

EPO 182

Formula 1: Analytics behind the tracks to the podium

Pallabi Deb, Piyush Lashkare, Oklahoma State University

Abstract

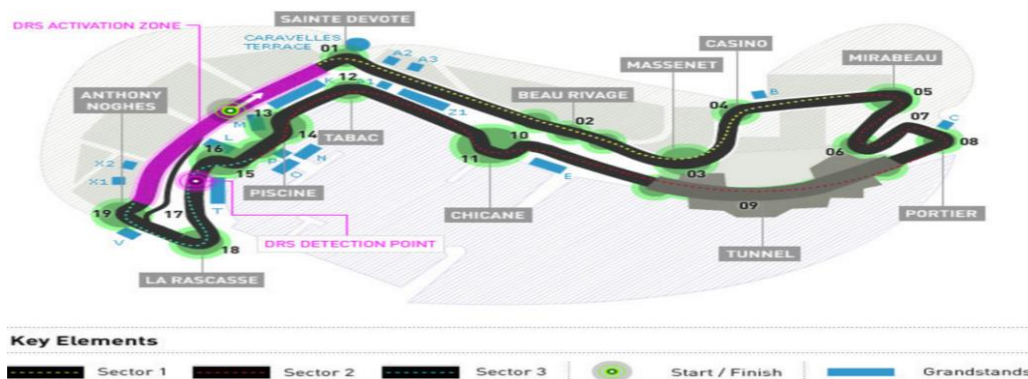
SPEED, POWER, INNOVATION, PERFORMANCE, FACTS, and STATISTICS are the words that distinguish Formula 1 racing, also known as “The Pinnacle of Motor Sport”. It generates yearly income of 1.2 billion dollars and involve many teams competing with their roaring turbocharged engines for the podium. With an average team’s budget of 300 million dollars, teams need not only show their engineering excellence but also a winning strategy. During a race, most drivers have an average heartbeat of 170 per minute, cars often go past the speed of 150 mph and a difference of milliseconds separate winners and losers. These characteristics make it different from most other sports.

Introduction

Today, Formula 1 racing is not only about flawless aerodynamics and legendary driving but is also fueled by DATA to accelerate the analysis. Currently, we have extracted the data for the entire 2011 Monaco Grand Prix race for all the sessions (FP1, FP2, FP3, Qualifying, and Final) for all the constructor teams. This paper reviews the analysis conducted to develop the models and to illustrate the significant attributes such as track conditions, tire types, fuel lap time, pit stops and stint length to predict the probable position of a driver in the final standing charts.

Predictive model criteria:-

- The probability of a driver taking the pole position and how it affects race results.
- What would be the probable race completion position on a particular circuit?
- Did pit stop timings really make a difference in winning?
- Factors affecting final race standings.
- Effect of tire compound on speed.



Data Preparation and Analysis

Official score reports are published by FIA® (Fédération Internationale de l'Automobile) after every race. Required data is extracted from these reports by executing Python® script and Tabula® software. These CSV data files are then sorted and prepared to be fed into SAS® for further analysis.

The final dataset consists of following variables-

Driver Name, Constructor Name, Laps, Tire Type, Number of Laps,
Pole Position, Stints, Tire Type, Pit Stop, Laps Behind,
Fuel Corrected Lap Time, Stop Count, Stop Time, Stopping Lap,
Lap in Stint and Time to Leader.

