Formula 1 Tire Choice Distribution Across Varying Racing Conditions

Abstract

Using the 2024 Formula 1 season data from the FastF1 repository, we examined how tire compound selection (soft, medium, hard) varied across 1303 race stints when accounting for circuit profile and track conditions. After exploratory analysis and stepAIC variable selection, we compared multinomial logistic regression models on a per-stint basis. Results showed that starting position, average track temperature, and number of turns were the strongest predictors of tire choice. The Generalized Hosmer-Lemeshow test confirmed excellent model fit ($p=0.57$), while ANOVA against the null model demonstrated significant predictive power ($p=9.537\times10^{-7}$). Teams showed a clear preference for medium and hard compounds (80-90% of selections), with soft tires more likely at starting positions further back on the grid and less likely on high-temperature, high-turn circuits. These patterns align with conservative modern F1 strategies, favoring tire longevity over immediate speed. Overall, our model provides a quantitative framework for predicting opponent tire strategies and supporting data-driven race planning in Formula 1.

1. Introduction

Formula 1 (F1) is an international motorsport founded in 1950 that has gained increasing popularity in the past decades due to globalization. It is a quintessential example of engineering, human performance, and data science pushed to the extreme as teams pursue a competitive advantage. At its core, F1 is a data-driven strategic sport in which team success depends not only on engineering but also on the track profile, weather conditions, the qualifying position of drivers, and the racing strategy of other teams [1-3, 5-6].

Our project focuses on one of the most critical variables that decides the outcome of any F1 race: tire selection. There are currently three types of tire compounds: soft, medium, and hard. While teams each have their own real-time data analytics informing them of the best moment to perform a pit stop, public-facing research on overall trends and strategies employed by F1 teams throughout the years has been sparse. Nevertheless, there is abundant static data concerning historical race results, conditions, and team strategies [4]. Therefore, this raises an important question: are there generalized tire strategies employed by Formula 1 teams across key covariates such as circuit profile, track conditions, and starting grid position?

2. Methods

2A. Dataset Description

To assess the relationship between tire-based strategies and race outcomes in Formula 1, a dataset that captures detailed information was required. We thus chose the FastF1 project, a community-maintained open-source Python library that provides access to and tools for analyzing F1 lap times, vehicle telemetry, race times, results, and strategy data. FastF1 serves as a comprehensive, unofficial repository for F1 statistics, as there is no official database for F1 statistics. We focused on the 2024-2025 season which includes 1303 unique stints. Each year brings new teams and strategies, so focusing on recent data will more accurately captures tire selection trends.

For this study, races were categorized by circuit profile and track conditions to assess the significance of tire strategy across different environments. Variables including event information, timing data, and track markers were used to characterize circuit profiles, while weather conditions and race control messages were used to identify track conditions.

2B. Statistical Modeling Framework

As tire type is a nominal variable, we chose to employ the multinomial logistic regression model. This approach was ideal as it allowed us to incorporate multiple explanatory variables to estimate the probability of selecting each specific tire compound. In addition, in recognition of the selection of new tires at each possible pit stop or “stint,” we decided to model our tire selection on a per-stint basis.

We modeled a multinomial logistic framework using separate probability equations for each tire type, with hard tires being treated as the baseline category. This enabled direct interpretation of coefficients in comparison to hard tire selections. We initially considered a broad set of possible predictors from the dataset, including stint length, starting position, distance traveled, and average stint speed. To assess their relevance, we used a combination of plotting key predictors, exploratory visualization, and the stepAIC algorithm to narrow our model down to three explanatory variables: average track temperature, starting position, and number of turns in the track. In Figure 1 we show clear differences in tire compound selection by starting position. From this, we fitted our final multinomial logistic regression.

