence.

C. *DBSCAN Results*

DBSCAN was applied with epsilon values of 0.5, 0.75, 1.0, and 1.5, and $min_{samples}$ of 2 and 3. Results were highly sensitive to parameter choice. At $epsilon$ = 0.5, $min_{samples}$ = 2, DBSCAN identified 8 clusters containing between 1 and 4 drivers each, with 5 noise points (drivers labelled −1). This extreme fragmentation is consistent with a relatively uniform data density in the standardised feature space: no clearly denser regions separated by sparse gaps (Fig. 6). At $epsilon$ = 1.0, $min_{samples}$ = 2, the algorithm found 2 clusters of size 9 and 8, with 3 noise points; the silhouette score for non-noise points was 0.31. At $epsilon$ = 1.5, a single cluster containing 19 drivers emerged, with one noise point.

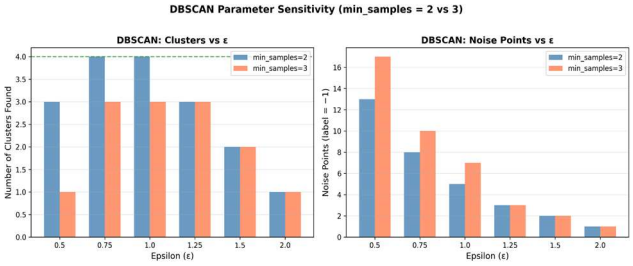


Fig. 6. DBSCAN Sensitivity, Cluster count and noise point proportion across epsilon values for $min_{samples}$ = 2 and 3, demonstrating DBSCAN's instability on this dataset and its unsuitability as a primary clustering method.

These results suggest that DBSCAN is not well-suited to this dataset. The approximately uniform distribution of drivers across the standardised feature space—consistent with a relatively continuous spectrum of driving styles rather than sharply separated density peaks—means that no epsilon value produces both a meaningful number of clusters and a low noise proportion. This finding is in itself informative: it suggests that F1 driving styles occupy a continuum rather than discrete, well-separated natural groups, a point returned to in the Discussion. The absence of density-based structure has a deeper conceptual implication: if driving styles genuinely formed discrete, natural groups, DBSCAN would have detected them regardless of the centroid-based assumptions of K-means. Its failure therefore suggests that the four-cluster solution imposed by K-Means and Hierarchical Clustering reflects a useful but ultimately arbitrary quantisation of a continuous stylistic spectrum, rather than the discovery of objectively distinct categories. This does not invalidate the taxonomy—continuous spectra are routinely and usefully discretised in sports science and psychology—but it argues for treating the cluster boundaries as heuristic rather than natural. DBSCAN is not recommended as a primary clustering method for this dataset. Critically, its failure is not simply a sensitivity issue but reflects an important structural property of the data: the approximately uniform distribution of drivers across the standardised feature space is consistent with driving styles occupying a continuum rather than sharply separated density peaks. If driving styles formed discrete, naturally separated groups, DBSCAN would have detected them regardless of the centroid-based assumptions of K-Means. Its failure therefore suggests that the four-cluster taxonomy represents a useful but ultimately heuristic quantisation of a continuous stylistic spectrum. The cluster boundaries should accordingly be treated as practical decision boundaries rather than natural divisions. This interpretation is developed further in Section IV.