Fig 2: Data preparation

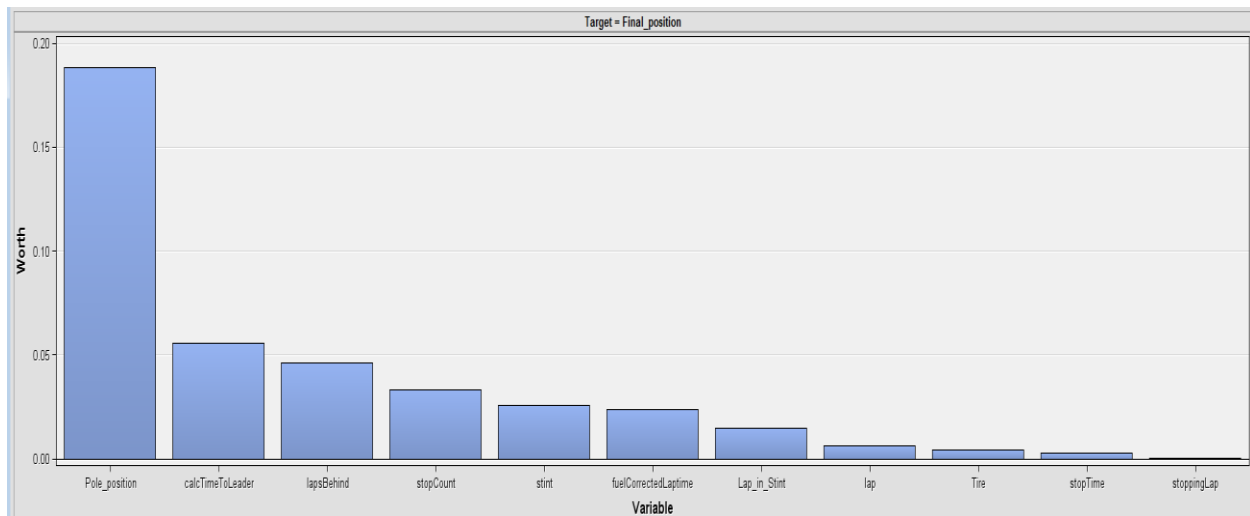
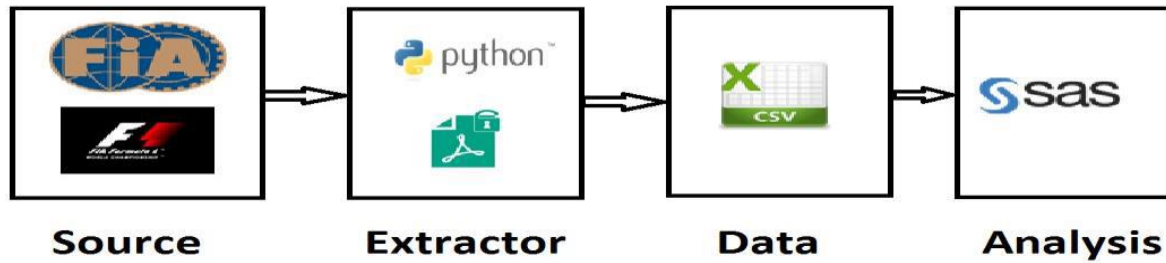


Fig 3: Variable worth

The Variable Worth (Fig 3) plot ranks inputs based on their worth with the target variable (i.e. Final position). We see that pole position has the highest chi-square value followed by time to leader and lap time.