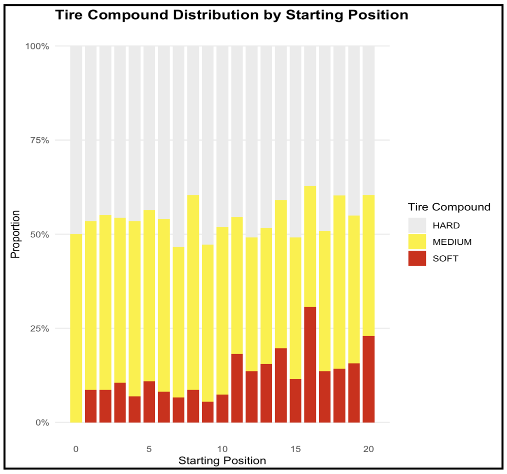


Figure 1. Stacked proportion bar chart of tire compound choice distribution by starting position.

3. Analysis

We developed a stepAIC-refined BCLM multinomial logistic regression model that incorporated our three key predictors: starting position, average track temp in C°, and number of turns. This model strikes a balance between complexity and predictive utility. A central component of our analysis involved evaluating how well this model captured the distribution of tire choices across stints. The Generalized Hosmer-Lemeshow goodness-of-fit test yielded a p-value of 0.57, indicating no evidence of any systematic deviation between the model's predicted probabilities and the observed frequencies of soft, medium, and hard tires. This suggests that the multinomial formulation with the selected predictors accurately aligns with tire choice behavior from our dataset. Furthermore, an ANOVA comparison with the null model produced a significant p-value of 9.537×10^{-7}, demonstrating strong evidence that at least one of the predictors contributes meaningfully to tire selection. These two diagnostics show that our three-predictor model provides substantial predictive power in estimating what tire compound an F1 team will choose for any given stint or track conditions.

4. Discussion

Our analysis showed that largely medium and hard tires are predicted to be selected, as evidenced by Figure 2:

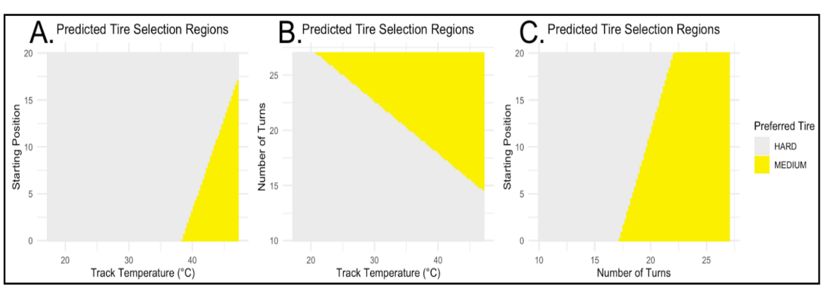


Figure 2. A. Predicted tire selection regions across starting position and average track temperature (number of turns held at median = 16). **B**. Predicted tire selection regions across number of turns and average track temperature (starting position held at median = 11). **C**.Predicted tire selection regions across starting position and number of turns (average track temperature held at median = 35.86)

While soft tire compounds are picked between 10-20% of the time, depending on the track conditions, there is a clear bias towards medium and hard tire compounds, which aligns with modern racing strategies that lean towards selecting more efficient, longer-lasting tires. Soft tires are known to be the fastest, yet wear out the quickest, thus making them a riskier choice.

Our analysis also indicates under which conditions certain tires are more likely to be selected. With reference to table 1 and because our model uses hard tires as a baseline, in comparison to hard tire compounds, soft tires are more likely to be selected by drivers starting further down the starting grid (indicating a need for more immediate speed when behind), and are less likely to be selected the more turns and the higher track temperature there is (indicating those conditions possibly wear soft tires even faster). The medium and hard tires comparison shows the opposite pattern, with a negative coefficient for starting position and positive coefficients for increased number of turns and temperature, perhaps indicating a preference towards a more "balanced" strategy in more complicated race tracks.

Coefficients	Intercept	starting_position	avg_track_temp_c	n_turns
MEDIUM	-1.1330182	-0.008094655	0.01573264	0.033160786
SOFT	-0.9431744	0.06189879	-0.02692431	-0.006040924

Table 1. BCLM Model Coefficients

Finally, our model can take any current or future race and predict the selected tire type based on their starting position, the number of turns within the race, and the track temperature that day. One such example is showcased in the appendix below, which explores hypothetical future track and race conditions by varying the number of turns and the average track temperature for that day. The data indicates a preference towards hard compounds at a lower number of turns, while medium tire compounds become the most popular prediction with a large (>21) number of turns. Higher temperatures also indicate a trend towards medium tires, with the highest probability occurring when both a large number of turns and high temperatures are present. Furthermore, a more forward starting position suggests a greater likelihood of choosing medium tires, while hard tires become more common for starting positions further back.

