Real-Time Decision Making in Motorsports: Analytics for Improving Professional Car Race Strategy

by
Christopher Ledesma Weisen Choo

Bachelor of Science in Information Systems Management
Singapore Management University, 2007

Master of Computing in Infocomm Security
National University of Singapore, 2013

Submitted to the System Design and Management Program in
Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management
at the
Massachusetts Institute of Technology

June 2015

Author's signature
……………………………………………………………………………………………………
System Design and Management Program
May 21, 2015

Certified by
……………………………………………………………………………………………………
Cynthia Rudin
Thesis Supervisor
Associate Professor of Statistics, CSAIL and Sloan School of Management

Accepted by
……………………………………………………………………………………………………
Patrick Hale
Senior Lecturer, Engineering Systems Division
Executive Director, System Design and Management Program

To Meryl and Zachary, in our lifelong pursuit of happiness.

To my Dad Malcolm, for introducing me to the world of motorsports.

Real-Time Decision Making in Motorsports: Analytics for Improving Professional Car Race Strategy

by
Christopher Ledesma Weisen Choo

Submitted to the System Design and Management Program on May 21, 2015
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

Abstract

We discuss features contained in a machine learning software developed at MIT for professional car racing, to improve the predictions of track position changes within a race. We study pit crew performance and driver performance within selected races, and find that good combined performance for both correlates to better finish positions.

Secondly, we classify tracks based on tire wear and the ratio of 2 versus 4 tire change decisions for pit stops. We find that a driver's performance in early stages of the race is similar to performance in later stages, suggesting that final race outcomes may be inferred from earlier stages of the race.

Thirdly, we look at how tire change decisions vary from track to track depending on tire wear, caution periods, and stages of the race to understand how teams adapt their tire change strategies as each race progresses. We propose heuristics based on these observations that may be used to improve the software.

Next, we test whether the construction of the machine learning dataset using similar and different track characteristics has a discernable impact on the predictive capability of the software. Our tests indicate that it may be preferable to aggregate different races together because there is no distinct difference in the results when compared to only selecting similar races.

Finally, we cover ideas about how new features could be implemented in the software, and touch on other factors affecting pit stop strategy in the quest for better predictive capability in the software.

Thesis Supervisor: Cynthia Rudin
Title: Associate Professor of Statistics, CSAIL and Sloan School of Management

Acknowledgements

This thesis is the culmination of several months of work done with Prashan Wanigasekara, with whom I spent many weekends and nights pouring through MATLAB code and piecing together statistics from the wealth of data we had. Prashan was instrumental in encouraging me to take up this topic. I hope that my insights from the world of Formula One have helped to guide some of our research directions for this project, as well as the work that lies ahead.

I would also like to thank Cynthia Rudin, the advisor for this thesis. Prashan and I spent many hours with Cynthia over our regular fortnightly meetings to dissect and analyze data from previous races. Her guidance, especially in the realm of statistical analysis, helped to strengthen the foundation for the features that we worked on for the machine learning model, as well as the experiments that we ran thereafter.

Thanks are also due to Theja Tulabandhula, the original developer of the machine learning software. Prashan and I exchanged many ideas with him over e-mail, and his pointers were critical in helping us understand the strengths of the software and potential for further improvement.

Matt, the program engineer who worked with us on this project, has also been an excellent source of technical insight into the rules and norms of NASCAR racing. His passion and dedication to the sport are something that all motorsports enthusiasts should aspire to emulate.

I would also like to thank the Singapore University of Technology and Design, which kindly sponsored my program at MIT. I look forward to contributing positively to the research environment back at SUTD when I return to Singapore a few weeks after the conclusion of this thesis.

Next, my career in motorsports and subsequent enrollment at MIT would not have been this possible if not for the support of my mentors Colin Syn from Singapore GP, Leong Yue Kheong from the Singapore Tourism Board, and Lawrence Foo. The Formula One Singapore Grand Prix is an amazing spectacle that I am proud to have been a part of for so many years. I am eternally thankful for their guidance, and wish the Singapore Grand Prix many more successful years.

Finally, I would like to thank Andrew James and Stephen Taylor from Formula One Management. The brief discussions we had during the 2014 Singapore Airlines Formula One Singapore Grand Prix were particularly insightful in developing some ideas that I used in this thesis. I also thank Douglas Hamilton, a brilliant classmate in the MIT SDM program who gave me much-needed inspiration on how to structure this thesis.

Contents

1. Introduction

1.1 United States Professional Car Racing

The National Association for Stock Car Auto Racing (NASCAR) runs a series of motorsports events in the United States. The most prestigious of these is the Cup Series, which is in its 67th season as of 2015. The series comprises a field of up to 43 stock cars powered by V8 engines, and drivers compete in up to 36 races held throughout the country, mostly on oval tracks.

Cars must regularly pit to refuel and change tires throughout each race. Decisions on when to do so depend on factors such as tire wear, fuel consumption, the driver's race strategy in relation to his rivals, and most importantly, unique to NASCAR, the incidence of caution flags waved frequently throughout the race.

Similar to other racing categories, caution flags are waved when there are hazards on the track. They could be caused by accidents, loose debris that affect the safety of the cars, spillage causing slippery surfaces, light rain, intrusions onto the track, or scheduled yellow flags. In such scenarios, cars are required to slow down and follow a pace car that is dispatched to bring the race under control while the hazard is attended to.

Typically, because racing effectively stops when caution flags are waved, teams take the opportunity to call drivers into the pits to change tires and refuel their cars. Tires must be changed regularly because of rapid degradation under racing conditions – this rate of degradation varies from track to track.

A number of other factors also affect pit stop strategy. For example, cars that delay their pit stops may end up at the front of the field if they are willing to sacrifice putting on fresh tires. This happens in part because of the scoring system.

Drivers who win a race are awarded 43 points, with each successive driver receiving 1 point less than the one ahead of him or her. Winners of each race also receive 3 bonus points. In addition, 1 bonus point is awarded to any driver who has led any one lap, and another 1 bonus point is awarded to the driver who has led the most laps in the race. In total, drivers may score a maximum of 48 points in a race – 43 for finishing first, 3 for winning the race, 1 for leading a lap, and 1 for leading the most number of laps. The minimum points a driver can score is 1 point, for finishing last.

The variety of different race strategies under different racing conditions makes it difficult for NASCAR crew chiefs to determine what strategy their drivers should take to maximize the points they get from each race. Essentially, crew chiefs must ensure that their drivers remain as far in front as possible, while having their race cars sufficiently fueled and outfitted with fresh tires.

Decisions that crew chiefs make during the race are based on intuition that is informed largely by their historical understanding of previous races, their drivers' performance in the days leading up to the race such as during the practice and qualifying sessions, and real-time information that is streamed to them during the race itself. Balancing the pros and cons of their options, such as deciding to change 2 tires instead of 4 is done on-the-fly, without a significant amount of software tools. This is in order to get their driver ahead of the pack, but at the cost of increased tire wear.

1.2 Data Analytics and Machine Learning

Data analytics can be broadly categorized into three types (Gartner, 2014). The first is descriptive analytics, which illustrates trends based on historical results. The second is predictive analytics, which uses those trends to predict outcomes for the future. The third is prescriptive analytics, which recommends actions that one should take to influence outcomes.

Top Formula One teams such as McLaren already use data analytics intensively (MeKenna, 2013) , but this is not prevalent in NASCAR because of regulations limiting the amount of technology that can be installed on the cars, such as the use of sensors and detailed telemetry that sends diagnostic information from the car to the garages during a race. To work within existing regulations, NASCAR teams are focusing on different technologies that help improve race outcomes without mounting sophisticated sensors on their cars.

For example, Michael Waltrip Racing works with a company known as Zebra for deploying RFID technology for pit crew training (Hartigan, 2014). However, such efforts are typically limited to off-track situations, outside of races. NASCAR has strict regulations controlling the amount of technology that can be used on race cars, and items such as tire sensors are prohibited for use during race weekends (Spencer, 2014).

Machine learning is one method in which computers can assist with decision-making during a NASCAR race. Underpinning this is the reliance on pattern recognition, particularly in the area of using decisions in previous races to enhance crew chiefs' decision-making capabilities during the course of current and future races. In essence, the computer can act as a quick-thinking advisor to the crew chief, using inputs from previous races and the live data stream from the current race. These inputs are structured in a way that allows the computer to detect patterns within the data, using institutionalized knowledge that race engineers have developed over the years.

1.3 Motivation

Data analytics has become an increasingly important part of motorsports, especially because teams and drivers need to continually make decisions during a race. How many tires should be changed during a pit stop? What race strategies are rivals going to take? What is the projected performance of each car for the duration of the race, based on their historical and current on-track performance? These questions need to be answered continually as the race progresses.

Although big data analytics is prevalent in a motorsports series such as Formula One, there are aspects exclusive to NASCAR that make the wholesale import of the Formula One approach infeasible. For instance, while race teams in the 2014 Formula One United States Grand Prix collected approximately 243 terabytes of data (Bi, 2014), each of them drawing data from a variety of sensors installed on the car and along the race track, NASCAR teams are unable to do so. Moreover, budgets in NASCAR are much more constrained, with teams running on an estimated 10 to 20 percent the budget of a top Formula One Team (Stevens, 2013).

In terms of race strategy, tire wear is particularly important in NASCAR racing because of the car's high weight-to-tire-size ratio. NASCAR cars weigh more than double that of IndyCar and Formula One cars and they cover much longer race distances. A typical NASCAR race could run a distance of 500 miles, compared to just 190 miles in Formula One. As a result, there is a high premium placed on tire wear and tire management.

Furthermore, the rules for the NASCAR Cup Series are different from other motorsports races. Caution flags are waved much more frequently, resulting in cars pitting mostly during caution laps instead of when the race is in full swing. Drivers are also awarded a bonus point for leading any one lap of the race, tempting some to remain on track while their rivals pit for fresh tires and refuel.

Moreover, the decision to pit for either 0, 2, or 4 tires is also a special feature in NASCAR, and only 6 pit crew are allowed to work on the car. This substantially slows down tire changes compared to the typical 3-second pit stop in Formula One, where about 21 pit crew are involved. In conjunction with that, the decision to pit for either 0, 2, or 4 tires means that the time taken for each car during a pit stop could vary by several seconds. This results in an adversarial scenario where teams must make calculated decisions based in part on what their rivals do in the pit.

Making real-time decisions with several considerations in mind is not straightforward as cars can lap a NASCAR track in as little as 19 seconds. Such is the case in Southeast_E, which is NASCAR's shortest track with a lap distance of 0.526 miles. On such a track, the time between a caution flag and when the pit road is open to receive cars can be particularly brief. This makes it difficult for NASCAR teams to perform detailed calculations to determine their optimal tire change strategy.

Collectively, these factors make it useful to incorporate some form of machine learning to enable crew chiefs to make more effective decisions. Machine learning algorithms are able to crunch massive amounts of data from previous races, analyze current trends, and incorporate new data from ongoing races to generate predictions for expected changes in track position based on pit stop strategy. In addition, these insights can be calculated during a race session. The software essentially acts as an assistant to the crew chief, who needs all the help he or she can get to make the final call.

1.4 Primary Research Objectives

This thesis focuses primarily on generating useful prescriptive analytics by developing descriptive and predictive analytics using data from previous seasons of the NASCAR Cup Series. These results are then formatted and implemented in software developed by MIT for an organization involved in the series. Currently, the software runs a variety of machine learning algorithms on data from the 2012 to 2014 NASCAR Cup races. Predictions are made over the course of a race, and the results are compared against the actual race outcomes to determine the accuracy of those algorithms.

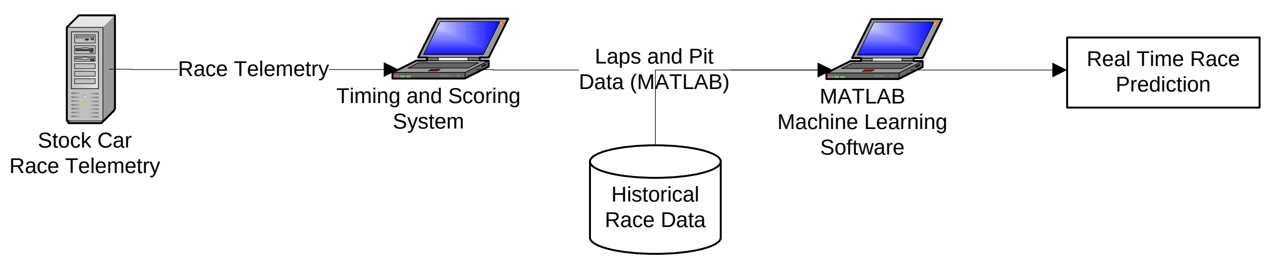


Figure 1. Components of the Machine Learning Software and Flow of Data

Further work needs to be done to convert these predictions into actionable decisions. Fundamentally, there are three key questions that need to be answered as a race progresses:

1. When should a driver pit during the course of a race?
2. How many tires should be changed during that pit stop?
3. If a driver pits and makes a certain tire change decision, what is the predicted change in track position before the next pit stop?

Because of the adversarial relationship between the driver and his or her rivals, a supplementary question that is useful for decision-making should also be considered:

1. What are the likely pit stop decisions made by the driver's rivals?

2. Current Work and Literature Review

2.1 Research Methods & Approaches

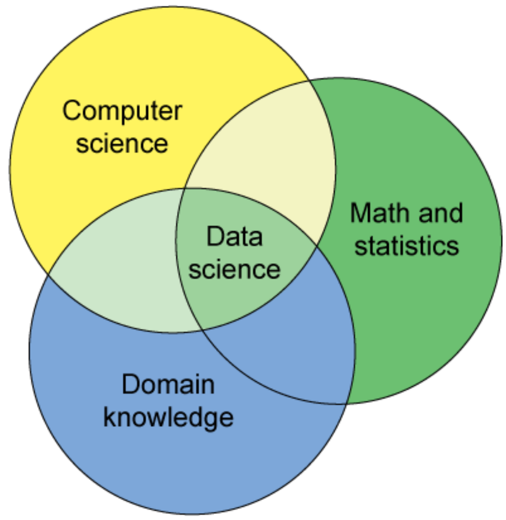


Figure 2. Data Science as a combination of Mathematics, Domain Knowledge, and Computer Science (Jones, 2013)

The successful application of machine learning in the NASCAR context requires a combination of distinct skillsets. In any data science project, domain knowledge, mathematics and statistics, and computer science work hand-in-hand, as shown in Figure 2. In this project, we combine the skills of a mathematics professor, a master's candidate in computer science, and a race circuit project manager to jointly work on the problem.