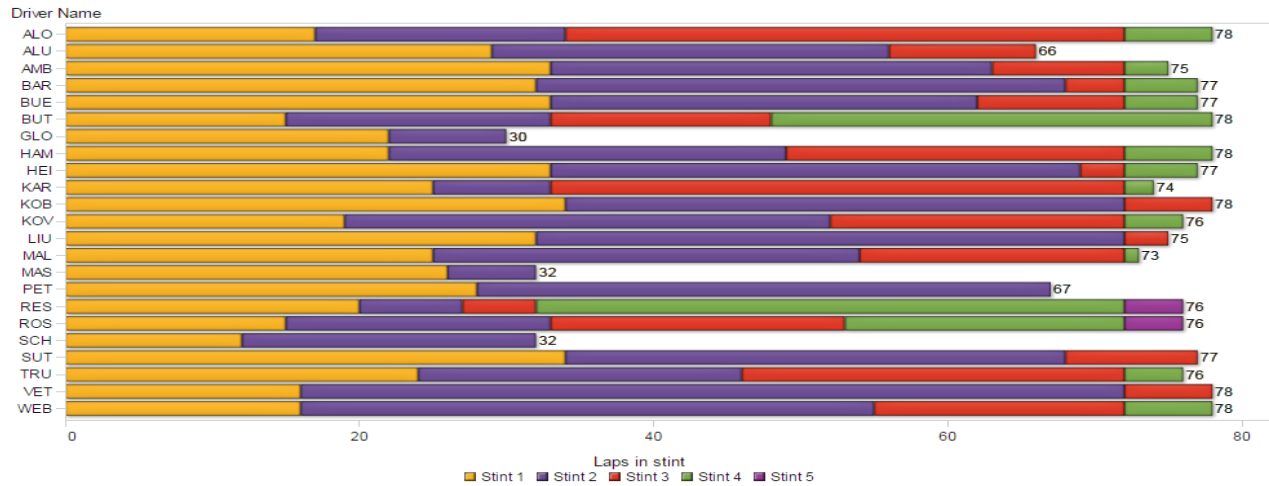


Fig 4: Laps in stint

Laps in stint gives us an overview of constructor's pit stop strategy. We see similar laps in stints for few constructors on both the drivers. (Fig.4)

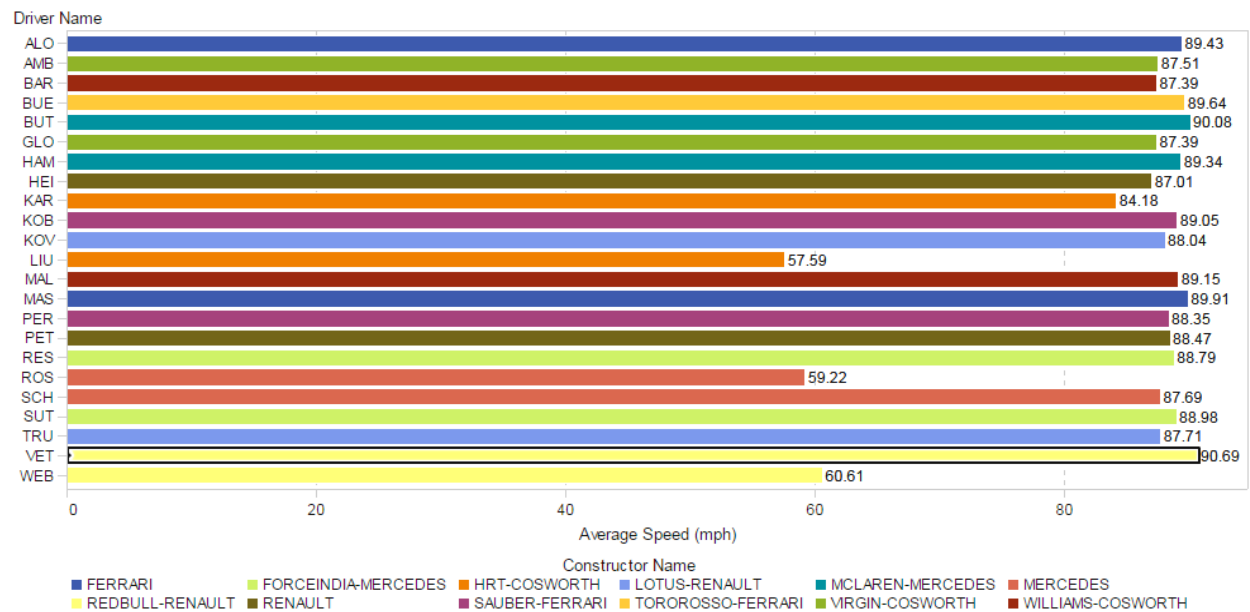


Fig 5: Speed across free practice session

Driver's speed across free practice sessions give us probable standings in the qualifying race and thus, helps us in determining grid positions for the final race. (Fig.5)