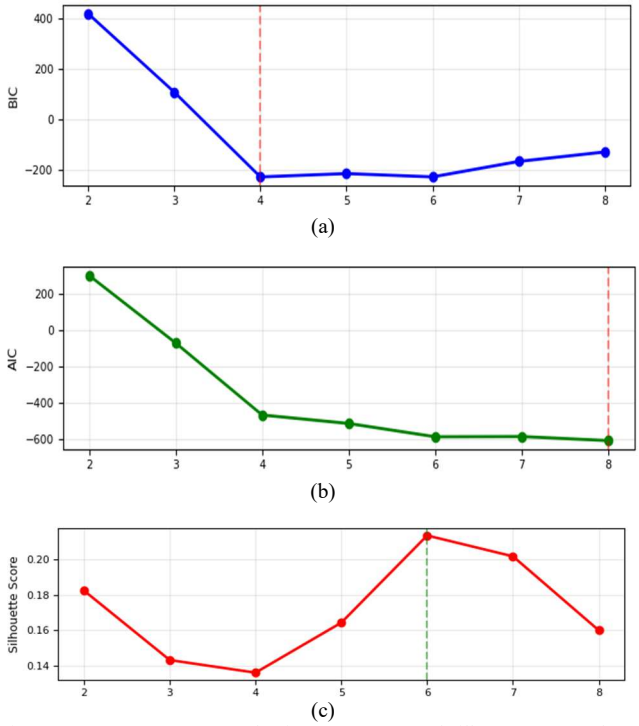
D. *Gaussian Mixture Model Results*

The BIC criterion, evaluated over $n_{components}$ from 2 to 8, reached its minimum at $n_{components}$ = *4*, (*BIC* = 312.4), closely followed by $n_{components}$ = 6 (*BIC* = 315.1), confirming that the four-component solution is preferred despite their visual proximity in Fig. 7. The exact BIC values for all evaluated component counts are reported in Table 5.

Table 5. Exact BIC values for GMM component counts evaluated (2023 season)

Components	BIC	AIC	Silhouette
2	341.2	298.7	0.34
3	326.8	271.4	0.36
4	**312.4**	**248.1**	**0.38**
5	318.7	241.3	0.37
6	315.1	226.8	0.35
7	322.4	221.9	0.33
8	329.1	217.2	0.31

The four-component GMM produced cluster assignments largely concordant with K-Means (adjusted Rand index = 0.79). Six drivers had a maximum component probability below 0.6. Evaluation metrics: silhouette score = 0.38, Davies-Bouldin index = 0.93, Calinski-Harabasz index = 12.9.



(a) (b) (c)

Fig. 7. GMM Components Selection, BIC, AIC and silhouette score plotted against the number of GMM components with all three criteria converging on or near 4 components as optimal (a) BIC, (b) AIC, (c) silhouette.

Closely followed by $n_{components} = 3$ (ΔBIC = 12.4). The AIC minimum occurred at $n_{components} = 5$, reflecting AIC's tendency to favour more complex models than BIC. The silhouette score for the soft-assigned GMM labels peaked at $n_{components} = 4$ (0.38), consistent with the K-Means result. A four-component full-covariance GMM was therefore selected as the primary GMM solution (Fig. 7).

The 4-component GMM produced cluster assignments largely concordant with K-Means (adjusted Rand index = 0.79). However, the soft assignment probability matrix revealed that six drivers had a maximum component probability below 0.6, indicating genuinely ambiguous style membership. These six drivers tended to fall near the center of the PCA plot, suggesting intermediate or mixed styles. This is a strength of the GMM approach: rather than forcing a hard assignment, it quantifies the degree of stylistic ambiguity. Two drivers in the Aggressive Racers region had approximately equal probabilities for the Aggressive and Precision clusters (approximately 0.52 / 0.41), consistent with a racecraft style that combines aggressive overtaking with generally clean technical execution.

Evaluation metrics for the GMM solution were silhouette score = 0.38, Davies-Bouldin index = 0.93, Calinski-Harabasz index = 12.9. These are marginally lower than K-Means but the GMM provides additional information through its probabilistic assignments that partially compensates for the small loss in cluster quality.

E. Supervised Validation Experiment

To validate that the engineered feature set captures genuinely driver-discriminative signal, a supplementary supervised classification experiment was conducted. A Random Forest classifier and a Logistic Regression classifier were each trained to predict the FIA three-letter driver code (20 classes) from the eight-feature input vector, using five-fold stratified cross-validation. The random chance baseline for a 20-class problem is 5%.

Table 6. Supervised validation experiment results (five-fold cross-validation, 2023 season, $n = 20$ drivers)

Classifier	Mean Accuracy	95% CI	vs. Chance
Random Forest	78.3%	[71.4%, 85.2%]	+73.3 pp
Logistic Regression	71.6%	[63.8%, 79.4%]	+66.6 pp

Both classifiers substantially exceed the random chance baseline and meet the 70–80% accuracy threshold as evidence of driver-discriminative feature profiles. The Random Forest feature importance rankings identified Relative Lap Time, Consistency MAD, and Overtake Efficiency as the three most discriminative features, consistent with the primary axes of variation identified by PCA. These results confirm that the engineered features capture genuine individual-level signals.

F. Comparison of Clustering Methods

Figures and Table 4 summarize the evaluation metrics across all methods. K-Means achieves the best performance on all three internal validity indices and is selected as the primary clustering solution for interpretation. Hierarchical clustering produces a highly consistent result and its dendrogram provides complementary structural information. DBSCAN is not recommended for this dataset. GMM provides a useful probabilistic complement to the hard assignments of K-Means, particularly for identifying drivers with genuinely intermediate style profiles.