There are several tasks that have already been completed. For instance, a working machine learning software has already been developed in MATLAB (Tulabandhula & Rudin, 2014). In that piece of software, certain race dynamics such as the "fresh air effect", where drivers lap quicker at the front and at the back of the field, has been considered. However, although several race characteristics have been factored into the model, more can be done to improve the predictive capability of the software because there are many more parameters that influence changes in track position.

The first step in our research is to account for the factors that determine how tire decisions are made during pit stops in a NASCAR race. This is done by watching NASCAR races, discussing race strategy with our industry sponsor for the project, reading research material, and reviewing features already included in the machine learning software.

Secondly, further research is done to understand additional factors that influence pit stop and tire change strategy, in the context of making improvements to the machine learning software. In particular, part of the model used in the current software written for the NASCAR team is documented. This involves reading the code, tracing the program's functions, and documenting its processes and logic. Next, we look at how tire change decisions are made in some races, and attempt to explain the reasons for those decisions. Combining insight from both the software's predictions and actual races, we are better able to understand the strengths and limitations of the software.

Thirdly, we will identify areas for further improvement in the software. We will run tests against our assumptions about dataset construction and its relation to predictive capability of the software. We also propose new features for the software, taking into account additional data gleaned from the most recent NASCAR season as well as those from previous seasons.

2.2 Factors Influencing Racing Performance

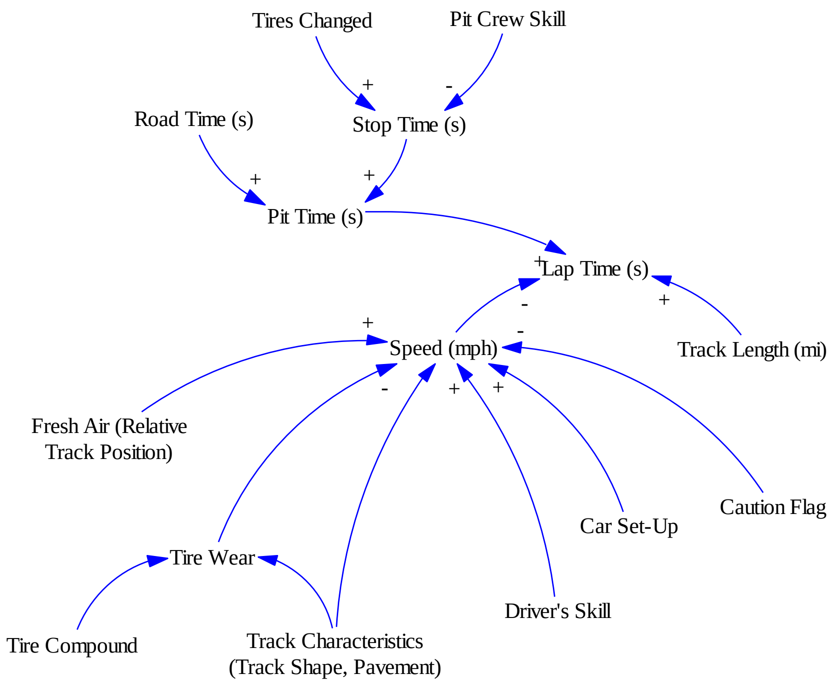


Figure 3. System Dynamics Diagram of Race Characteristics

Figure 3 shows a basic system dynamics model that maps the various parameters of a race and how they work together to influence a driver's lap time. Generally speaking, the driver with the lowest cumulative total lap time throughout the course of a race will emerge the victor, and those with higher cumulative lap times will fall behind the winner in order of their cumulative race timing.

It is important to decompose the cumulative lap times into individual laps, because lap times are recorded and reported to the team garages every time a driver completes a lap. We then proceed to break down lap times into its respective components:

$$Lap\ Time = \frac{Track\ Length}{Speed} + Pit\ Time$$

Track length is usually measured in miles, and speed is measured in miles per hour. We need to also account for Pit Time, which is the time taken for a car to drive into the pits (road time) and change its tires (stop time). This time is added to the Lap Time whenever a car leaves the track and enters the pit. Although this point appears obvious, lap times must be represented in a way that allows different races to be compared against each other within the machine learning model.

The speed that a driver is able to achieve during each lap depends on a variety of factors:

1. Fresh Air (Relative Track Position)
2. Driver's Skill
3. Car Set-Up
4. Track Characteristics
5. Tire Wear
6. Caution Flags

Fresh air refers to the track position of the car in relation to the rest of the field. Cars at the front and at the back of the pack tend to achieve higher speeds because of the larger gaps between them and their neighbors. Small gaps result in aerodynamic disadvantages because the air flow around the cars becomes more turbulent due to the close proximity of their neighbors and the interaction of cars with the surrounding airflow. Conversely, larger gaps are prevalent nearer the front and the back of the field, resulting in cleaner air that allows drivers to race faster.

A driver's skill is another factor in determining the speed that he or she can lap the track. Better drivers are able to drive faster because of their ability to maneuver around corners quickly and hit optimal speeds along the straight portions of the track. Due to the duration of a NASCAR race, a driver also needs endurance to maintain high levels of concentration throughout the race, which often runs for 3 hours.

The car set-up is also crucial in determining performance. Generally speaking, only the top 10 to 15 drivers have a realistic chance of winning a race. Even though NASCAR cars are meant to be fairly standardized, regulations allow for many tweaks that allow a better-financed team to build more improvements into the car. Teams that develop their cars better are able to build sustainable advantages that are carried on to subsequent races in the season.

Next, the track characteristics determine the speed at which drivers can lap the track. Oval tracks such as the Southeast_C are faster when compared to road courses such as the West_B and the Northeast_A. Drivers are able to hit much higher speeds on oval tracks because the tracks have fewer turns and are banked around the corners, compared to road courses which have many turns and flatter corners. Track characteristics also affect tire wear. Specifically, coarser and older asphalt pavements result in higher tire degradation, compared to pavements constructed with finer aggregate and are repaved more recently.

Tire wear also affects a driver's speed. Tires wear out at different rates based on the characteristics of the track's pavement. In conjunction with that, the tires deployed for the race are made up of different compounds, and affect how quickly the tires degrade.

Lastly, NASCAR has a relatively high frequency of caution flags, which denote the period in which a pace car is deployed onto the track. This brings the race under control for officials to clear debris from the track, to make it safer for racing. The extremely high lap times clocked under cautions encourage drivers to pit as soon as possible, as the positive effect of fresh tires far outweighs the time lost for the pit stop.

Together, these race characteristics act in concert with a variety of additional factors to determine the performance of a driver and his or her track position throughout the race. This is captured in the more comprehensive system dynamics causal loop diagram in Figure 4, which contains variables that will be discussed further in the thesis.

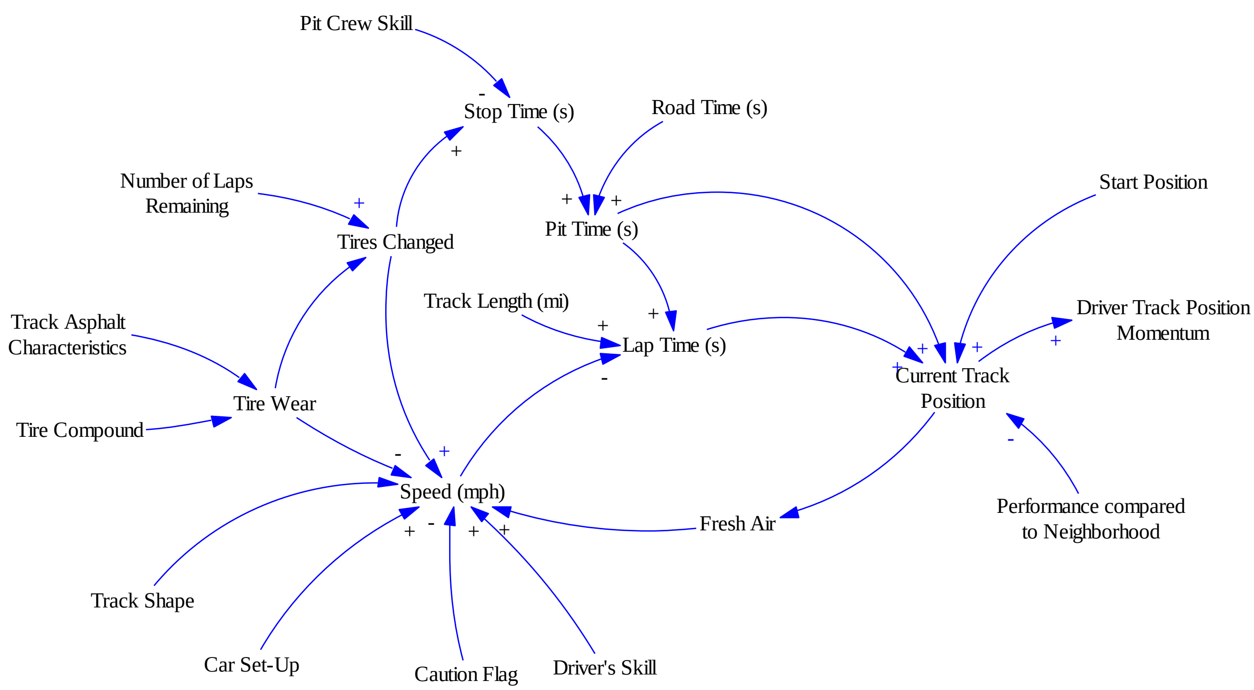


Figure 4. System Dynamics Causal Loop Diagram of Race Characteristics Affecting Track Position

2.3 Race Definitions and Measuring Race Performance

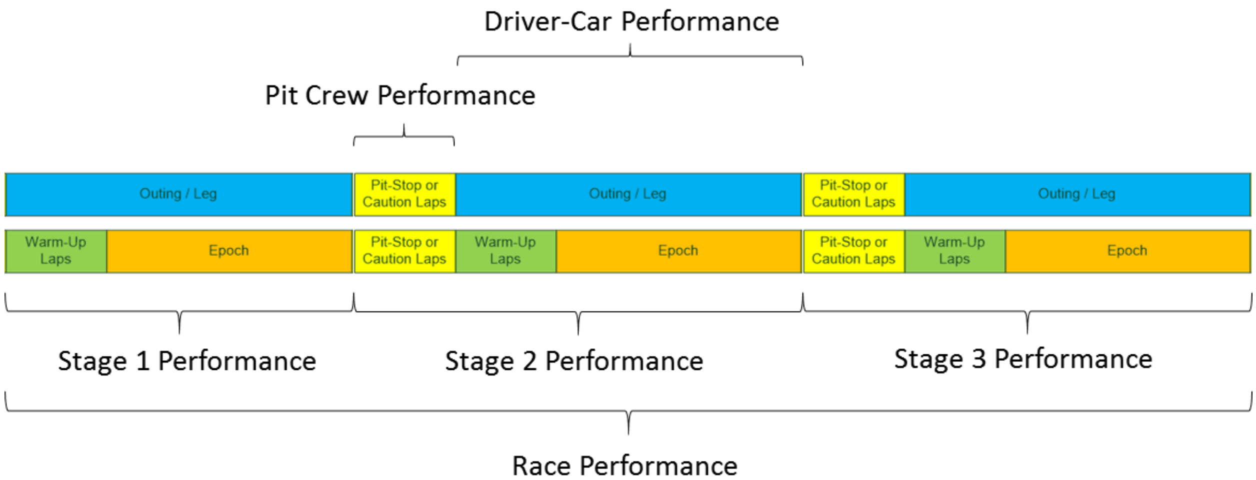


Figure 5. Composition of Race Performance based on Leg, Driver-Car, and Pit Crew based on a hypothetical Race comprising 3 Stages

In the machine learning software, a race comprises a series of outings and caution periods as depicted in the blue and yellow boxes in Figure 5. Each outing comprises warm-up laps, which are marked in green if a pit stop is made, and epochs which are marked in orange. In aggregate, a race is composed of several outing-caution sequences.

In terms of measuring performance, we look at the race performance of a driver as an aggregate of stage performances over the course of the race. Apart from the first stage, each of these stages consists of cautions, warm-up laps, pit stops if any, and epochs.

Cars typically pit during caution laps, and pit crew performance is measured during that period by studying stop times, which are defined by the amount of time each car is stationary in the pits. Crew chiefs also make strategic decisions during this period by deciding whether to swap 0, 2, or 4 tires.

Performance during an epoch is largely influenced by the skill of the driver and the set-up of the car, as touched on in the system dynamics diagram in Figure 3. The driver-car performance should ideally be measured across the warm-up laps and epoch. However, the machine learning software measures driver-car performance based only on what happens during the epoch, with inputs from the caution laps and pits selectively included. Warm-up laps are disregarded in the machine learning model as cars do not run at their optimal speeds until their tires have warmed up.

2.4 Developing Predictions for the Race

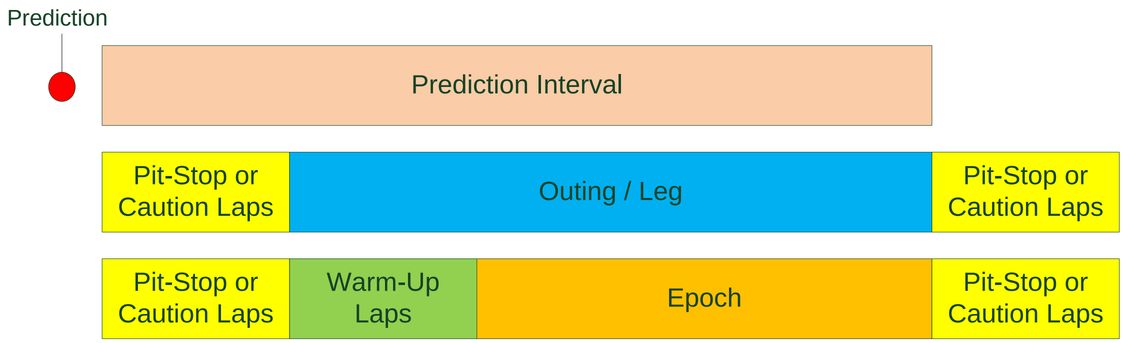


Figure 6. Definition of Prediction Interval in relation to Outings/Legs, Warm-Up Laps, Cautions, Pit Stops, and Epochs

The machine learning software makes a prediction about the track position of a driver over the span of a Prediction Interval as described in Figure 6. This prediction is made prior to the period when a pit stop or caution laps occur, so as to inform crew chiefs of the possible outcomes of different tire change decisions when a pit stop is made.

