Journal of Interdisciplinary Science Topics

Does History Matter? Applying Predictive Analytics to Formula 1

Kay Davis
Natural Sciences (Life and Physical Sciences), School of Biological and Biomedical Sciences, University of Leicester
08/04/2026

Abstract
Predicting reliability in Formula 1 is challenging, because team performance is determined by both technical conditions and long-term organisation within the teams. This study tests if the historical data of a constructor (and its lineage in F1) could be used in predictive analytics to forecast 2026 DNF (Did Not Finish) rates. Using these, I will be attempting to create short term forecasts by three different modelling methods, assessing the most accurate, and then applying it to the upcoming season to attempt to apply predictive analytics to Formula 1.

Keywords: *Sports; Probability/Statistics; Mathematical modelling; Forecasting; Team Performance; Formula 1 (F1)*

Introduction
In Formula 1 (F1), a team's performance depends on a delicate balance of speed, control, and reliability. As a result of this, the skill of a driver or the pace a car has are often the main focuses in race discussions, but it is widely recognised that the ability of a team to finish the race is a major factor that often comes into play in dramatic ways. A DNF (Did Not Finish) can come from mechanical failure, damage from contact with a rival car, or most recently in Aston Martin's case the car vibrating so much the driver, Fernando Alonso, could not handle the pain caused from it [1].

However, reliability in F1 often comes from experience and organisation. Newer teams entering the sport almost always experience higher DNF rates than older teams, at least at first [1]. There is also correlation between money spent or better facilities owned by certain teams improving reliability [2]. This is due to being able to produce more data to build on easier. However, most teams on the grid have long histories, as teams in modern F1 are more likely to be bought and rebranded than started from scratch (with Haas and Cadillac being recent exceptions) [3]. As such, they inherited data and insight to build upon from the team's previous iteration [4].

In this paper, I will explore whether from this historical data, from when teams ran under different names and ownership, can be used to predict future results. The purpose is not simply to see which team is most or least reliable, but whether this historical data is still playing a part on modern F1 through predictive analytics.

Methods
Race result data was collected for each team, including the results under all previous names of that team [5]. The names for each team were grouped, and all results were pulled from 1950 to 2025. For each team, a seasonal DNF rate was produced (proportion of race starts each season that ended in a retirement for any reason).

Three forecasting approaches were compared: a carry-forward model, simple exponential smoothing (SES), and Holt's linear trend method [6,7]. I chose these three since they range in the sophistication of their modelling, so testing all three gives a good range to find which model fits best. I assessed the results made by all three models by testing them against data that was withheld when producing the models to make sure I chose one based only on its predictive ability. The model with the best performance was then used to produce a forecast for the 2026 season.

Results

Team	Model	mae	rmse
Alpine	Naive	0.275	0.394493
Aston Martin	Naive	0.096296	0.096864
Ferrari	SES	0.091899	0.111341
McLaren	Naive	0.274074	0.312299
Mercedes	SES	0.1292	0.137531
Racing Bulls	Naive	0.266667	0.408248
Red Bull Racing	SES	0.2	0.244949
Sauber	SES	0.305547	0.340344
Williams	SES	0.213434	0.292918

Table 1 [5,8,9,10] - Table comparing the findings for which model was most accurate in predictions for each team's data. The model specified is the one with lowest prediction error for that team (naïve refers to the carry-forward modelling). MAE refers to mean absolute error, showing average difference between the predicted and observed DNF rates. RMSE (root mean squared error) measures the root of the mean squared difference between predicted and observed DNF rates. Gives more weight to larger errors than MAE to help analyse the effectiveness of the method [1].

Comparing the forecasting models (shown in Table 1) suggested that SES was the best overall predictive model for this use. This makes logical sense, since it means more recent historical data was more important, which makes sense since some teams have histories greater than 20 years. Data from 1980 for instance is unlikely to have an impact in modern F1 reliability.

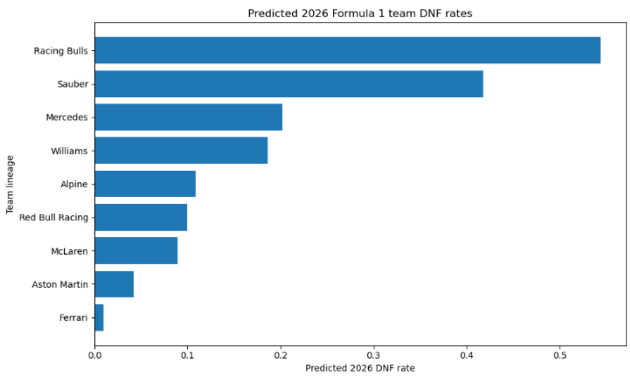

Figure 1 – Bar graph showing the results of predictive analytics using SES, and the output produced for each team in a short-term forecast for the 2026 season [5,8,9,10].