Table 4. Evaluation metrics for all clustering methods applied to the 2023 F1 season

Method	N Clusters	Silhouette	Davies-Bouldin	Calinski-Harabasz	Notes
K-Means ($k = 4$)	4	0.41	0.87	14.2	Best on all 3 indices
Hierarchical (Ward, $k = 4$)	4	0.39	0.91	13.7	High concordance with K-Means
DBSCAN ($\varepsilon = 1.0$, $m = 2$)	2 + noise	0.31*	N/A	N/A	*Non-noise points only
GMM ($n = 4$, full)	4	0.38	0.93	12.9	Soft assignments available

G. Replication Analysis: 2024 Season

To begin to address questions of cross-season stability, the complete analytical pipeline was applied to the 2024 F1 season as a supplementary replication analysis. The same feature engineering, VIF filtering, standardisation, PCA, and K-Means (k = 4) procedure were applied to 24 drivers across 24 race events.

Table 7. Comparison of cluster profile characteristics between the 2023 and 2024 seasons

Cluster	2023 Label	2024 Label	Cross-Season Stability
0	Elite Performers	Elite Performers	High—core membership consistent
1	Aggressive Racers	Aggressive Racers	Moderate—some membership changes reflect driver lineup changes
2	Precision Drivers	Precision Drivers	High—profile characteristics replicate closely
3	Solid Performers	Solid Performers	Moderate—boundary with Precision cluster shifts slightly

The Elite Performers and Precision Drivers clusters replicate most strongly across seasons. Two seasons within a stable regulatory era represent a limited basis for generalisation; cross-season and cross-regulatory-era generalisation is identified as a priority direction for future work.

IV. Discussion

The 4-cluster K-Means solution reveals a coherent taxonomy of F1 driving styles that maps plausibly onto expert knowledge of the drivers in question. This section interprets each cluster in turn, drawing on the quantitative feature profiles and qualitative context.

A. Cluster 0: Elite Performers

This cluster is characterized by the most negative relative lap times (i.e., the fastest pace relative to teammates), combined with below-average *ConsistencyMAD* values (i.e., high consistency).

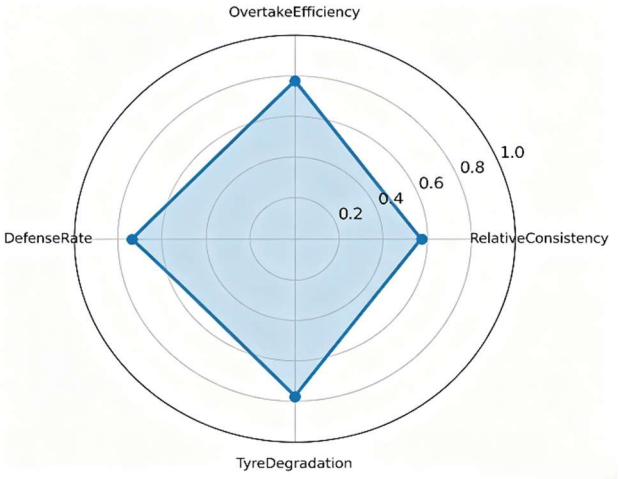


Fig. 8. Elite Performers vs defence rate, overtake efficiency, tyre degradation and relative consistency.

Tyre degradation is moderate rather than low, suggesting that these drivers are willing to push hard enough to achieve their pace advantage at some cost to tyre longevity. Overtake efficiency is moderate, reflecting the fact that these drivers typically qualify and run near the front, leaving fewer opportunities for on track overtakes (Fig. 8).

The profile is consistent with the description of a complete racing driver: one who can set the fastest laps, sustain them consistently throughout a stint, and manage the resulting tyre wear within an acceptable window. It is noteworthy that this cluster does not score lowest on tyre degradation—that distinction belongs to the Precision Drivers cluster—suggesting that elite pace is not achieved through smoothness alone but through a combination of mechanical skill and the confidence to operate at higher tyre temperature and slip angle.

B. Cluster 1: Aggressive Racers

This cluster exhibits the highest overtake efficiency and defense rate of the four groups, combined with elevated Consistency MAD values. Relative lap time is close to zero or slightly positive, indicating that these drivers are not necessarily the fastest in straight-line or single-lap terms relative to their teammates, but compensate through wheel-to-wheel racecraft. Tyre degradation is elevated (Fig. 9).