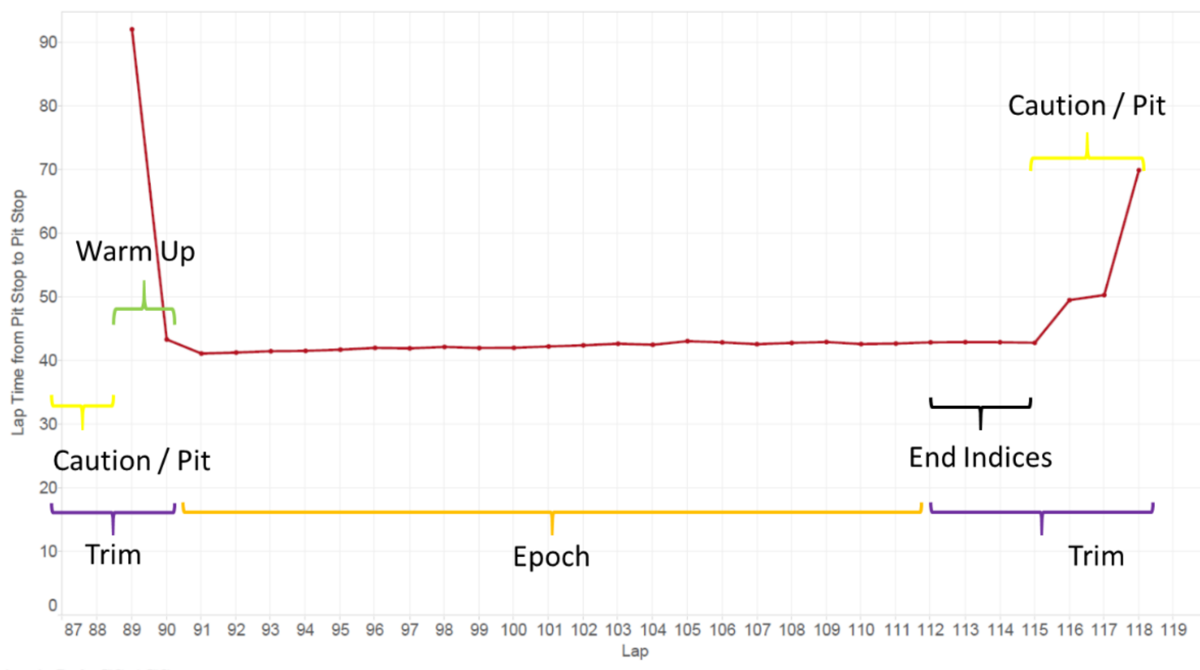


Figure 7. How Epochs are generated for the Feature Matrix in the Machine Learning Software

The software makes this prediction by building a feature matrix using epochs from races in the past, as well as epochs from the ongoing race. These epochs only consider a subset of the characteristics of an outing, as laps are trimmed from the start and end as shown in Figure 7, which plots a driver's lap times over a caution-outing-caution. Epochs are written into the feature matrix only after laps from the front and back of the epoch have been trimmed.

At the front of the epoch, towards the left side of the graph, the caution laps, pits, and warm-up laps are trimmed. The number of warm-up laps trimmed is based on a predetermined number of laps at the beginning of the epoch. Lap times which are slower than the fastest lap recorded within that predetermined number of laps are removed.

At the back of the epoch, towards the right side of the graph, the caution laps and pit stops are removed. In addition to that, a predetermined number of laps known as the end indices are trimmed. This is to account for slower lap times recorded as the caution flags are waved, as well as racing incidents that immediately precede those caution laps.

2.5 Data Structures in the Machine Learning Software

Figure 8 below shows a snapshot of selected data structures that form part of the machine learning software. *Race* is the superset which comprises *parameters*, *pittimes*, *epoch*, *predictions*, and other associated details.

Track-specific characteristics are captured in the *parameters*.

Epoch consists of the feature matrix, which is represented by 165 features in its columns, and all of the race's epochs for all drivers in its rows.

Predictions are the set of predictions made during the course of the race.

Pittimes capture details of all drivers and their respective pit stop and tire-related information.

Lastly, *other* contains supplementary data used for intermediate calculations as the machine learning software is run.

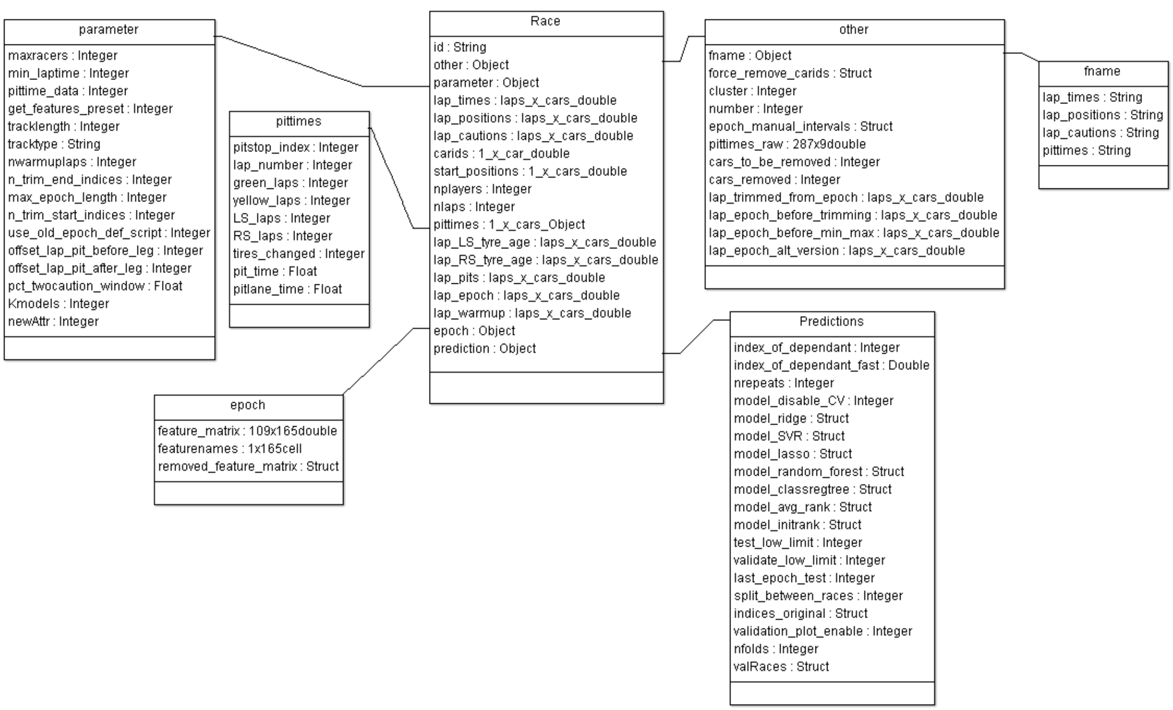


Figure 8. Data structures used by the Machine Learning Software

2.6 Components of the Feature Matrix and Feature Selection

The machine learning software uses a variety of inputs from race telemetry sent to the team garages as the race progresses. Specifically, the following details are captured for all drivers before they are processed to construct the feature matrix:

1. Lap Times
2. Pit Stops
3. Track Positions
4. Race Cautions

The feature matrix in Figure 9 consists of 165 features, but only a subset of 40 are selected as independent variables for the machine learning algorithm, with the objective to predict the track position of drivers at the end of the Prediction Interval. This prediction is represented as the dependent variable, or Feature ID 33.

A number of the independent variables can be classified according to the type of race characteristics that the machine learning software attempts to model. For example, Feature IDs 34 and 35 refer to the track position of the driver in relation to the rest of the field. This is meant to capture the “Fresh Air” effect (Tulabandhula & Rudin, 2014), in which drivers near the front and back of the fields tend to reach faster speeds.

Feature IDs 56, 59, 60, 61, and 69 attempt to capture characteristics related to drivers in the neighborhood. For instance, the number of tires changed by rivals in the vicinity, the respective age of their tires, and their relative track positions in the field of drivers all have an impact on the predictions for a driver.

Other race characteristics are also accounted for in the remaining features, but it is up to the machine learning algorithm to find a linear combination of features such as using regression, that minimizes the error in the predictions. The following are the dependent and independent variables in the regression equation. Detailed descriptions have been removed for this public document.

ID.	Classification	Description	Dep. Var	Ind. Var
33	current outing	delrank (leg end - pre pit)'	X	
9	before pit	Feature ID 9		X

13	before pit	Feature ID 13		X
26	before pit	Feature ID 26		X
28	before pit	Feature ID 28		X
34	before pit	Feature ID 34		X
35	before pit	Feature ID 35		X
37	before pit	Feature ID 37		X
40	before pit	Feature ID 40		X
41	before pit	Feature ID 41		X
43	before pit	Feature ID 43		X
44	before pit	Feature ID 44		X
45	before pit	Feature ID 45		X
56	before pit	Feature ID 56		X
59	before pit	Feature ID 59		X
60	before pit	Feature ID 60		X
61	before pit	Feature ID 61		X
69	before pit	Feature ID 69		X
75	before pit	Feature ID 75		X
76	after pit	Feature ID 76		X
77	after pit	Feature ID 77		X
82	before pit	Feature ID 82		X
85	after pit	Feature ID 85		X
89	before pit	Feature ID 89		X
127	before pit	Feature ID 127		X
132	current outing	Feature ID 132		X
136	current outing	Feature ID 136		X
139	after pit	Feature ID 139		X
140	after pit	Feature ID 140		X
146	before pit	Feature ID 146		X
149	before pit	Feature ID 149		X
150	before pit	Feature ID 150		X
151	before pit	Feature ID 151		X
153	before pit	Feature ID 153		X
157	before pit	Feature ID 157		X
158	before pit	Feature ID 158		X
160	before pit	Feature ID 160		X
161	before pit	Feature ID 161		X
163	before pit	Feature ID 163		X
164	before pit	Feature ID 164		X
165	before pit	Feature ID 165		X

Figure 9. Dependent and Independent Variables in Machine Learning Software

2.7 Predictive Capability of Current Model

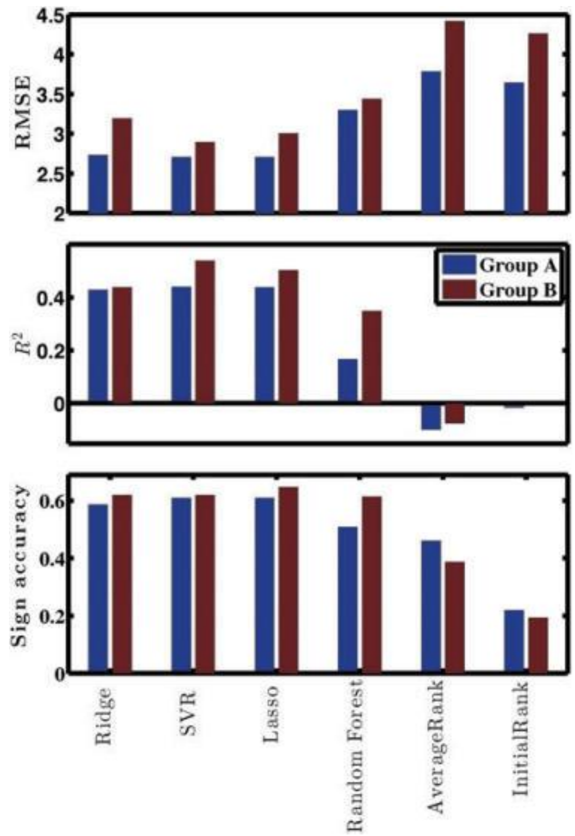


Figure 10. Predictive Performance of Machine Learning Software based on Root Mean Square Error, R^2, and Sign Accuracy (Tulabandhula & Rudin, 2014)

The machine learning software's implementation of ridge regression, support vector regression, lasso, and random forests performed significantly better than the baseline methods across 3 metrics – the root mean square error, R^2, and sign accuracy (Tulabandhula & Rudin, 2014).

For the Initial Rank baseline method, the predicted change in track position over the course of the epoch was 0. For the Average Rank baseline method, the track position at the end of the epoch was based on driver's average track position from the previous epochs.

Essentially, there was some evidence that the model possessed some predictive capability in estimating the track position of drivers at the end of the epoch compared to a naïve approach.

2.8 Observations and Suggested Improvements

To begin the journey of answering the questions posed in this thesis, we needed to predict the difference in track position at the end of the upcoming epoch, before the next pit stop occurred. By doing so, the software could simultaneously recommend a tire change decision for the current pit stop leading into the epoch. We began by looking at gaps we needed to fill for the current version of the software.

Firstly, the machine learning model required accurate inputs for the tire change decisions made by the entire field of drivers. It had not been designed to make predictions for all drivers, and additional features were needed to provide some degree of predictability for the actual tire decisions expected of individual drivers. In this regard, we understood that NASCAR teams typically had a good understanding of the preferred tire change strategy for each track and there were some observable patterns that we could include into the model.

Secondly, teams understood that certain tracks such as Southeast_I necessitated 4-tire changes very frequently because of the very high tire wear of the track. As a result, they rarely made 2 or 0-tire changes during a pit stop. The lack of such samples where contrarian decisions were taken made it difficult for the machine learning model to make precise predictions for such strategies.

Thirdly, although there was a sizeable number of features incorporated into the machine learning model, there was significant scope to introduce additional features that could improve the predictive capability of the software. For example, there should be strong penalties applied for pitting under regular racing conditions, particularly if there is no caution likely to occur anytime soon.

In order to address these gaps, we explored a few areas in the thesis. Firstly, we sought to understand the broader tire change strategies employed on different race tracks. Secondly, we learned more about specific circumstances in which 0 or 2-tire pit stops were made. Thirdly, we attempted to draw correlations between different features in the model to strengthen the software's predictive capabilities.

3. Analysis of Machine Learning Features

3.1 Performance After 2-Tire Pit Stop

A series of 17 races from the 2012 NASCAR Cup were classified into 2 groups for the machine learning software. The first group of races represented those where pit stops to change 2 tires led to a loss of track position, whereas those in the second group of races saw drivers maintain or improve their track positions (Tulabandhula & Rudin, 2014). The classification of races is shown in the following table.

Group A (Loss of Track Position with 2-Tire Stop)	**Group B** (Maintain Track Position with 2-Tire Stop)
1. West_A	1. Southeast_C, Race 1
2. Southeast_H	2. Midwest_C, Race 1
3. Southeast_D, Race 1	3. Northeast_B, Race 1
4. Midwest_A, Race 1	4. Midwest_D
5. Midwest_B	5. Northeast_C, Race 1
6. Southwest_A, Race 2	6. Midwest_A, Race 2
	7. Southeast_C, Race 2
	8. Northeast_C, Race 2
	9. Midwest_C, Race 2
	10. Southeast_E, Race 2
	11. Southeast_G

Figure 11. Classification of Races based on Machine Learning Software

We selected Midwest_B from Group A and Southeast_G from Group B to illustrate the change in track position between two pit stops whenever 2-tires were changed. For the two graphs on the following page, the horizontal axis represents the lap of the race. The vertical axis represents the track position of the driver 4 laps after pitting.