That being said, there are several limitations that may affect those exploring similar topics. Several variables could have been included, such as stint length, safety-car periods, team-specific strategies, and other factors that either are not collected in data or are impossible to completely encapsulate. Due to constraints in our regression modeling capabilities, we opted for a simpler model, which limited our ability to capture those variables. We also recognized that our data collection used the starting position, but a model that updated the position as the race went on would be more efficient for predicting tire compound selection later in the race. Finally, model accuracy was constrained by the limited amount of data for specific tire types. The data presented had a large selection of hard and medium tire compounds, yet rarely selected soft, wet, or intermediate compounds. The latter two lacked sufficient evidence to make our model, and the soft compound was infrequently used enough that the model rarely predicted it as most likely.

Although our model was inherently simplistic by design, it revealed patterns that align with real F1 team strategies. F1 is a sport where experience and intuition are critical when making decisions such as tire choice or pit stop timing. Despite this, quantitative models such as ours can be critical in analyzing and predicting other teams' behavior and adapting one's own strategies as a result. More analyses like these can help F1 teams make optimal decisions in a data-driven fashion.

5. References

1. **Brusik, F.** (2024). *Predicting lap times in a Formula 1 race using deep learning algorithms: A comparison of univariate and multivariate time series models* [Master's thesis, Tilburg University]. Tilburg University Research Portal. https://arno.uvt.nl/show.cgi?fid=180319
2. **Fry, J., et al.** (2024). Faster identification of faster Formula 1 drivers via time-rank duality. *Economics Letters*, *235*, Article 111542. https://doi.org/10.1016/j.econlet.2024.111542
3. **Msakamali, B.** (2024). *F1 data analysis and tactical insights: Exploring Formula 1 race performance strategies* [Bachelor's thesis, Tampere University of Applied Sciences]. Theseus. https://www.theseus.fi/bitstream/handle/10024/856650/Msakamali_Baraka.pdf?sequence=6
4. **Oehrly, T.** (2020). *FastF1* [Computer software]. GitHub. https://github.com/theOehrly/Fast-F1
5. **Patil, A., Jain, N., Agrahari, R., Hossari, M., Orlandi, F., & Dev, S.** (2023). A data-driven analysis of Formula 1 car races outcome. In L. Longo & R. O'Reilly (Eds.), *Artificial intelligence and cognitive science* (Vol. 1662, pp. 136–148). Springer. https://doi.org/10.1007/978-3-031-26438-2_11
6. **van Kesteren, E.-J., & Bergkamp, T.** (2023). Bayesian analysis of Formula One race results: Disentangling driver skill and constructor advantage. *Journal of Quantitative Analysis in Sports*, *19*(3), 165–183. https://pmc.ncbi.nlm.nih.gov/articles/PMC10660124/

AI Statement

AI was not used in this submission.