Using SES to produce the results seen in Graph 1, the highest predicted DNF rate for 2026 was Racing Bulls at 0.544, or 54.4%. This is followed by Sauber, at 0.418 or 41.8%. Then an intermediate group was formed with Mercedes and Williams, with 20.2% and 18.6% respectively. The lower forecast range had Alpine (10.8%), Red Bull (10%), and McLaren (9%). Aston Martin and Ferrari formed a surprising outlier group with the lowest forecasts of 4.2% and 1% respectively.

These forecasts didn't make a uniform risk distribution across the whole grid, instead indicating a clear spread with four distinct groups. This supports the idea that the historical results can be used in predictive analytics.

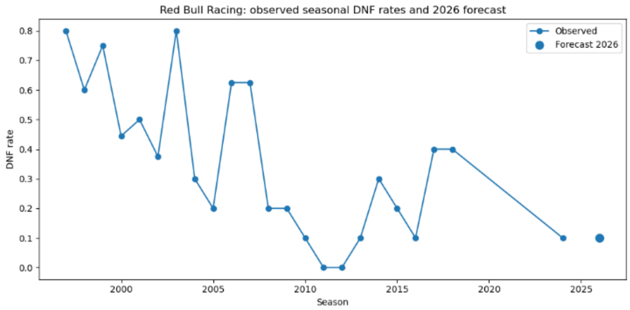

Figure 2 – Graph showing the individual output of a single team (used Red Bull Racing as an example) to show the data measured and the predicted data point created [3,8,9,10].

Due to the nature of this paper being predictive of things that have yet to happen, there is no concrete conclusion. However, the current data produced with this method suggests that predictive analytics can be applied to Formula 1. The results also suggest that the historical data of teams does have an impact on the current events of the sport in a measurable way.

Limitations

This study does have several limitations. First, the forecasts were made using just the historical DNF data, so it ignores things like important reasons for reliability change (such as change in regulation, staffing and driver changes).

This paper also assumes that the modern constructor and the previous team under different owners are highly related. While rare, there are occasions when the team is sold with very few of its previous resources (such as bankruptcy).

Finally, data relating to Haas F1 Team was unable to be produced, and such was not present in this study.

References

[1] Cash, J. (2026) *Motor-racing-Alonso 'couldn't feel hands or feet' during Chinese Grand Prix*. Reuters. [Online]. Available at: https://www.reuters.com/sports/formula1/motor-racing-motor-racingalonso-couldnt-feel-hands-or-feet-aston-martin-retires-2026-03-15/ [Accessed: 3 April 2026].

[2] Loník, I. & Kotrba, V. (2024) *The key importance of spending on research and development in Formula One*, Managerial and Decision Economics, 45(1), pp. 70-79. DOI: 10.1002/mde.3983.

[3] Formula 1 (2026) *The family tree of F1's 11 teams and how they came to be*. Formula 1. Available at: https://www.formula1.com/en/latest/article/the-family-tree-f1-11-teams-and-how-they-came-to-be.2QBA1PPMf0bC8mp2xxqZeq [Accessed: 3 April 2026].

[4] Kanal, S. (2023) *TEAM GUIDE: Aston Martin's complex F1 roots – and how they're aiming to fulfil lofty title ambitions*. Formula 1. Available at: https://www.formula1.com/en/latest/article/team-guide-aston-martins-complex-f1-roots-and-how-theyre-aiming-to-fulfil.465QgZKQ9XcBDcD54VT4xR [Accessed: 3 April 2026].

[5] Jolpica F1 contributors (2026) *Jolpica F1*. Available at: https://github.com/jolpica/jolpica-f1 [Accessed: 16 March 2026].

[6] Hyndman, R.J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, OTexts, Chapter 8.

[7] Holt, C.C. (1957) *Forecasting seasonals and trends by exponentially weighted moving averages*, ONR Memorandum, Vol. 52, Carnegie Institute of Technology, pp. 1-52.

[8] Seabold, S. & Perktold, J. (2010) *Statsmodels: Econometric and statistical modeling with Python*, Proceedings of the 9th Python in Science Conference, pp. 92-96. DOI: 10.25080/Majora-92bf1922-011

[9] statsmodels developers (2026) *statsmodels.tsa.holtwinters.SimpleExpSmoothing*. Available at: https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.SimpleExpSmoothing.html [Accessed: 16 March 2026].

[10] statsmodels developers (2026) *statsmodels.tsa.holtwinters.Holt*. Available at: https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.Holt.html [Accessed: 16 March 2026].

[11] Fédération Internationale de l'Automobile (2024) *FIA 2026 Formula 1 Financial Regulations for F1 Teams*. Available at: https://api.fia.com/sites/default/files/fia_2026_f1_regulations_-_section_d_financial_-_f1_teams_-_iss02_-_2024-12-11.pdf [Accessed: 3 April 2026].