Lines are drawn from the 2-tire change pit stops, either to the subsequent pit stop, or to the end of the race. These lines are then colored depending on the number of tires changed. Red represents 2 tires, blue represents 0 tires, and green represents 4 tires. If a line starts with red and ends in blue, it means that the driver pitted for 2 tires for the first stop and 0 tires in the next. The end of each line is also labeled with an ID number (i.e. “3”) and the car number associated with that line (i.e. “#22”).

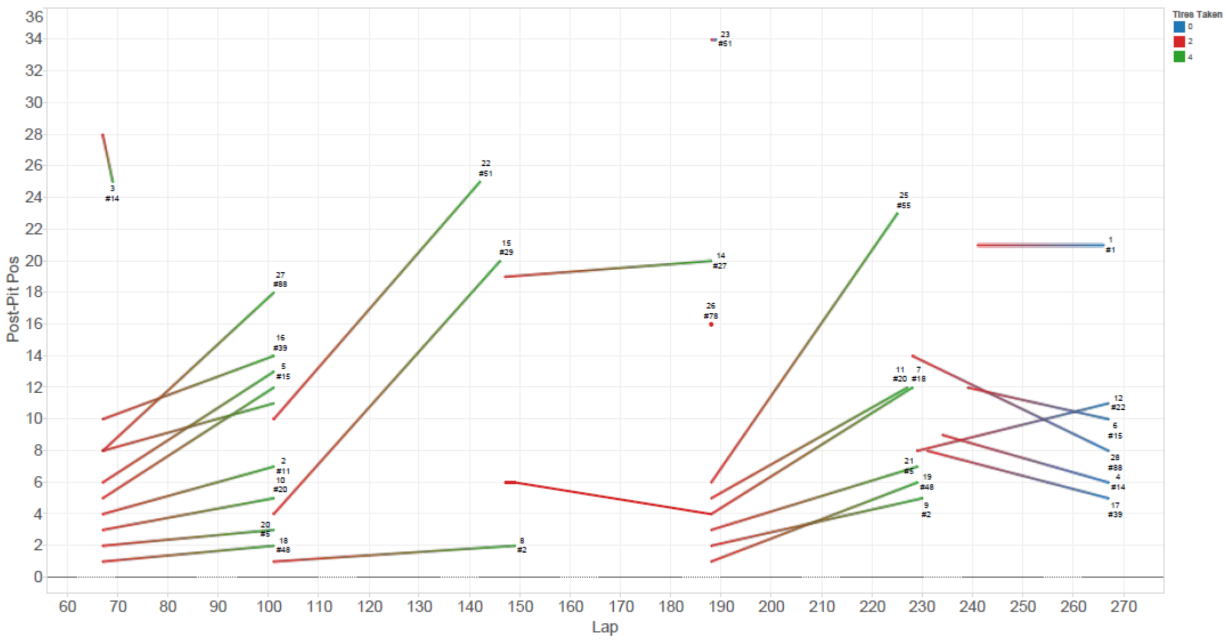


Figure 12. Graph of Midwest_B 2012 Race depicting Change in Track Position for Pit Stop with 2 Tires Changed

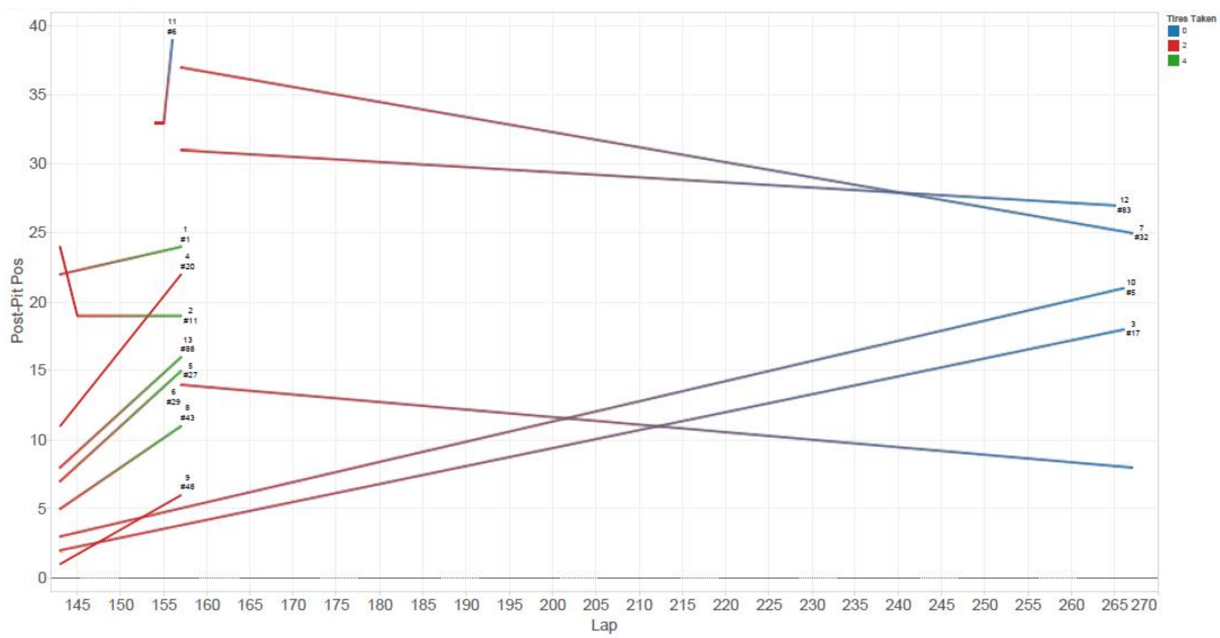


Figure 13. Graph of Southeast_G 2012 Race depicting Change in Track Position for Pit Stop with 2 Tires Changed

By visually comparing the two races from different groups, it was fairly obvious that drivers in the Midwest_B race lost track position after pitting for 2 tires. However, this wasn’t apparent in the

Southeast_G race. There was a fair number of upward-trending lines, suggesting that it may not be advisable to change two tires in the Southeast_G race either.

Importantly, what stood out were the number of pit stops where 2 tires were changed. There was a greater frequency of such pit stops in the Midwest_B race compared to the Southeast_G race, suggesting that teams had more flexibility to change their tire strategy in the Midwest_B race. We attempted to produce visualizations of the Midwest_B races from 2013 and 2014, to understand whether 2-tire pit stops were similarly frequent across a 3-year period.

The following graphs in Figure 14 and Figure 15 depict the Midwest_B 2013 race and Midwest_B 2014 race respectively.

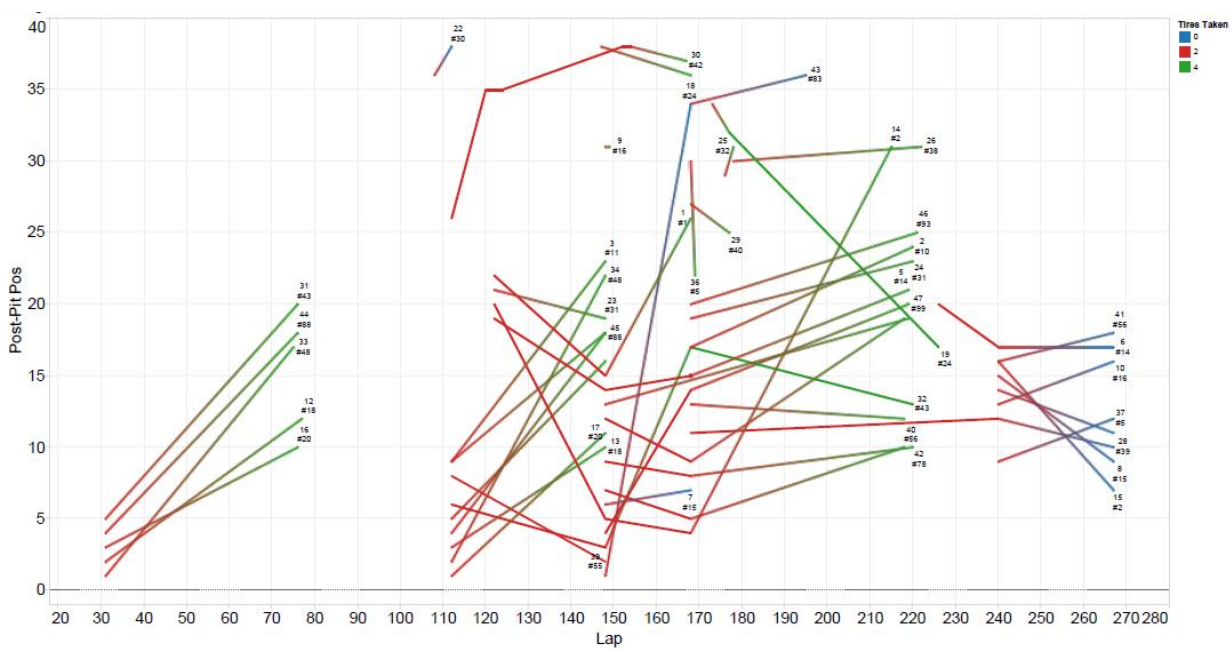


Figure 14. Graph of Midwest_B 2013 Race depicting Change in Track Position for Pit Stop with 2 Tires Changed

Comparing the 3 Midwest_B races made it clear that drivers were often able to change 2 tires at different stages of the race. However, the frequency of such pit stops differed substantially across all 3 races. The 2014 race featured the fewest 2-tire pit stops while the 2013 race featured the most. This implied that the absolute number of 2-tire pit stops was not a useful indicator of the tire strategy. We needed to look a step further, and perhaps compare the ratio of 2 to 4 tire pit stops to get a better understanding of the track's characteristics.

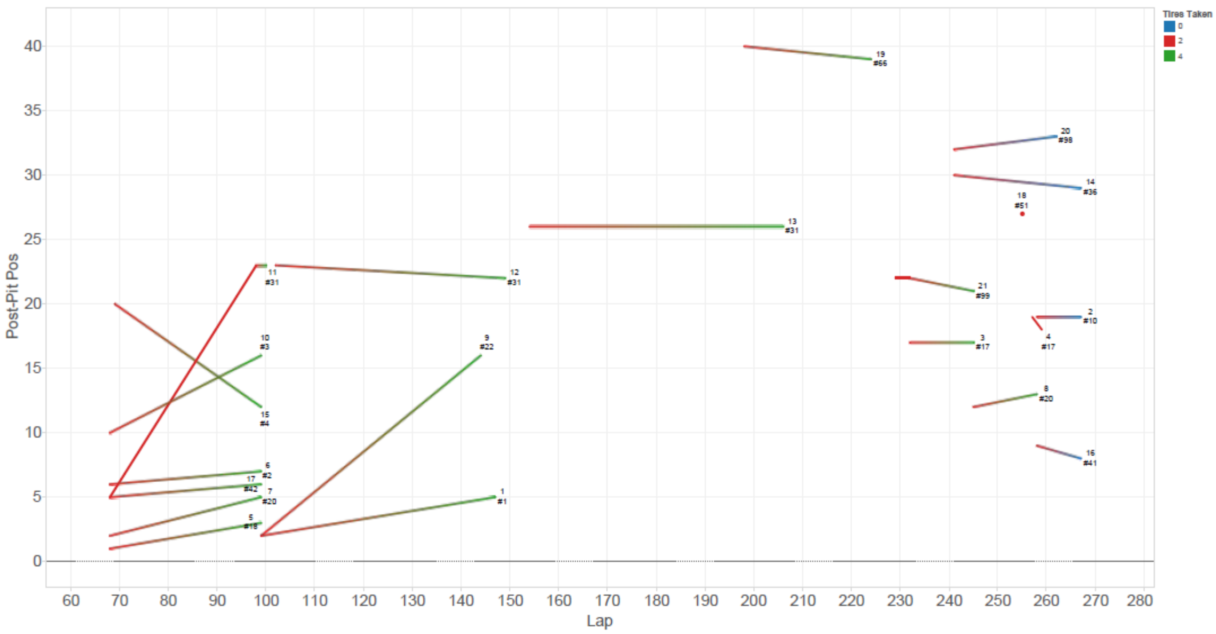


Figure 15. Graph of Midwest_B 2014 Race depicting Change in Track Position for Pit Stop with 2 Tires Changed

In practical terms, tracks with higher tire wear due to its asphalt characteristics mean that drivers have to pit more frequently for 4-tire changes instead of 2. During a review with the industry sponsor, we were informed that some tracks in the NASCAR season had a very strong bias in favour of 4-tire changes. This appeared to run counter to how some tracks were classified in Figure 11. In particular, the two races held in Midwest_A were categorized into different groups. If tracks had a strong bias to one or another tire change strategy, why would Midwest_A appear under both groups? Were the groupings not sufficiently detailed, or could there be more categories not accounted for?

This led us to plot the graph of the Southeast_I race in 2014, which was outside the group of classified races. We were advised that Southeast_I was a track where almost all pit stops involved 4-tire changes because of the extremely high tire wear. We needed to verify whether drivers who pitted for 2 tires truly performed poorly almost all the time.

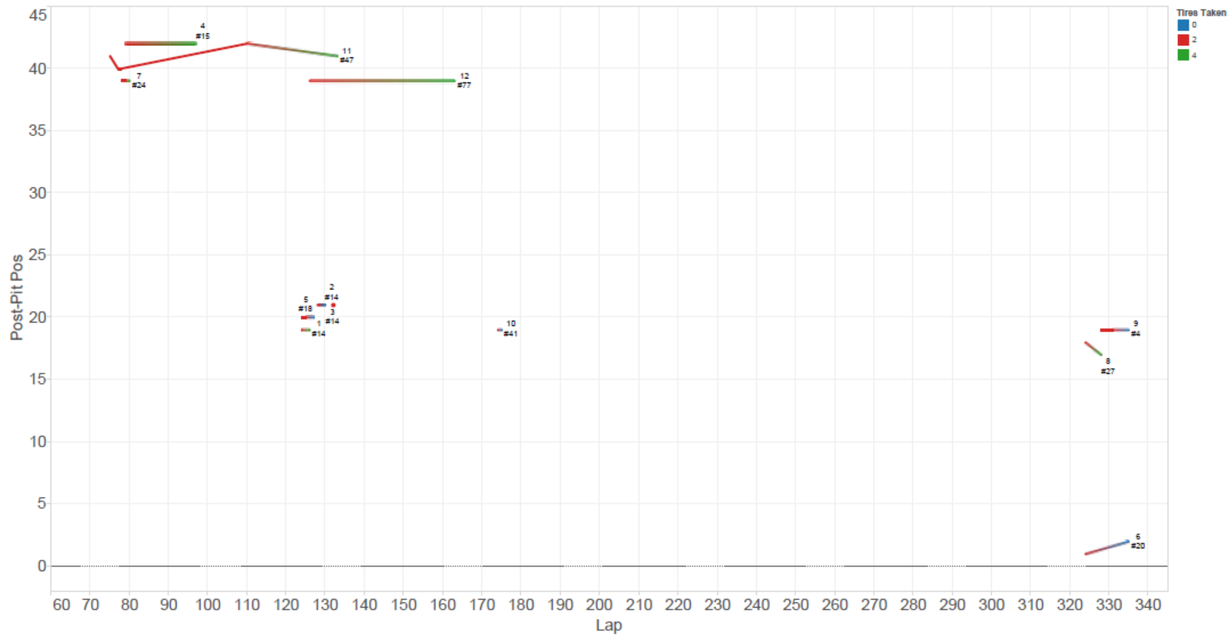


Figure 16. Graph of Southeast_I 2014 Race depicting Change in Track Position for Pit Stop with 2 Tires Changed

Based on the plot for the Southeast_I race, it was evident that 2-tire pit stops were indeed very infrequent. However, not all 2-tire pit stop decisions were bad, and if we were to use the original classification as suggested in the paper, Southeast_I would be considered a Group B race because the 2-tire pit stops produced reasonable results, contrary to what was the intuitive decision.