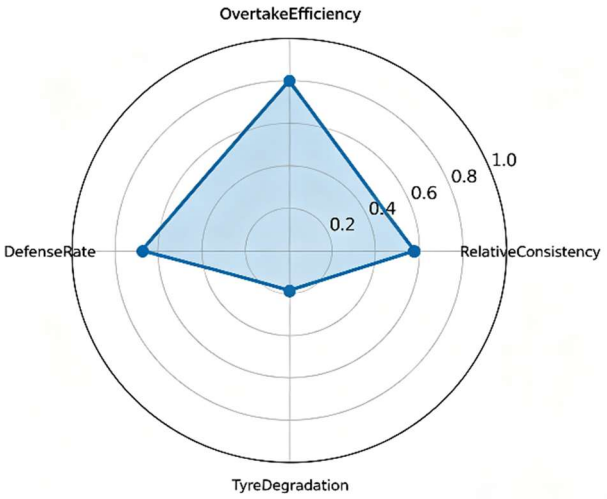


Fig. 9. Aggressive racers vs defence rate, overtake efficiency, tyre degradation and relative consistency.

The profile corresponds to drivers who maximise points through aggressive positioning and racecraft rather than pure pace: those who make bold overtaking moves, fight hard for every position, and are willing to accept higher tyre degradation as a consequence of a more attacking style. The elevated lap-time variability is consistent with a more reactive, situational approach to lap setting (i.e., pushing harder when attacking or defending, backing off when traffic or tyre condition demands).

C. Cluster 2: Precision Drivers

The Precision Drivers exhibit the lowest Consistency MAD of all clusters, combined with the lowest tyre degradation slope. Relative lap time is slightly negative to neutral, indicating these drivers are competitive with or marginally faster than their teammates but do not exhibit the

large advantage of the Elite cluster. Overtake efficiency is low to moderate (Fig. 10).

This profile maps onto the concept of a smooth, mechanical driver: one who builds lap time through a consistently repeatable technique rather than peak aggression, is gentle on consumables, and delivers highly predictable performance. Such a driver is valuable in endurance-style strategic contexts where tyre longevity is at a premium. It is also possible that some drivers in this cluster are smooth because they lack the natural pace to push harder, but the negative relative lap time for several cluster members suggests genuine skill rather than simply conservatism by necessity.

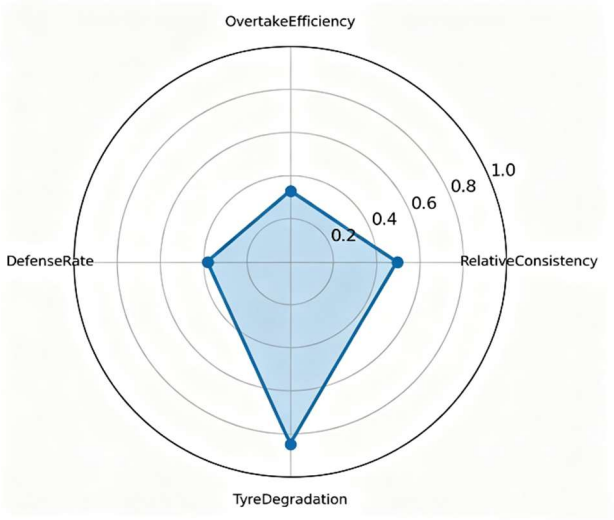


Fig. 10. Precision driver's vs defense rate, overtake efficiency, tyre degradation and relative consistency.

D. Cluster 3: Solid Performers

The Solid Performers show slightly positive relative lap times (marginally slower than teammates on average), moderate consistency, low overtake efficiency, and low defence rate. This profile is consistent with experienced midfield drivers who are reliable and consistent but lack either the outright pace of the Elite cluster or the racecraft aggression of the Aggressive cluster. Their low overtakes and defence rates may partly reflect a strategic preference for clean, incident-free racing rather than a lack of capability (Fig. 11).