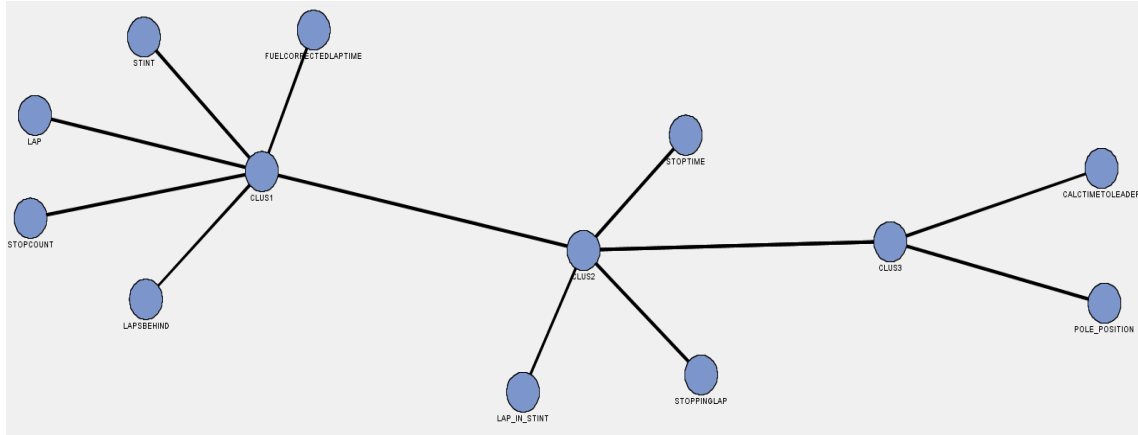


Fig 6: Variable clustering

As we have various variables of similar type, we used PROC VARCLUS to effectively group the variables and the result in Fig.6 has three clusters with relevant variables in each group.

Results

Tire Degradation

The degradation is analyzed using PROC REG procedure, taking into account the delta (difference in lap times) and driver's speed, measured against the response variable Laps.

The calculation yields how long a particular set of tires can be used keeping up with the competitive speed.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	4.34388	4.73303	0.92	0.3589
lap	1	-0.11551	0.10997	-1.05	0.2937

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	94.24840	0.89287	105.56	<.0001
lap	1	-0.13105	0.02101	-6.24	<.0001

Fig 7: Estimates for speed and delta – Soft

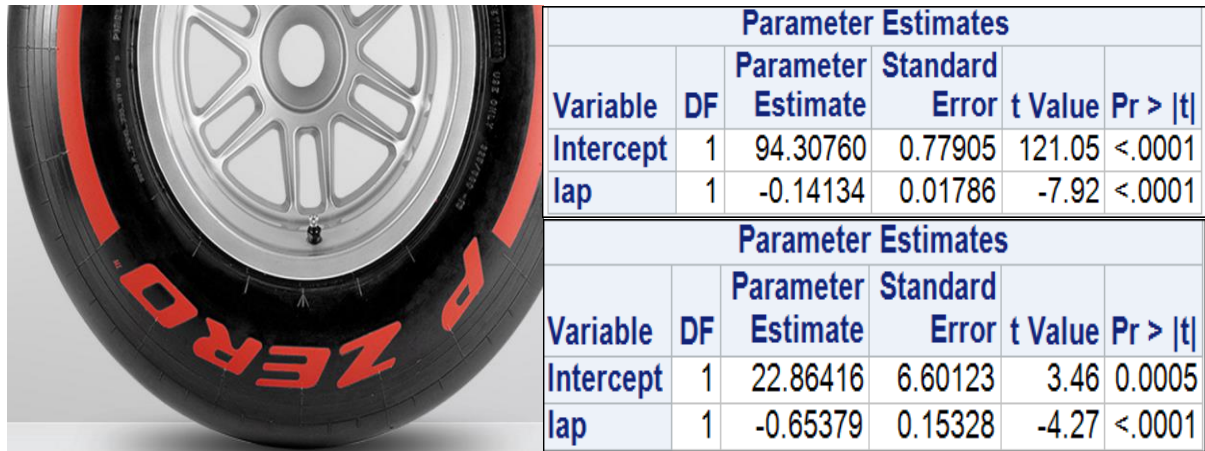


Fig 8: Estimates for speed and delta – Super Soft

Tire life estimation

Considering the maximum and minimum speed achieved, fuel load with time, degradation level, delta time, track temperature we calculated the tire life of soft and super soft tire when used on a competitive speed. Soft compound- 62 laps and Super-soft compound- 44 laps.