Our review of the plots based on the machine learning software's track classification did not yield any definitive conclusion, except to suggest that the race classifications could be further refined. However, we noted the following:

1. Races where 2-tire pit stops usually led to losses in track position did not necessarily mean that all drivers were forced to pit for 4-tires at all times. There were circumstances where taking a contrarian approach worked.

2. It may be possible to infer the characteristics of a track and its impact on tire change decisions by looking at the ratio of 4-tire to 2-tire pit stops.

3.2 Performance based on Stop Time in Pit

We wanted to understand the importance of pit crew performance during pit stops, in relation to the drivers' finishing position at the end of the race. Although the machine learning software currently shaves off caution periods and warm up laps because of the high variability of track positions and lap times during the caution periods, it retains the stop time of the cars during their pit stops. These stop times, which do not include the time that drivers spend driving down the pit lane, indicate the time that vehicles are stationary in the pit. They solely measure the time taken for pit crew to refuel cars and change tires.

It takes approximately 4 seconds to change 0 tires and only refuel a car, 6 to 7 seconds to change 2 tires and refuel, and about 13 to 14 seconds to change 4 tires and refuel. The frequency of 4-tire pit stops tends to be significantly higher than other types of pit stops, so we focused purely on pit crew performance during those types of stops. We selected the following races from the 2014 NASCAR Cup for review:

1. West_A
2. Southeast_C, Race 1
3. Northeast_B, Race 1
4. Midwest_A, Race 1
5. Midwest_C, Race 2
6. Southwest_A, Race 2

The following box plots were generated by reviewing the finishing position of the drivers on the horizontal axis, and the stop time taken in seconds on the vertical axis. The whiskers of the box plot extend to 1.5 times the interquartile range of the stop times for 4-tire pit stops.

We first looked at the West_A race as depicted in Figure 17. We noticed from the plot that the time taken for 4-tire pit stops were in fact higher for the top 5 drivers compared to those who finished from 6th to 10th. In fact, pit stop times for the 11th to 14th drivers were generally lower than the top 10 drivers. This suggested that the performance of pit crews during pit stops was not the sole determinant in determining performance of drivers. In fact, there was even a pit stop lasting more than 60 seconds for the 5th placed driver, which was about 4 times as long as a regular 4-tire pit stop.

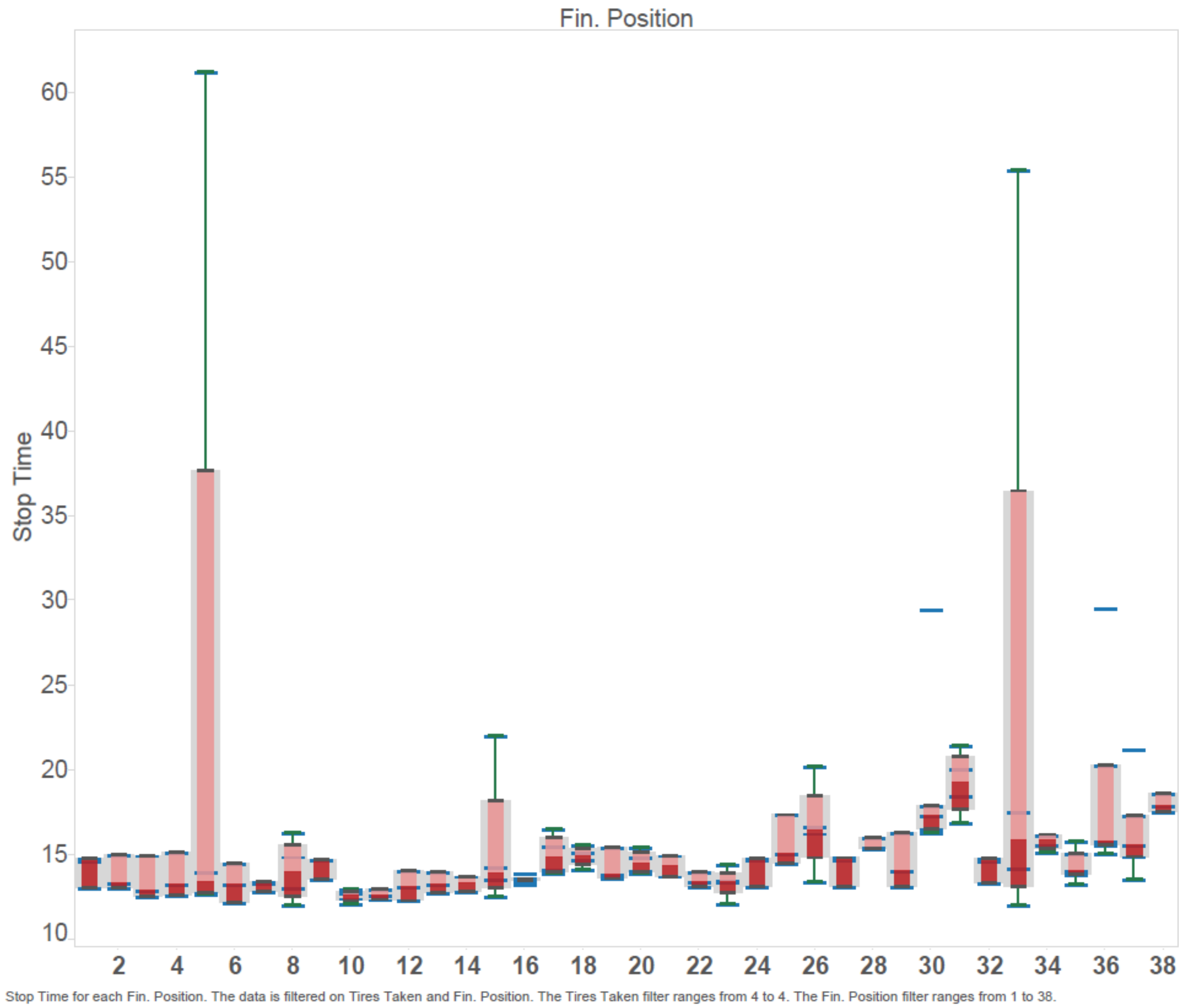


Figure 17. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for West_A 2014 Race

However, we noticed an upward slope in the box plots after the 23rd-placed driver in the race. Perhaps there was a weak correlation between the finishing position of the drivers and pit crew performance. It could be possible that past a certain point, pit crew performance was not a useful determinant of a driver's likely finishing position.

In order to determine whether the West_A race was an outlier or in fact a trend that could be seen across many races, we looked at the plots for the remaining races. The next plot in Figure 18 shows 4-tire changes in Southeast_C Race 1 to see whether the pattern seen in the West_A race repeats itself.

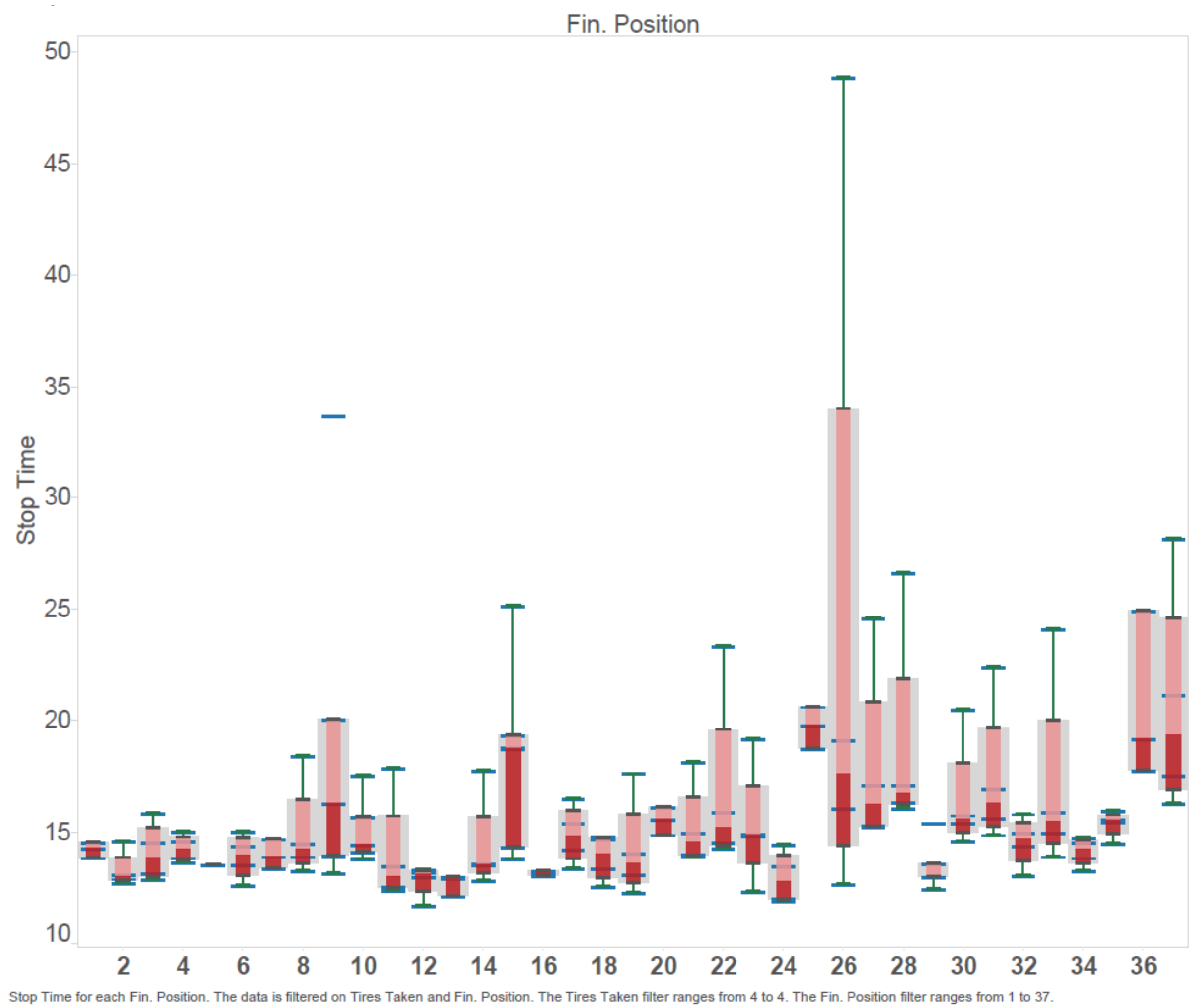


Figure 18. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for Southeast_C Race 1 2014

Based on Southeast_C Race 1 in Figure 18, there is a clearer upward-sloping trend when comparing the box plots from the driver that finished first to the one who finished last. However, like the previous West_A race, the top few drivers were not necessarily the ones who clocked the fastest 4-tire pit stops. Although the top 10 drivers clocked roughly the same stop times, the 12th, 13th, 16th, and 24th-placed drivers had among the best pit stop performances.

The Northeast_B Race 1 of 2014 as depicted in Figure 19 was a bit of an outlier in our analysis. The 6th and 12th-placed drivers had particularly poor 4-tire pit stops that lasted between 80 to 120 seconds, as indicated by their very tall boxes. Other drivers at the front of the field had relatively low 4-tire pit stop times compared to the lower-half, from the 20th position onwards.

There are exceptions where very poor pit stop times will not affect track positions too badly. These occur when caution periods last particularly long, or when subsequent caution periods are very closely separated such that drivers can regain lost laps.