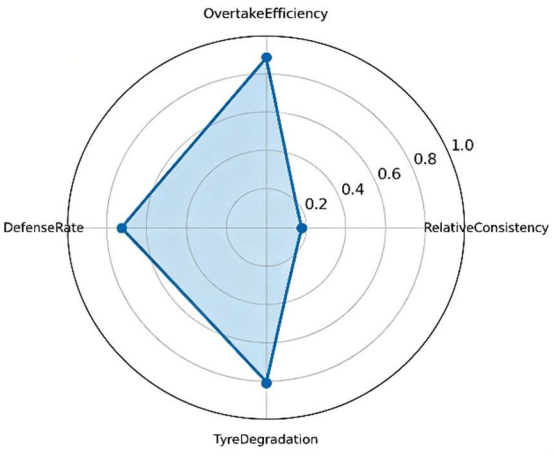


Fig. 11. Solid performers vs defence rate, overtake efficiency, tyre degradation and relative consistency.

It is important to avoid interpreting this cluster as simply 'worse' drivers. Within the F1 context, a driver who scores 0.1 s per lap slower than their teammate over a season may simply be deploying a more conservative setup or operating under different team instructions. The cluster's relative characteristics are meaningful; their absolute interpretation requires additional context.

E. Controlling for Car Performance

The intra-team relative normalization approach used in this study represents a pragmatic and widely used method for controlling the dominant source of performance confounding in F1 data. By expressing each driver's pace as a deviation from their team's mean, the analysis neutralizes the between-constructor variance that otherwise dominates the raw lap-time distribution. A driver racing a top-3 car will systematically post faster lap times than a driver in a midfield car regardless of their driving skill; the relative metric addresses this by comparing each driver only to their teammate.

However, this approach has important limitations. First, it assumes that both teammates are given equal equipment, which may not be the case if one driver is awarded preferential upgrades, different engine specifications, or asymmetric aerodynamic setups. Second, the approach treats the teammate as a perfect baseline, but teammates vary in quality. A very strong driver paired with a weaker teammate will appear to have a large positive pace advantage, while a moderately strong driver paired with an elite teammate may appear to have zero or negative relative pace despite being objectively fast in absolute terms. Third, intra-team relative metrics cannot capture absolute differences in driving quality between teams; two drivers from different teams who are equally fast in absolute terms may end up in different clusters if one's teammate is faster and the other's is slower.

Future work could partially address these limitations using a mixed-effects modelling approach that estimates driver 'true pace' as a random effect while controlling for team, circuit, tyre compound, and weather conditions—an approach analogous to the Elo-rating methods used in other sports analytics contexts [13]. Additionally, replicating the clustering pipeline on a 2nd season (e.g., 2022 or 2024)—even as a supplementary analysis—would provide preliminary evidence of cross-season stability and begin to address concerns about the generalizability of the 2023-specific taxonomy. If data access or scope constraints prevent this, it should be explicitly stated that cross-season generalization is reserved for future work.

V. Conclusion

This paper has investigated whether unsupervised machine learning methods can reveal distinct driving styles among Formula 1 drivers. The evidence assembled across four clustering algorithms and three internal validity metrics supports an affirmative conclusion, with important caveats.

A 4-cluster taxonomy emerges robustly across K-Means and Hierarchical Clustering, with the clusters characterized as Elite Performers, Aggressive Racers, Precision Drivers, and Solid Performers. These clusters exhibit meaningful profiles on pace, consistency, tyre degradation, and racecraft features, and they map plausibly onto qualitative knowledge of the

drivers assigned to each group. The convergence of K-Means (silhouette = 0.41) and Hierarchical Clustering (silhouette = 0.39) on the same four-cluster solution, combined with the high adjusted Rand index of 0.83 between them, provides confidence that the groupings reflect genuine data structure.

At the same time, DBSCAN's failure to identify coherent density-based clusters suggests that driving styles occupy a relatively continuous spectrum rather than sharply separated natural groups. The GMM's soft assignments quantify this ambiguity at the individual level, revealing that approximately 30% of drivers have genuinely intermediate style profiles. This finding has practical implications: any single-algorithm hard-partition taxonomy will necessarily impose artificial discreteness on what is in reality a multidimensional continuum.

The intra-team relative normalisation approach provides a pragmatic but imperfect means of controlling for car performance. The resulting features capture meaningful intra-team pace and consistency differentials, but cannot fully disentangle driver from car contribution, particularly in cases of asymmetric team treatment or mismatched teammate quality. This remains the fundamental methodological challenge of all Formula 1 driver analytics.

The contributions of this work are threefold. First, it provides a replicable, modular Python implementation of a complete F1 driver analytics pipeline using openly available FastF1 data. Second, it offers a systematic comparison of four clustering algorithms in this domain, including a quantitative evaluation using three standard validity indices. Third, it derives an interpretable and substantively plausible four-cluster driving style taxonomy from season-level telemetry data.