The below graphs demonstrate the change in speed and delta as per the laps progress for driver- Lewis Hamilton

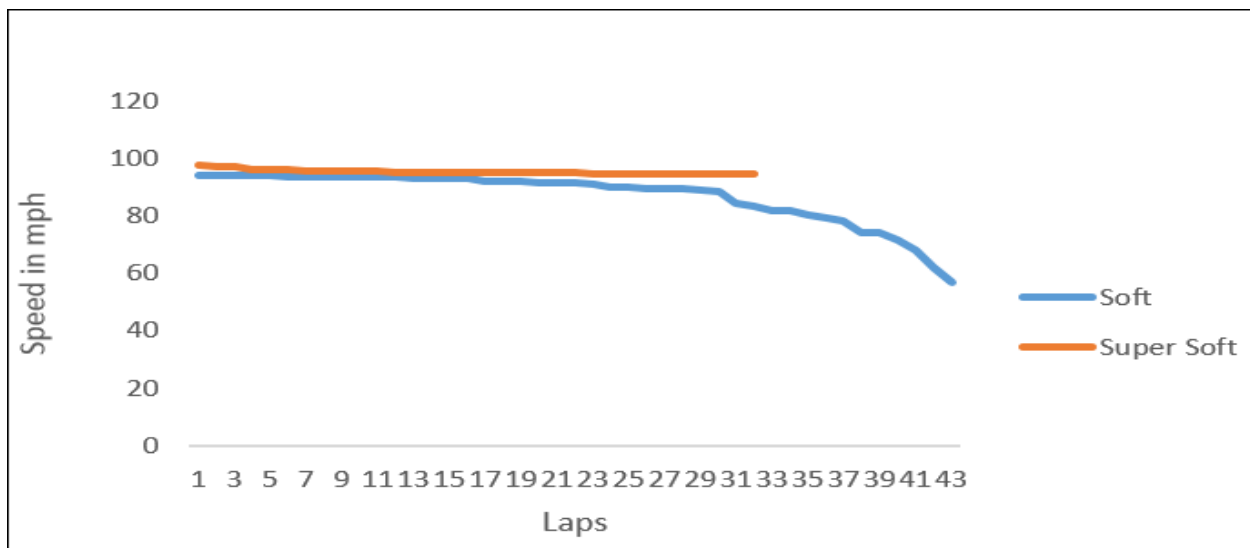


Fig 9: Change in speed over laps

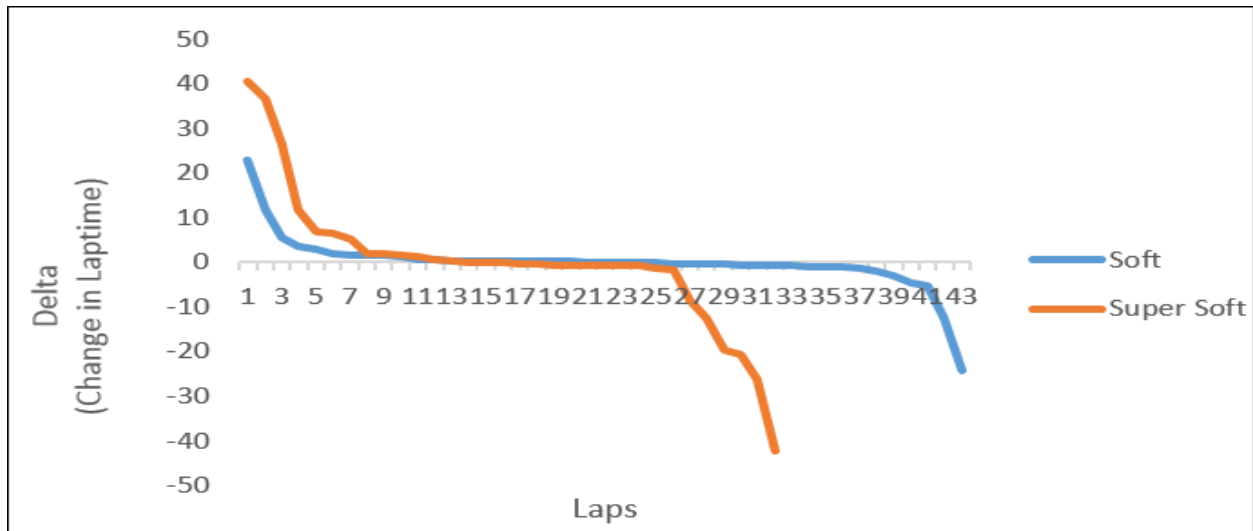


Fig 10: Change in delta over the laps

Race Analysis

Using PROC Logistic, we attained the significant variables that strongly relate to the response variable i.e. final positions. We analyzed the effect of each variable, taking into account its parameter estimates and then studying the trend over the interval.

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
Delta_Soft	17	6593.5743	<.0001
Pole_position	17	1316877.68	<.0001
Tire	12	253.7260	<.0001
fuelCorrectedLaptime	17	37.5862	0.0028
lap	17	180.3871	<.0001
lapsBehind	17	10220.9695	<.0001
stint	17	146.9388	<.0001
Constructor_name*Tire	140	6458094.86	<.0001

Fig 11: Variable Importance

Conclusions

Race Strategy

Analyzing the driver's performance throughout the free practice sessions gives us a comprehensive idea of the probable grid positions for the final race and also the stint strategy over the laps.

Interesting to learn the tire compound characteristics-

Soft- Durable, low degradation, high working range with optimal speed.

Super Soft- High speed, fast degradation, and maximum control over twisted circuits.

The results of tire degradation across the laps with the change of speed gives us an overview to plan a race.

Achievable plan- Degradation pattern unfolds the maximum laps for each tire with an optimum speed over the track under favorable conditions. Plan overlays on using the tires till the end of life, with least pit stops.

Optimal plan- Keeping up with the maximum competitive speed and required pit stops with optimal tire degradation across the laps. Plan overlays least time strategy with maximum achievable speed.