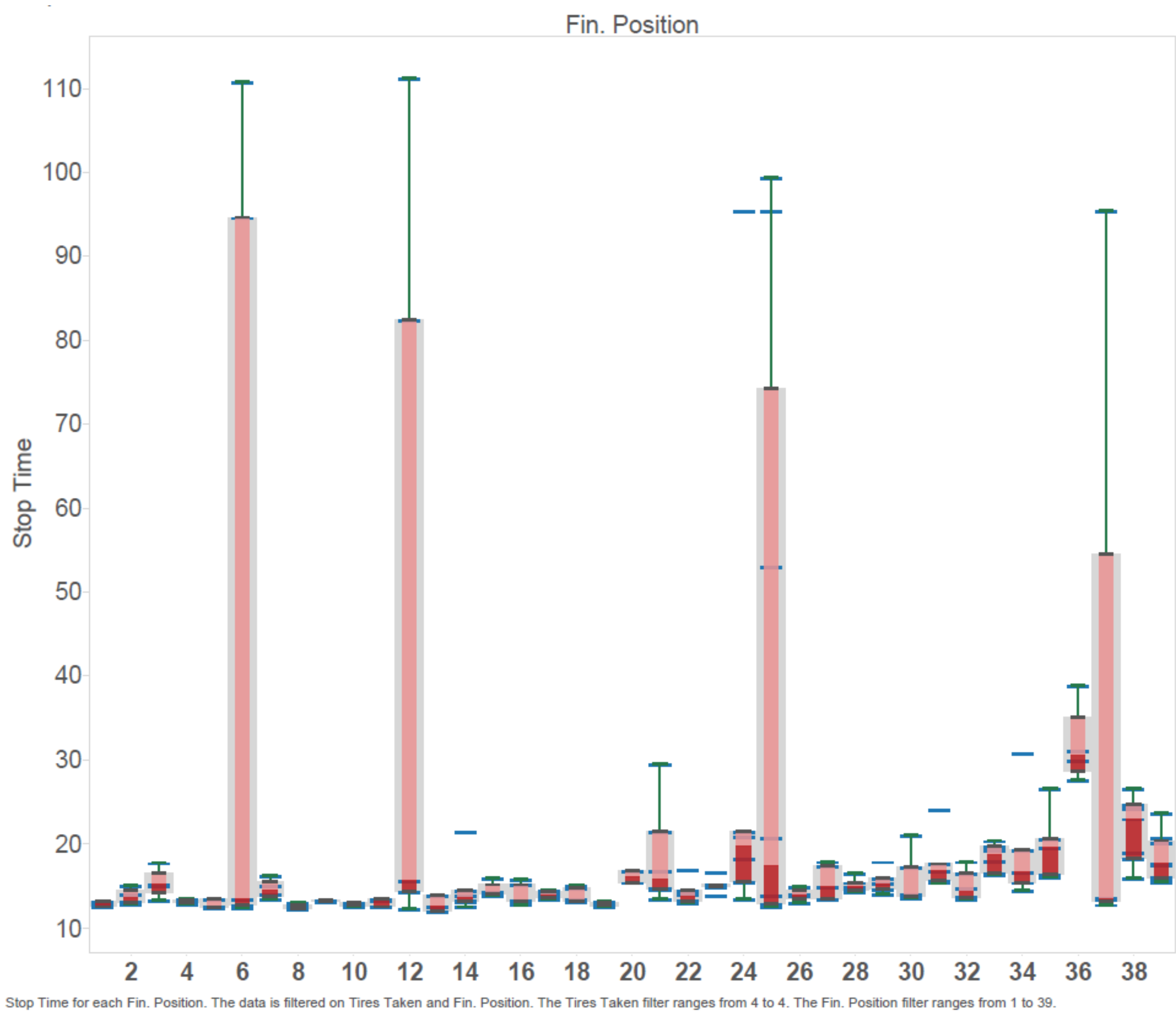


Figure 19. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for Northeast_B Race 1 2014

In this race, a horrendous 111-second pit stop dropped the 6th-position finisher by about 20 spots after lap 72, and further down the field by a few more spots in the subsequent 95-second pit stop on lap 80. The driver staged a miraculous recovery over the next leg of the race, moving from 30th to 14th-position by the next pit stop, and advancing to 6th by the next. This outlier suggested that pit stops had to be viewed in the context of overall performance by also including details about what happened during the regular racing period. In a similar situation, the car in 12th had to make up for the 82-second pit stop by only refueling and not changing tires at the subsequent stop 9 laps later. This helped regain some track position lost.

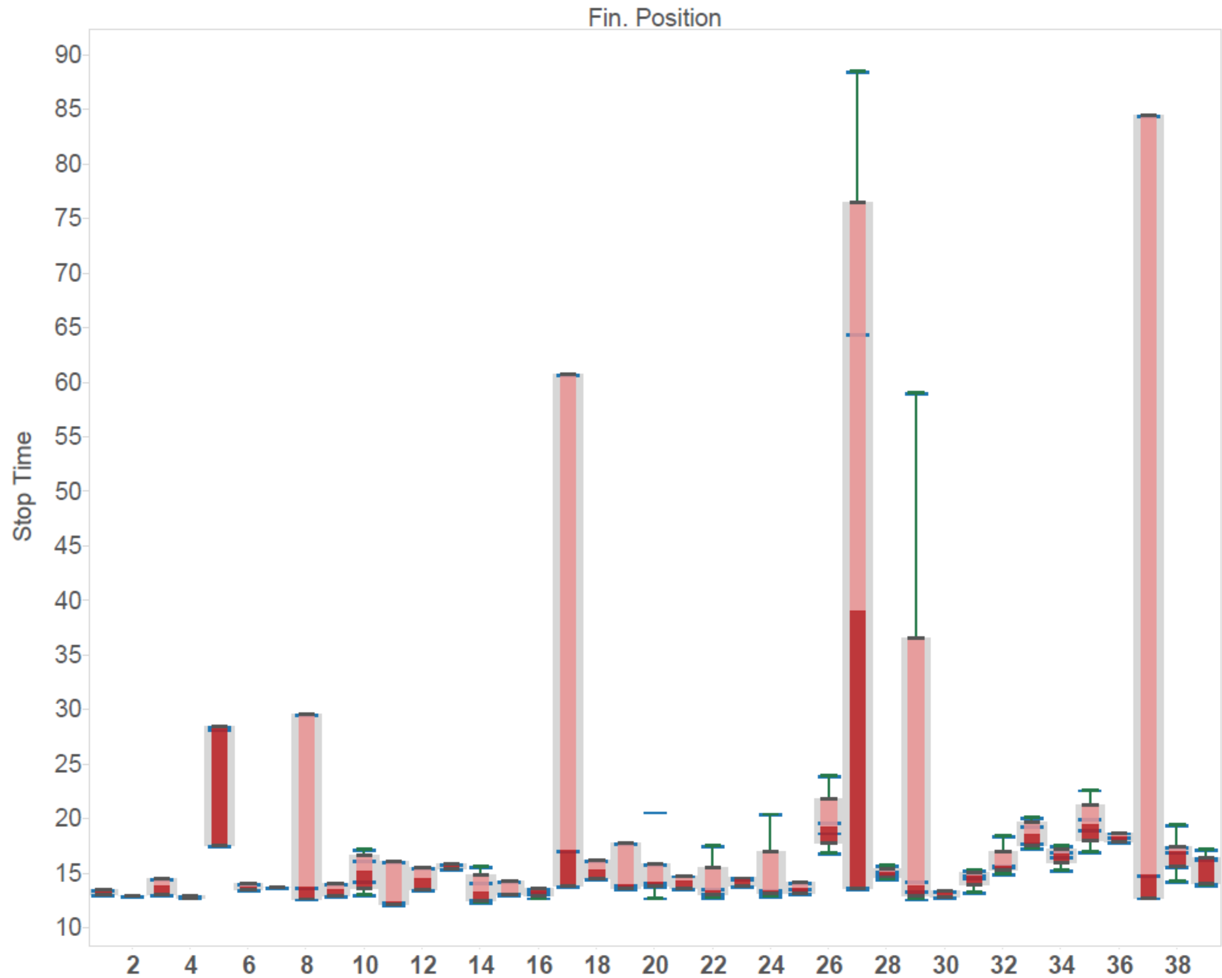


Figure 20. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for Midwest_A Race 1 2014

The Midwest_A Race 1 in 2014 had some very long 4-tire pit stops recorded that were outside the norm. Although the top 15 drivers had very consistent 4-tire pit stop times in general, drivers in 5th and 8th positions clocked a few longer pit stops in comparison. The lower half of the field, like in previous races we observed, had relatively longer stop times and poorer consistency.

In the case of the driver in 5th, the poor pit stops were recorded in the earlier legs of the race, particularly during stops on Laps 6, 8, 73, and 104 of the 200-lap race. Running at 19th position on Lap 108, the driver eventually overcame those deficits and slowly advanced in the field. From this driver's performance, it is reasonable to assume that poor performance in the early part of the race can be overcome in the later stages. The driver's remaining pit stops were only for 2-tire changes on laps 110, 148, and 184, and stop times during those pit stops were reasonably good.

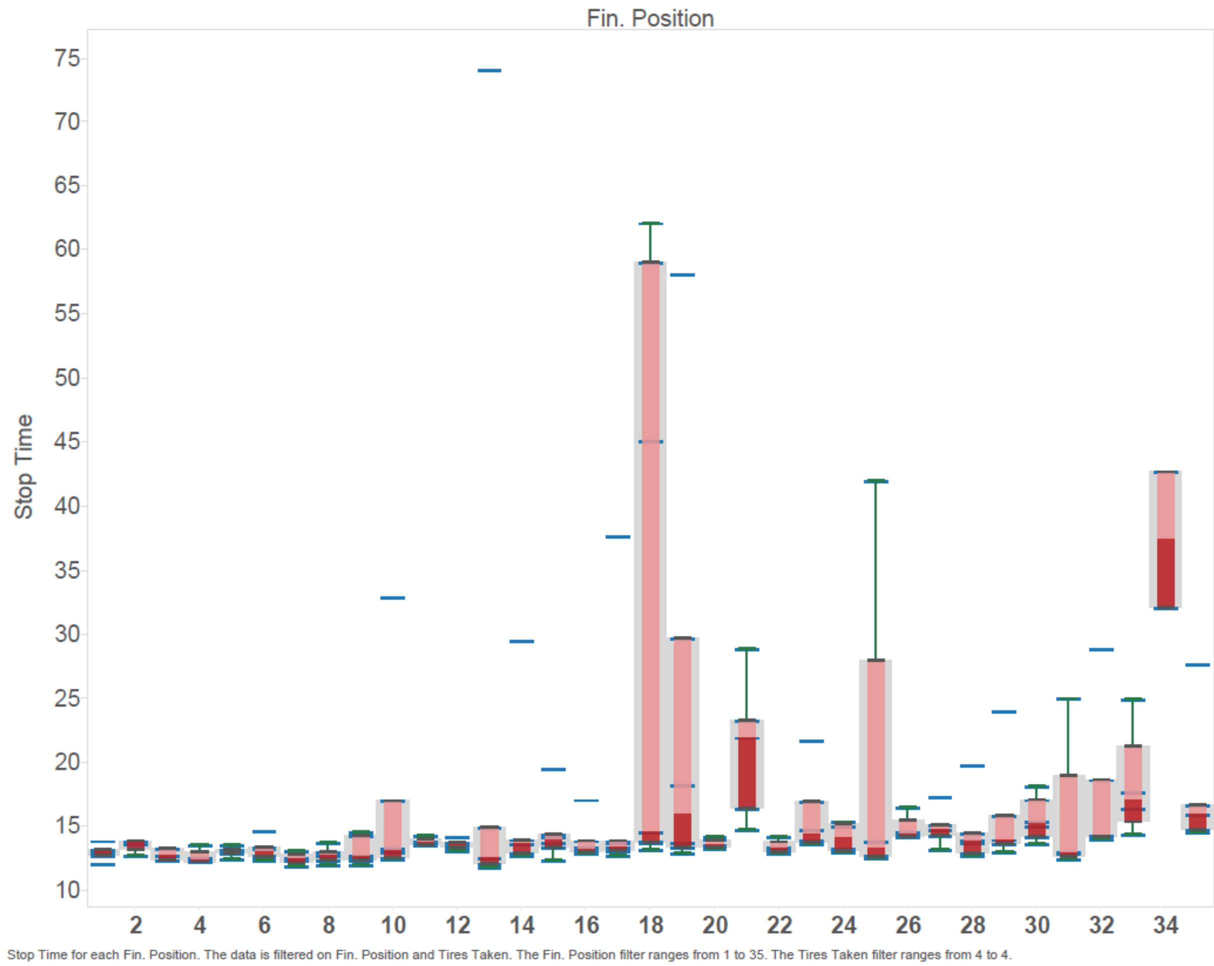


Figure 21. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for Midwest_C Race 2 2014

Next, we studied the Midwest_C Race 2 in 2014 as depicted in Figure 21, to see whether there was any discernable trend in the pit stop times and the finishing position of drivers in that race. Again, we noticed that the top few drivers had stop times that were tightly-clustered together. In addition, they had relatively low and consistent 4-tire pit stop times compared to the drivers that were further down the order. For instance, the driver in 10th had one particularly long 4-tire pit stop, and drivers in the 14th to 19th positions had stops that were a few seconds off the interquartile range.

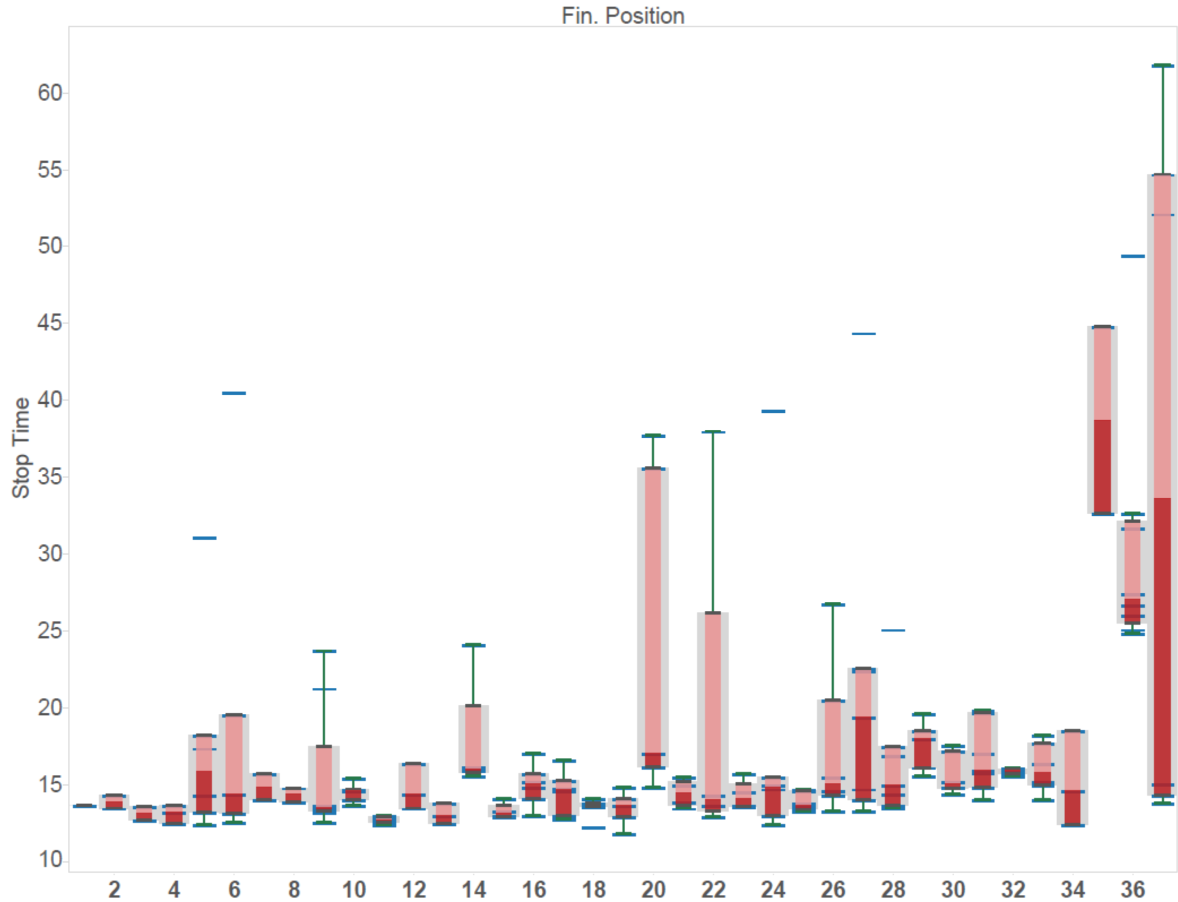


Figure 22. Box Plot of 4-Tire Pit Stop Times based on Final Finishing Position of Drivers for Southwest_A Race 2 2014

Finally, we reviewed the Southwest_A Race 2 in 2014 in Figure 22 and noticed similar trends. Cars in the front generally had more consistent and lower stop times than those at the back, although this was not universal. Similar to what was discovered in previous races, poor stop times could be made up through a combination of good performance outside of the caution periods, as well as good strategic decisions on tire changes in subsequent pit stops.