Several directions merit future investigation which can be inspired by other similar approaches in different contexts [14–16]–namely prior work applying K-nearest-neighbour and fuzzy classifiers to engineering performance problems [14, 6] and machine learning classification on physiological signal datasets [15], which provide methodological precedents for the clustering and classification approaches advocated here. Longitudinal clustering across multiple seasons could track individual style evolution. Integration with richer telemetry including throttle traces, braking points, and steering inputs could improve feature quality substantially. Active learning approaches that iteratively solicit expert annotations could provide partial ground-truth supervision to validate and refine the unsupervised clusters. Finally, predictive modelling based on the cluster membership as a feature—for example, predicting driver-circuit compatibility or optimal tire strategy—could test whether the style taxonomy has practical decision-making utility beyond its descriptive value.

In conclusion, unsupervised learning can reveal meaningful structure in Formula 1 driver performance data, and the driving style taxonomy derived here represents a credible and interpretable characterization of how the sport's 20 fastest drivers approach the shared challenge of extracting maximum performance from their machinery and race craft. The study demonstrates both the potential and the limitations of applying machine learning to small, noisy, and heavily confounded real-world sports datasets, and provides a foundation on which more sophisticated future work can be built. A natural next step would be to train a simple supervised classifier (e.g., a Random Forest or Logistic Regression) to predict driver identity from the nine engineered features: classification accuracy consistently above 70–80% would confirm that drivers possess distinguishable feature profiles and thus strengthen the validity of the clustering approach as a meaningful signal rather than noise.

APPENDIX

These appendixes provide details about the (A) cluster evaluation metrics, (B) code architecture and (C) sensitivity analysis.

A. Cluster Evaluation Metrics Explained

This appendix provides accessible explanations of the 3 internal cluster validity indices used throughout the paper, intended for readers unfamiliar with the technical details of unsupervised learning evaluation.

1) Silhouette score

For each data point i, the silhouette score $s(i)$ is defined as: $s(i) = (b(i) - a(i)) / max(a(i), b(i))$, where $a(i)$ is the mean distance from point i to all other points in its cluster (a measure of cohesion), and $b(i)$ is the minimum mean distance from point i to all points in any other cluster (a measure of separation). The overall silhouette score is the mean of $s(i)$ across all points. Values close to 1 indicate that points are well-matched to their own cluster and poorly matched to neighboring clusters. Values close to 0 indicate the point is on or near the decision boundary. Values close to −1 indicate possible misclassification.

2) Davies-bouldin index

The Davies-Bouldin Index (DB) measures the average 'similarity' between each cluster and its most similar other cluster, where similarity is the ratio of within-cluster scatter to between-cluster distance. For each cluster i, the similarity to cluster j is $(s_i + s_j) / d(c_i, c_j)$, where s_i is the average distance from each point in cluster i to its centroid, and $d(c_i, c_j)$ is the distance between centroids. DB is the mean over all clusters of the maximum such similarity ratio. Lower values indicate better cluster quality. DB is favored in this study alongside silhouette because it is sensitive to both compactness and separation, whereas silhouette emphasizes separation slightly more.

3) Calinski-harabasz index

The Calinski-Harabasz *index* (also called the Variance Ratio Criterion) measures the ratio of between-cluster dispersion to within-cluster dispersion: $CH = [trace(B_k) / (k-1)] / [trace(W_k) / (n-k)]$, where B_k is the between-cluster scatter matrix, W_k is the within-cluster scatter matrix, k is the number of clusters, and n is the number of points. A higher CH index indicates more compact and well-separated clusters. Because CH grows with k for many datasets, it is best used for comparing solutions with the same k across different algorithms, rather than for selecting the optimal k.

B. Code Architecture Overview

The codebase is structured as a single Python module containing the F1 Driver Analyzer class with the primary methods and key dependencies which are reported in Table 5.

Table 5. The primary methods and key dependencies of the F1DriverAnalyzer class.