6. Appendix

Baseline Categorical Logits Model Code and Analysis

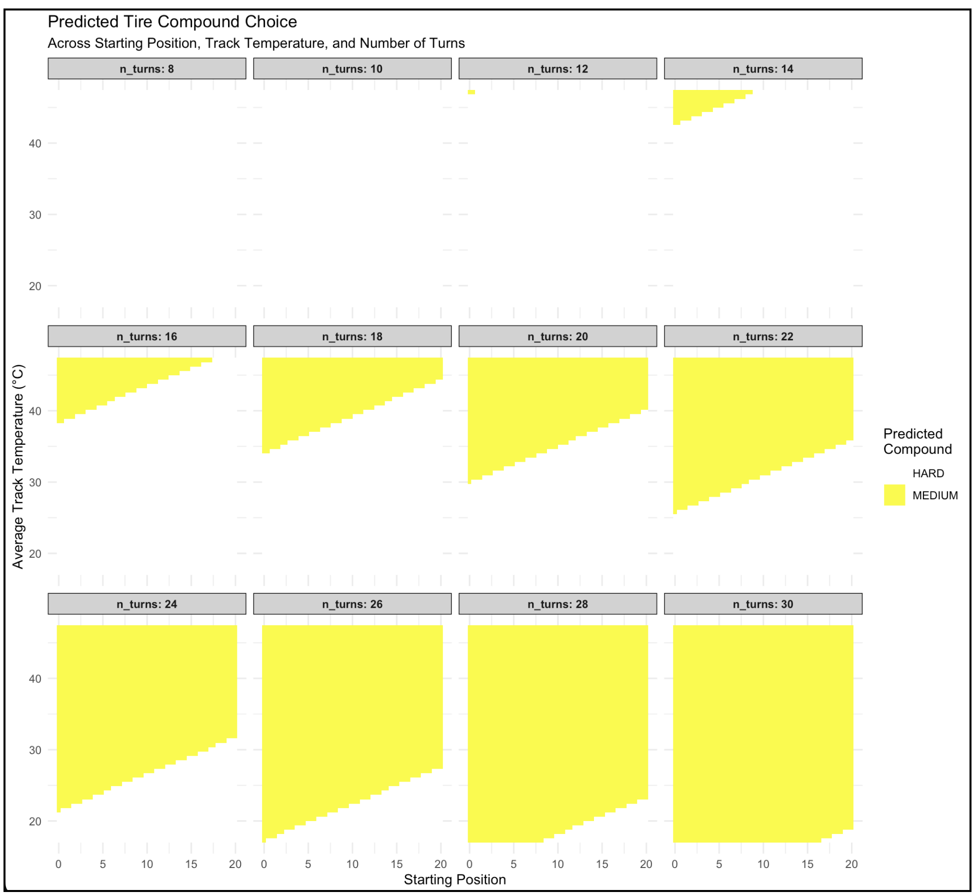
```
Model:
multinom(formula = compound_label ~ starting_position + avg_track_temp_c + n_turns,
         data = stints_clean)

Coefficients:
       (Intercept) starting_position avg_track_temp_c      n_turns
MEDIUM  -1.1330182      -0.008094655       0.01573264  0.033160786
SOFT    -0.9431744       0.061896879      -0.02692431 -0.006040924

Std. Errors:
       (Intercept) starting_position avg_track_temp_c    n_turns
MEDIUM   0.4444886        0.01108952      0.007134742 0.01901775
SOFT     0.6599580        0.01686937      0.010043086 0.03026212

Residual Deviance: 2247.026
AIC: 2263.026
```

Predicted Tire Choice Over Various Starting Positions, Track Temperatures, and Number of Turns in a Track



Tire Compound Distribution By Starting Position Figure Code

```
library(ggplot2)
library(scales)

# Create stacked bar plot
ggplot(stints_clean, aes(x = starting_position, fill = compound_label)) +
  geom_bar(position = "fill", width = 0.8) +
  scale_fill_manual(values = c("HARD" = "#EBEBEB",
                               "MEDIUM" = "#FFF200",
                               "SOFT" = "#DC0000"),
                    name = "Tire Compound") +
  scale_y_continuous(labels = percent) +
  labs(title = "Tire Compound Distribution by Starting Position",
       x = "Starting Position",
       y = "Proportion") +
  theme_minimal() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(face = "bold", size = 14))
```