We had the following conclusions by reviewing plots of the selected races:

1. Drivers in the front tended to have more consistent and lower stop times compared to those the back. However, this was not universally applicable as other factors influenced their overall race performance. A more holistic view of the race was likely to give a more complete picture of the driver's performance.
2. Drivers with very high stop times tended to lose track position. However, such losses could be mitigated by making strategic tire change decisions later on in the race, and clocking good pit stop times thereafter.

3.3 Momentum of Track Positions Gained or Lost during a Race

Over the course of a race, drivers tend to move up and down the order depending on the performance of their cars. We wanted to investigate whether the changes in their track position were consistently positive or negative as the race progressed, particularly during each epoch as shown in Figure 5. An epoch is defined by the leg, less the warm-up laps and caution periods. Warm up laps are trimmed because drivers lap more slowly during these legs as their tires need time to warm up before they can perform optimally.

A driver's performance during each epoch was captured in the machine learning software under the feature that measured the momentum of track positions gained or lost during a race. Specifically, this measurement is referred to in the software as the *average previous rate of change in rank*, or what we call the Driver Momentum. This was calculated for each leg of the race by tracking the average change in track position for each outing and obtaining the median:

$$x_i = \frac{Track\ Position\ at\ the\ End\ of\ Epoch_{i-1} - Track\ Position\ at\ the\ Start\ of\ Epoch_{i-1}}{Number\ of\ Laps\ in\ Epoch_{i-1}}$$

$$Driver\ Momentum = Median\ of\ \{\, x_2, x_3, \ldots, x_k \,\}$$

where *i* = outing number, and *k* = number of outings. x_1 is ignored in the calculation of Driver Momentum as there are no positions prior to the first lap.

In the box plot shown in Figure 23, we studied the Southwest_B Race 2 race in 2014. The vertical axis represents the average rate of change in track position for the previous epoch of the race. The horizontal axis represents the final track position of drivers in that race. Each point in the box plot represents the driver momentum at each point of the race. A negative momentum represents an improvement in track position, while a positive momentum represents a loss in track position.

Looking across the field, we noticed that the momentum for most drivers tended to be either in the positive or negative realm. This meant that most drivers who improved their track position during a leg of the race tended to maintain that momentum for subsequent legs of the race. Those who lost track position during a leg of the race continued to lose track position in the following legs of the race. This observation applied to a majority of the drivers in the race, suggesting that one could predict the relative performance of most drivers by simply studying their momentum in the earlier stages of the race. Since Southwest_B was classified as a track with high tire wear, we needed to check if this was simply an isolated case, or whether such driver momentum patterns occurred elsewhere as well.

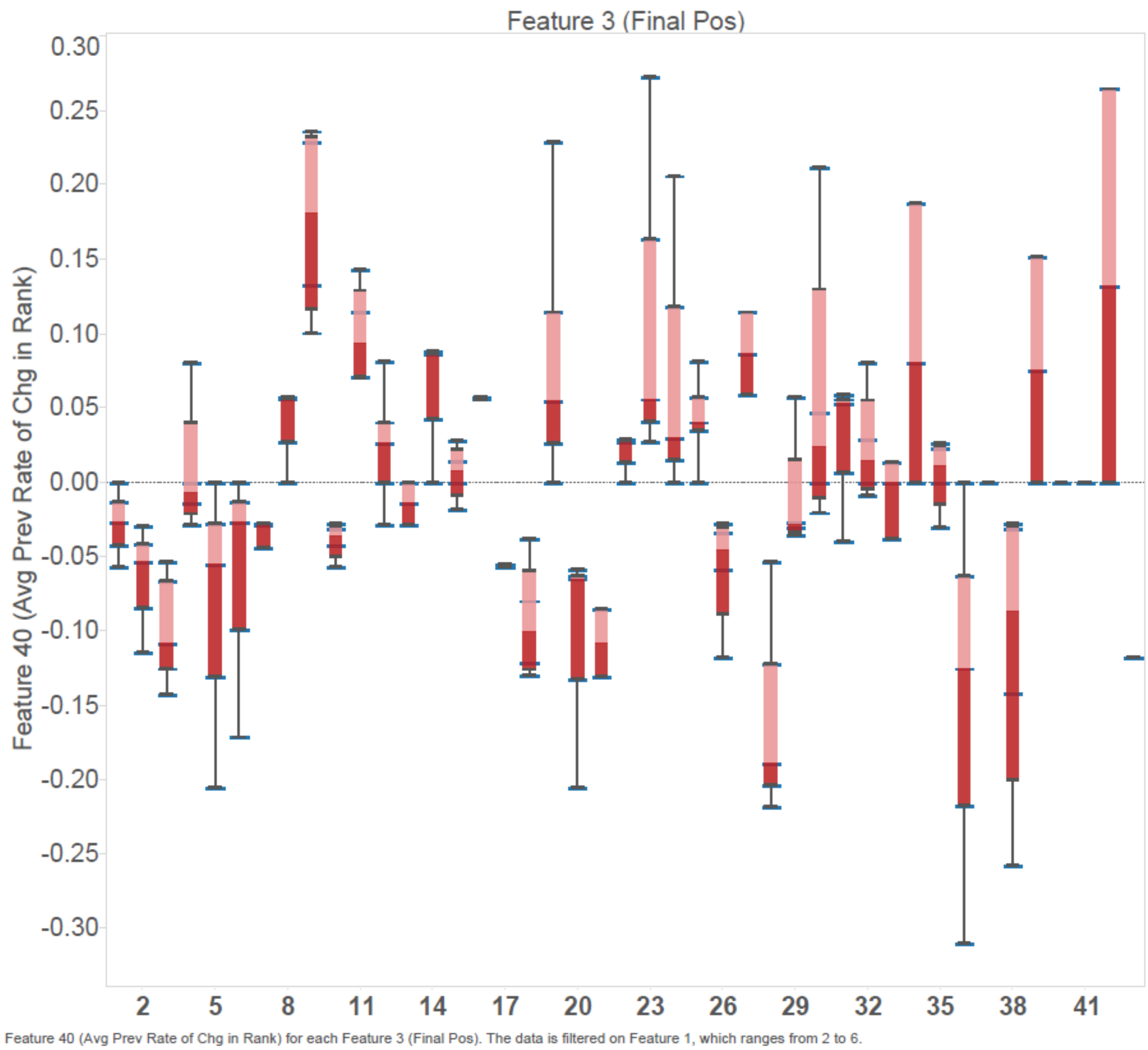


Figure 23. Box Plot of Momentum of Track Position for Southwest_B Race 2 2014 for Drivers based on Finishing Position

We looked at the Southeast_F Race 2 in 2014 to study whether the pattern applied to another track with high tire wear. In Figure 24, we noticed that the pattern remained consistent for most of the drivers. Those who gained track position in one epoch of the race continued to do so in others. Those who lost track position in one epoch of the race continued to lose track position in others.

Next, we attempted to perform the same analysis across a larger number of tracks, covering those which had low and medium tire wear as well to see if the pattern was generalizable across a larger group of races in the NASCAR Cup season.

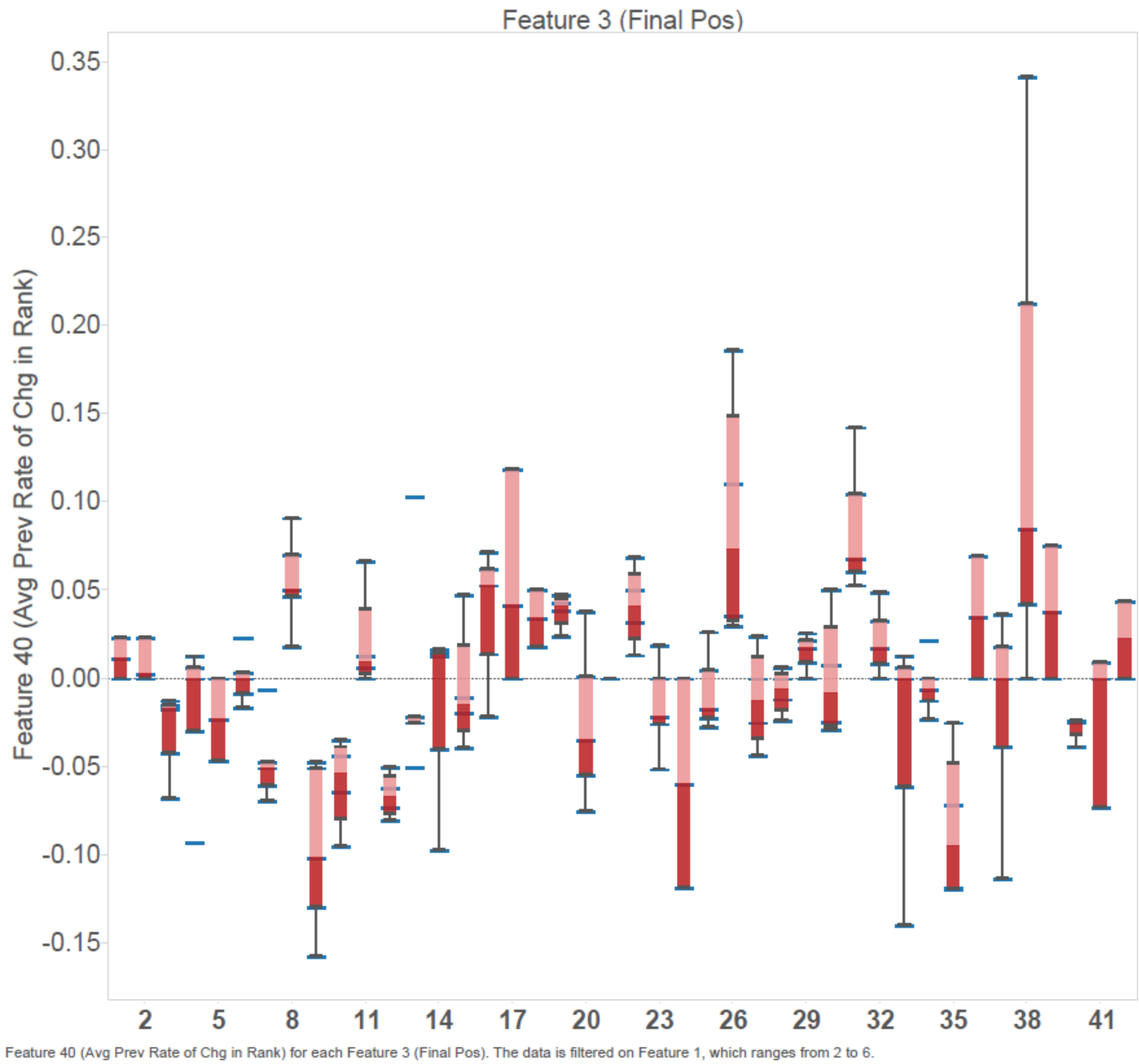


Figure 24. Box Plot of Momentum of Track Position for Southeast_F Race 2 2014 for Drivers based on Finishing Position

The following **low tire wear** tracks were selected for analysis in the subsequent pages:

1. Northeast_C, Race 1
2. Northeast_C, Race 2
3. Midwest_A, Race 1

4. Southwest_A, Race 1

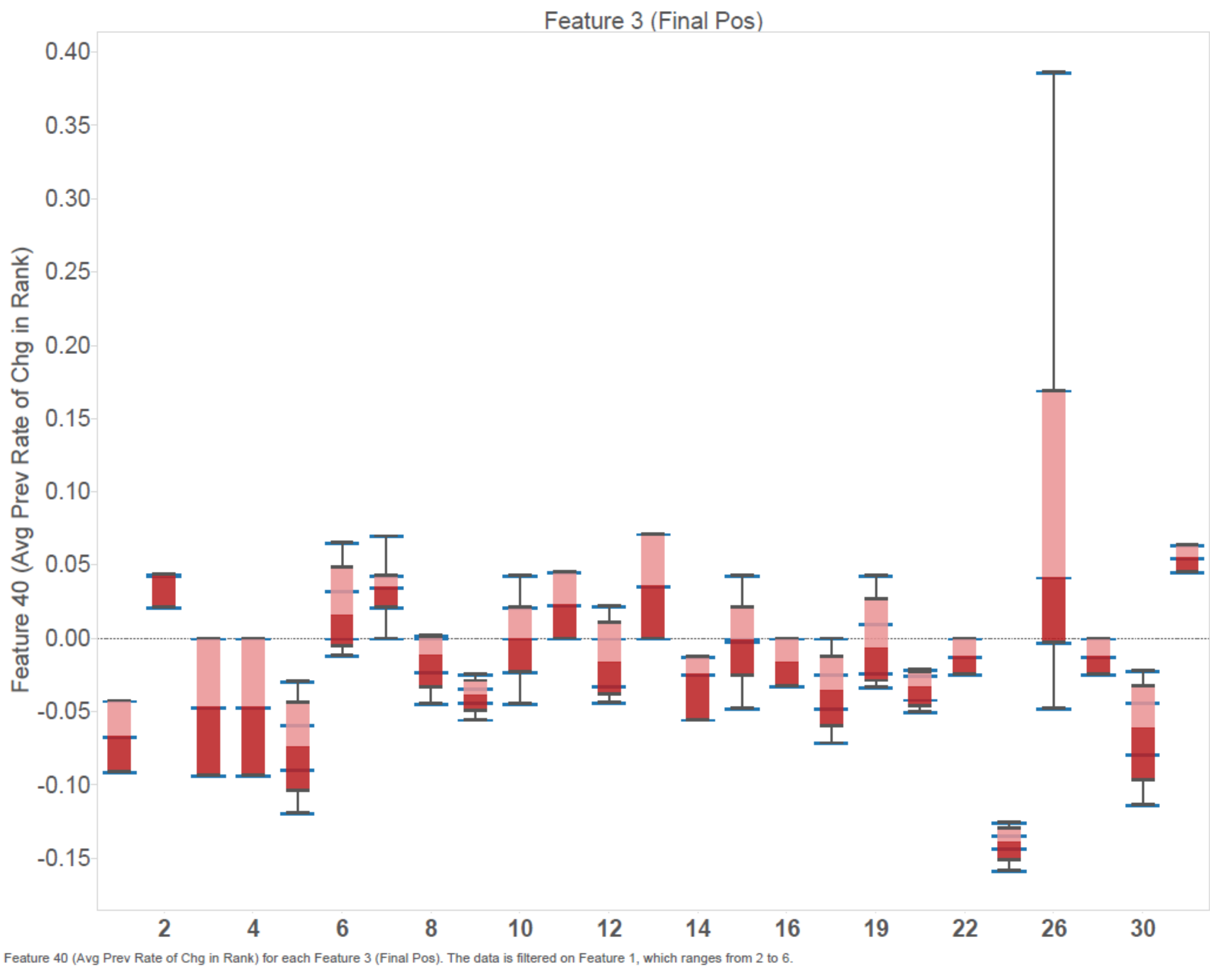


Figure 25. Box Plot of Momentum of Track Position for Northeast_C Race 1 2014 for Drivers based on Finishing Position

The Northeast_C Race 1 in 2014 demonstrated that most drivers had a consistently positive or negative driver momentum. Driver momentum was calculated across roughly 5 epochs of the 305-lap race for each driver. Taking reference from the first driver, driver momentum was captured at around laps 104, 147, 203, 243, and 292. The results from this race correlated with our previous observation that driver momentum was largely fixed throughout the race.