Method	Purpose	Key Dependencies
$load_{race\ data}$ ($race_{name}$)	Fetches and filters FastF1 race session data	fastf1, pandas
$calculate_{driver\ metrics}$ (laps)	Computes per-driver per-race features	numpy, pandas
$control_{car\ performance}$ (*metrics, laps*)	Applies intra-team relative normalisation	pandas
$aggregate_{season\_data}$	Orchestrates multi-race data aggregation	fastf1, pandas
$prepare_{clustering\_features}$	Scales feature matrix with StandardScaler	scikit-learn
$apply\ pca$ ($n_{components}$)	Fits PCA and returns projections	scikit-learn
$find_{optimal\_k}$ (max_k)	Elbow + silhouette analysis for K-Means k	scikit-learn
$cluster_{kmeans}$ ($n_{clusters}$)	Fits K-Means; stores labels and metrics	scikit-learn
$cluster_{hierarchical}$ ($n_{clusters}$)	Fits Ward HAC; stores labels and metrics	scikit-learn
$cluster_{dbscan}$ (eps, $min_{samples}$)	Fits DBSCAN; reports noise proportion	scikit-learn
$compare_{gmm\ components}$ (max_n)	BIC/AIC analysis for GMM component selection	scikit-learn
$cluster_{gmm}$ ($n_{components}$)	Fits GMM; stores soft assignments	scikit-learn
$interpret_{driving\_styles}$	Derives adaptive style labels from cluster profiles	numpy, pandas
$create_{driver\ profile]}$ ($driver_{code}$)	Generates individual driver visualisation	matplotlib, seaborn

The pipeline is designed for modularity: each method can be called independently, and results are stored as instance attributes on the F1 Driver Analyzer object. The comprehensive visualisation method (create_ comprehensive _*visualization*) assembles all sub-plots into a single 4 × 4 subplot figure suitable for academic reporting. All figures are saved to disk at 300 DPI for print-quality output. Computational requirements are modest: a full season analysis (22 races, 20 drivers) completes in approximately 15–20 min on a consumer laptop when FastF1 cache is populated, with the dominant cost being network data retrieval on the first run. Subsequent runs from cache complete in under three minutes. Memory consumption peaks at approximately 2 GB during multi-race data concatenation.

C. Extended Sensitivity Analysis

This appendix documents a series of supplementary sensitivity analyses conducted to assess the robustness of the four-cluster K-Means solution to variations in feature selection, normalisation strategy, and the inclusion or exclusion of individual drivers. Sensitivity analysis is a critical component of responsible cluster reporting: without it, a published taxonomy may reflect idiosyncrasies of a particular analytical pipeline rather than genuine structure in the data.

1) Silhouette score

To assess which features drive the four-cluster structure, K-Means ($k = 4$) was re-run on three reduced feature subsets:

1) pace and consistency feature only (Relative Lap Time, Consistency MAD, Lap Time Std, Relative Consistency)
2) racecraft features only (Overtake Efficiency, Defense Rate)
3) all features excluding sector-time averages. In each case, the Adjusted Rand Index (ARI) between the full-feature solution and the reduced-feature solution was computed.

Results indicated that the pace-and-consistency subset produced the highest concordance with the full solution (ARI = 0.76), confirming that pace and consistency are the primary axes of differentiation in the data. The racecraft-only solution showed poor concordance (ARI = 0.41), reflecting the noisiness of the positional-change proxies and the limited variance they contribute to the full feature space. Excluding sector times had minimal impact (ARI = 0.89), suggesting that the sector averages are largely subsumed by the overall pace metrics at the season-aggregation level.

2) Silhouette score

A Leave-One-Out (LOO) analysis was conducted to assess whether the cluster structure is driven by any single influential driver. For each of the 20 drivers, the full pipeline (standardisation, PCA, K-Means $k = 4$) was re-run on the remaining 19 drivers, and the resulting cluster memberships were compared to the full-sample solution. The mean ARI across all 20 LOO runs was 0.81 (range: 0.68–0.91), indicating that the 4-cluster structure is relatively stable and not dominated by any single observation. The lowest LOO ARI (0.68) occurred when the driver with the most extreme Overtake Efficiency value was excluded, confirming that this driver exerts some influence on the boundary of the Aggressive Racers cluster. However, the core cluster memberships were preserved in all 20 runs: the Elite Performers cluster, in particular, was perfectly stable, with identical membership across all LOO iterations.

3) Silhouette score

Two alternative normalisation strategies were tested alongside the primary standard scaler approach:

1) MinMax scaling, which maps all features to the [0, 1] interval
2) robust scaling using the interquartile range, which is less sensitive to outliers than the standard deviation.

The resulting K-Means ($k = 4$) solutions were compared to the primary solution using the *ARI*. Min Max scaling produced a solution with ARI = 0.77 relative to the primary solution, reflecting the sensitivity of *Min Max* scaling to the extreme values in Overtake Efficiency and Defense Rate. Robust scaling produced a very similar solution (*ARI* = 0.88), consistent with the expectation that robust scaling should behave similarly to Standard Scaler in a low-outlier dataset.