Safety Car- In event of an accident/extremely poor conditions, drivers maintain a defined speed with no over-taking and may not pit, unless it is to change tires. Plan overlays taking pit stops when safety car appears within -5 laps of optimal planned stints.

Suspended Race- Severe conditions necessitate to suspend the race where cars proceed to pit lane without over taking and stop in a queued formation. Plan overlays changing tires that attained the maximum speed and least wear over the previous laps.

From the model, the major factors that contribute to the final position are- delta soft, pole position, tire, fuel corrected lap time, lap, laps behind, stint and the interaction effect of constructor and tire.

Future Work

Building the same model over entire season, would give us additional insights.

Getting the telemetry data would help in understanding the car's and driver's health, thus would contribute to the explanation of the model and then planning a strategy.

References

- [Formula one official website](#) [Pirelli tyre information guide](#)
- [FIA Regulations and glossary](#) [2011 Monaco formula one Wiki page](#)

Acknowledgement

We thank Dr. Goutam Chakraborty and Dr. Miriam McGaugh, Department of Business Analytics, Oklahoma State University for their help and guidance throughout the research.

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Paper Author 1

Name: Pallabi Deb

Enterprise: Oklahoma State University

Address: 127 N Duck St., Apt# F31

City, State ZIP: Stillwater, Oklahoma 74075

Work Phone: 405-334-9404

E-mail: pallabi.deb@okstate.edu

Paper Author 2

Name: Piyush Lashkare

Enterprise: Oklahoma State University

Address: 127 N Duck St., Apt# D24

City, State ZIP: Stillwater, Oklahoma 74075

Work Phone: 716-903-9472

E-mail: piyush.lashkare@okstate.edu