Predicted Tire Selection Regions R Code

```
library(ggplot2)
library(dplyr)
library(patchwork)

# Plot 1: Starting Position vs Track Temperature (hold n_turns at median)
grid1 <- expand.grid(
  starting_position = seq(min(stints_clean$starting_position),
                          max(stints_clean$starting_position),
                          length.out = 100),
  avg_track_temp_c = seq(min(stints_clean$avg_track_temp_c),
                         max(stints_clean$avg_track_temp_c),
                         length.out = 100),
  n_turns = median(stints_clean$n_turns)
)
grid1$predicted_compound <- predict(model_full, newdata = grid1, type = "class")

p1 <- ggplot(grid1, aes(x = avg_track_temp_c, y = starting_position,
                        fill = predicted_compound)) +
  geom_tile() +
  scale_fill_manual(values = c("HARD" = "#EBEBEB",
                               "MEDIUM" = "#FFF200",
                               "SOFT" = "#DC0000"),
                    name = "Preferred Tire") +
  labs(title = "Predicted Tire Selection Regions",
       x = "Track Temperature (°C)",
       y = "Starting Position") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

###Continue below
```

```
# Plot 2: Number of Turns vs Track Temperature (hold starting_position at median)
grid2 <- expand.grid(
  starting_position = median(stints_clean$starting_position),
  avg_track_temp_c = seq(min(stints_clean$avg_track_temp_c),
                        max(stints_clean$avg_track_temp_c),
                        length.out = 100),
  n_turns = seq(min(stints_clean$n_turns),
               max(stints_clean$n_turns),
               length.out = 100)
)
grid2$predicted_compound <- predict(model_full, newdata = grid2, type = "class")

p2 <- ggplot(grid2, aes(x = avg_track_temp_c, y = n_turns,
                       fill = predicted_compound)) +
  geom_tile() +
  scale_fill_manual(values = c("HARD" = "#EBEBEB",
                              "MEDIUM" = "#FFF200",
                              "SOFT" = "#DC0000"),
                   name = "Preferred Tire") +
  labs(title = "Predicted Tire Selection Regions",
       x = "Track Temperature (°C)",
       y = "Number of Turns") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

# Plot 3: Starting Position vs Number of Turns (hold avg_track_temp_c at median)
grid3 <- expand.grid(
  starting_position = seq(min(stints_clean$starting_position),
                         max(stints_clean$starting_position),
                         length.out = 100),
  avg_track_temp_c = median(stints_clean$avg_track_temp_c),
  n_turns = seq(min(stints_clean$n_turns),
               max(stints_clean$n_turns),
               length.out = 100)
)
grid3$predicted_compound <- predict(model_full, newdata = grid3, type = "class")

p3 <- ggplot(grid3, aes(x = n_turns, y = starting_position,
                       fill = predicted_compound)) +
  geom_tile() +
  scale_fill_manual(values = c("HARD" = "#EBEBEB",
                              "MEDIUM" = "#FFF200",
                              "SOFT" = "#DC0000"),
                   name = "Preferred Tire") +
  labs(title = "Predicted Tire Selection Regions",
       x = "Number of Turns",
       y = "Starting Position") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

combined_plot <- p1 + p2 + p3 + plot_layout(guides = "collect")

combined_plot
```

Predicted Tire Choice Over Various Starting Positions, Track Temperatures, and Number of Turns in a Track R Code

```
library(ggplot2)
library(dplyr)

# Create prediction grid across all three predictors
grid <- expand.grid(
  starting_position = seq(min(stints_clean$starting_position),
                          max(stints_clean$starting_position),
                          length.out = 50),
  avg_track_temp_c = seq(min(stints_clean$avg_track_temp_c),
                         max(stints_clean$avg_track_temp_c),
                         length.out = 50),
  n_turns = seq(8, 30, by = 2)  # Creates plots for 8, 10, 12, ..., 30 turns
)

# Get predictions
predictions <- predict(model_full, newdata = grid, type = "class")
grid$predicted_compound <- predictions

# Create faceted plot
ggplot(grid, aes(x = starting_position, y = avg_track_temp_c,
                 fill = predicted_compound)) +
  geom_tile() +
  facet_wrap(~n_turns, labeller = label_both, ncol = 4) +
  scale_fill_manual(values = c("SOFT" = "#FF0000",
                               "MEDIUM" = "#FFFF00",
                               "HARD" = "#FFFFFF"),
                    name = "Predicted\nCompound") +
  labs(title = "Predicted Tire Compound Choice",
       subtitle = "Across Starting Position, Track Temperature, and Number of Turns",
       x = "Starting Position",
       y = "Average Track Temperature (°C)") +
  theme_minimal() +
  theme(strip.background = element_rect(fill = "lightgray"),
        strip.text = element_text(face = "bold"))
```