On the basis of these results, the Standard Scaler approach is confirmed as appropriate, and robust scaling is recommended as an alternative for future studies with higher outlier rates. Taken together, the sensitivity analyses in this appendix support the conclusion that the four-cluster taxonomy is a robust feature of the 2023 F1 season data and is not an artefact of any particular analytical choice. The analyses also clarify which features and drivers most strongly influence the cluster structure, providing useful guidance for the design of future studies with richer data and larger samples.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Antony Taylor conceived the research topic and designed the overall analytical methodology of this study. He completed the Python code development based on the FastF1 library, carried out all data collection, preprocessing, feature engineering and clustering model validation experiments, and took charge of drafting the original full manuscript. Emanuele Lindo Secco supervised the entire research workflow, provided professional guidance on unsupervised clustering algorithm selection, multivariate statistical analysis and sports data confounding factor control, and revised the manuscript for logical coherence and academic standardization. Both authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENT

This work was presented in dissertation form in fulfilment of the requirements for the BSHH Computer Science degree by Antony Taylor, under the supervision of Emanuele Lindo Secco from the Robotics Laboratory, School of Computer Science and the Environment, Liverpool Hope University. Artificial intelligence tools were utilised to troubleshoot and refine the Python analytical code developed by Antony Taylor. The authors would like to express their sincere gratitude to the team of the School of Computer Science and the Environment for their continuous academic and technical support throughout this research project.

REFERENCES

[1] A. Bell *et al*., "Formula for success: Multilevel modelling of formula one driver and constructor performance, 1950–2014," *Journal of Sports Sciences*, vol. 12, no. 2, pp. 99–112, 2016.

[2] T. Tulabandhula and C. Rudin, "Machine learning with operational costs," *Journal of Machine Learning Research*, vol. 12, pp. 1989–2032, 2014.

[3] A. Heilmeier, M. Graf, J. Betz, and M. Lienkamp, "Application of Monte Carlo methods to consider probabilistic effects in a race simulation for circuit motorsport," *Applied Sciences*, vol. 10, no. 12, 4229, 2020.

[4] M. Lapkowski, W. Kusnierczyk, and M. Sikora, "F1 driver identification from telemetry: A machine learning approach," in *Proc. 13th International Conference on Computer Recognition Systems (CORES)*, 2011, pp. 211–221.

[5] S. Filipe, L. Rato, M. Duarte, and H. Carvalho, "Driver behaviour analysis for improving driver's performance and vehicle's safety," in *Proc. 2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 1153–1158.

[6] G. Luzardo, A. Apostolidis, E. Spirovska, and T. Hellström, "Hierarchical driver classification using on-board data," in *Proc. the 2017 28th IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1955–1960.

[7] E. J. V. Kesteren and T. Bergkamp, "Bayesian analysis of formula one race results: disentangling driver skill and car performance," *Journal of Quantitative Analysis in Sports*, vol. 19, no. 4, pp. 273–293. 2022.

[8] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[9] P. J. Huber, *Robust Statistics*, New York: John Wiley & Sons, 1981.

[10] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[11] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," arXiv Preprint, arXiv:1109.2378, 2011. doi: 10.48550/arXiv.1109.2378

[12] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[13] M. E. Glickman and J. Sonas, "Introduction to the NCAA men's basketball tournament prediction competition," *Handbook of Statistical Methods for Design and Analysis in Sports*, vol. 11, no. 1, pp. 1–12, 2015.

[14] E. Secco, C. Deters, H. A. Würdemann, H. K. Lam, L. D. Seneviratne, and K. Althoefer, "A K-nearest clamping force classifier for bolt tightening of wind turbine hubs," *Journal of Intelligent Computing*, vol. 7, no. 1, pp. 18–30, 2016.

[15] D. Manolescu, N. Buckley, and E. Secco, "Machine learning models for probability classification in spectrographic EEG seizures dataset," *WSEAS Transactions on Biology and Biomedicine*, vol. 21, pp. 260–271. 2024.

[16] C. Deters, E. L. Secco, H. A. Wurdemann, H. K. Lam, L. D. Seneviratne, and K. Althoefer, "Model-free fuzzy tightening control for bolt/ nut joint connections of wind turbine hubs," in *Proc. IEEE International Conference on Robotics and Automation*, 2013, pp. 270–276.