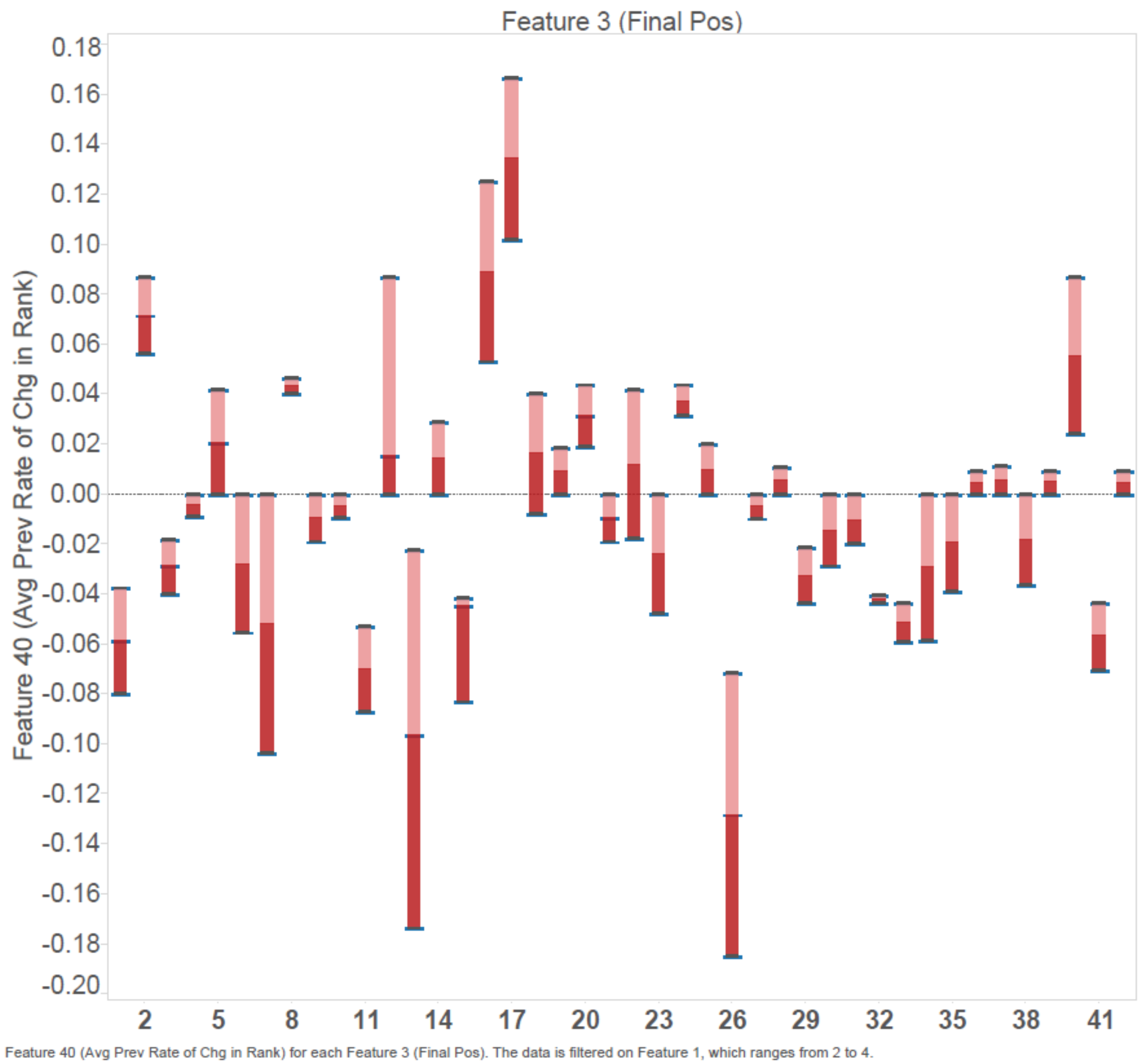
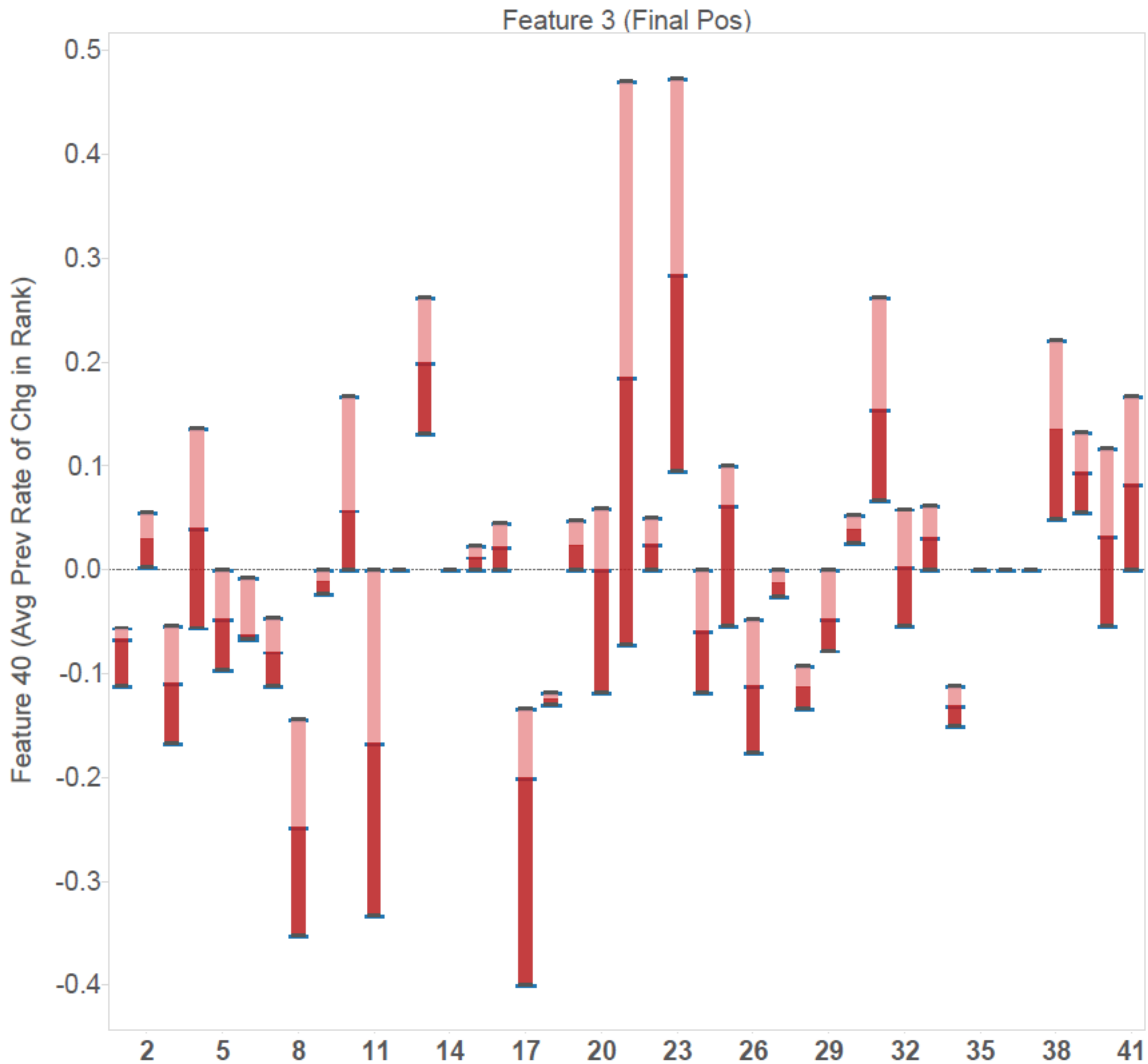


Figure 26. Box Plot of Momentum of Track Position for Northeast_C Race 2 2014 for Drivers based on Finishing Position

The Northeast_C Race 2 in 2014 also showed that most drivers had a consistently positive or negative driver momentum. Driver momentum was calculated across roughly 4 epochs of the 303-lap race for each driver. Taking reference from the first driver, driver momentum was captured at around laps 99, 162, 239, and 286. The results from this race also correlated with our previous observation that driver momentum was largely fixed throughout the race.



Feature 40 (Avg Prev Rate of Chg in Rank) for each Feature 3 (Final Pos). The data is filtered on Feature 1, which ranges from 2 to 4.

Figure 27. Box Plot of Momentum of Track Position for Midwest_A Race 1 2014 for Drivers based on Finishing Position

The Midwest_A Race 1 in 2014 also showed that most drivers had a consistently positive or negative driver momentum. Driver momentum was calculated across 5 epochs of the 200-lap race for each driver. Taking reference from the first driver, driver momentum was captured at around laps 66, 96, 140, 177, and 200. The results from this race again correlated with our previous observations that driver momentum was decided early on in the race.

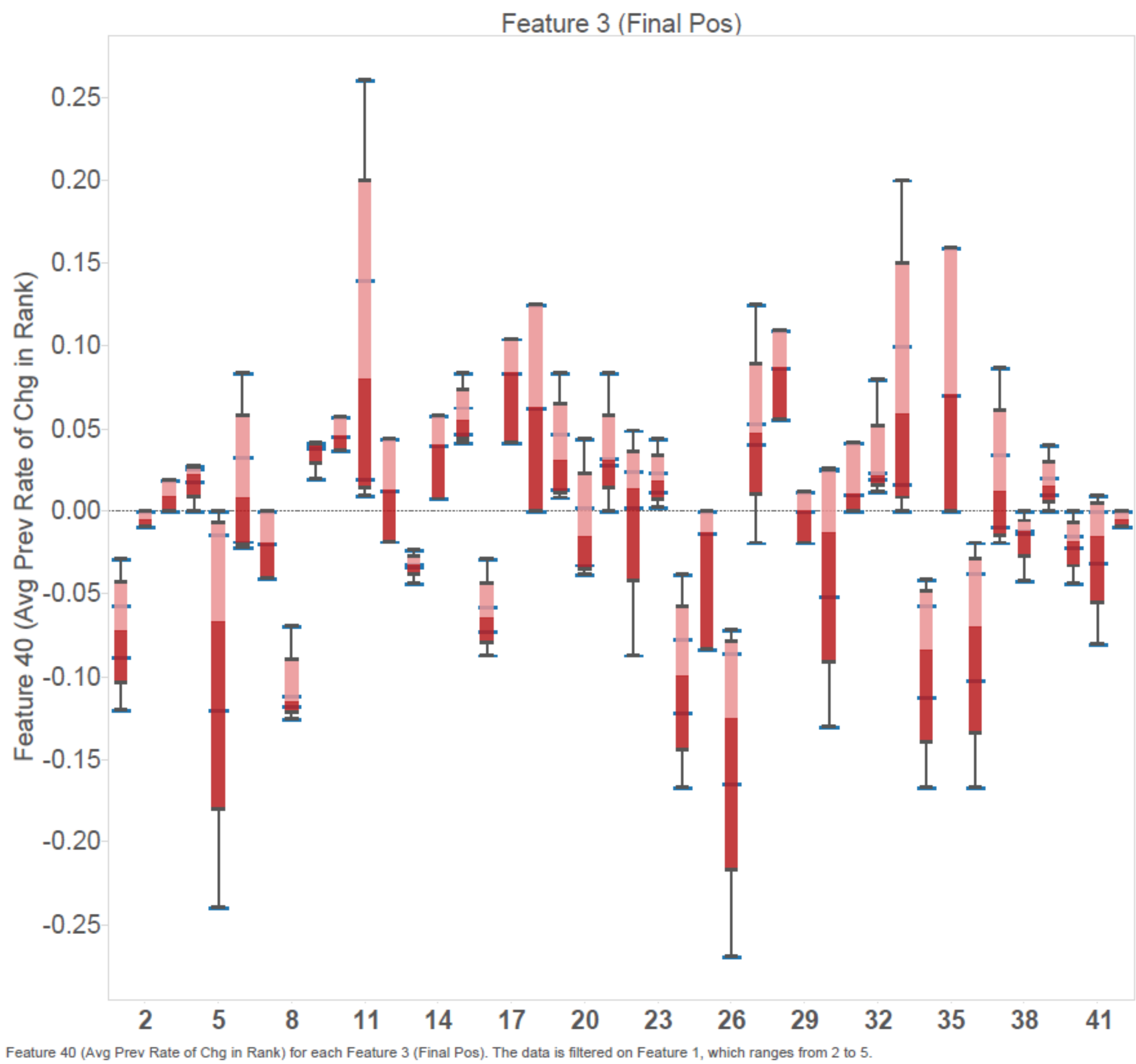


Figure 28. Box Plot of Momentum of Track Position for Southwest_A Race 1 2014 for Drivers based on Finishing Position

The Southwest_A Race 1 in 2014, like the previous box plots, also showed that most drivers had a consistently positive or negative driver momentum. Driver momentum was calculated across 3 or 4 epochs of the 312-lap race for each driver. Taking reference from the first driver, driver momentum was captured at around laps 98, 155, 231, and 268. The results from this race also correlated with our previous observation that driver momentum was largely fixed in the early part of the race.

It appeared that the pattern of driver momentum applied to low tire wear tracks. To understand whether this phenomenon applied across **high tire wear** tracks for races in 2014, we also plotted the graphs for the following races in addition to the Southwest_B Race 2 and Southeast_F Race 2 races that were plotted at the beginning of this section:

1. Southeast_F, Race 1
2. Southeast_I
3. West_C

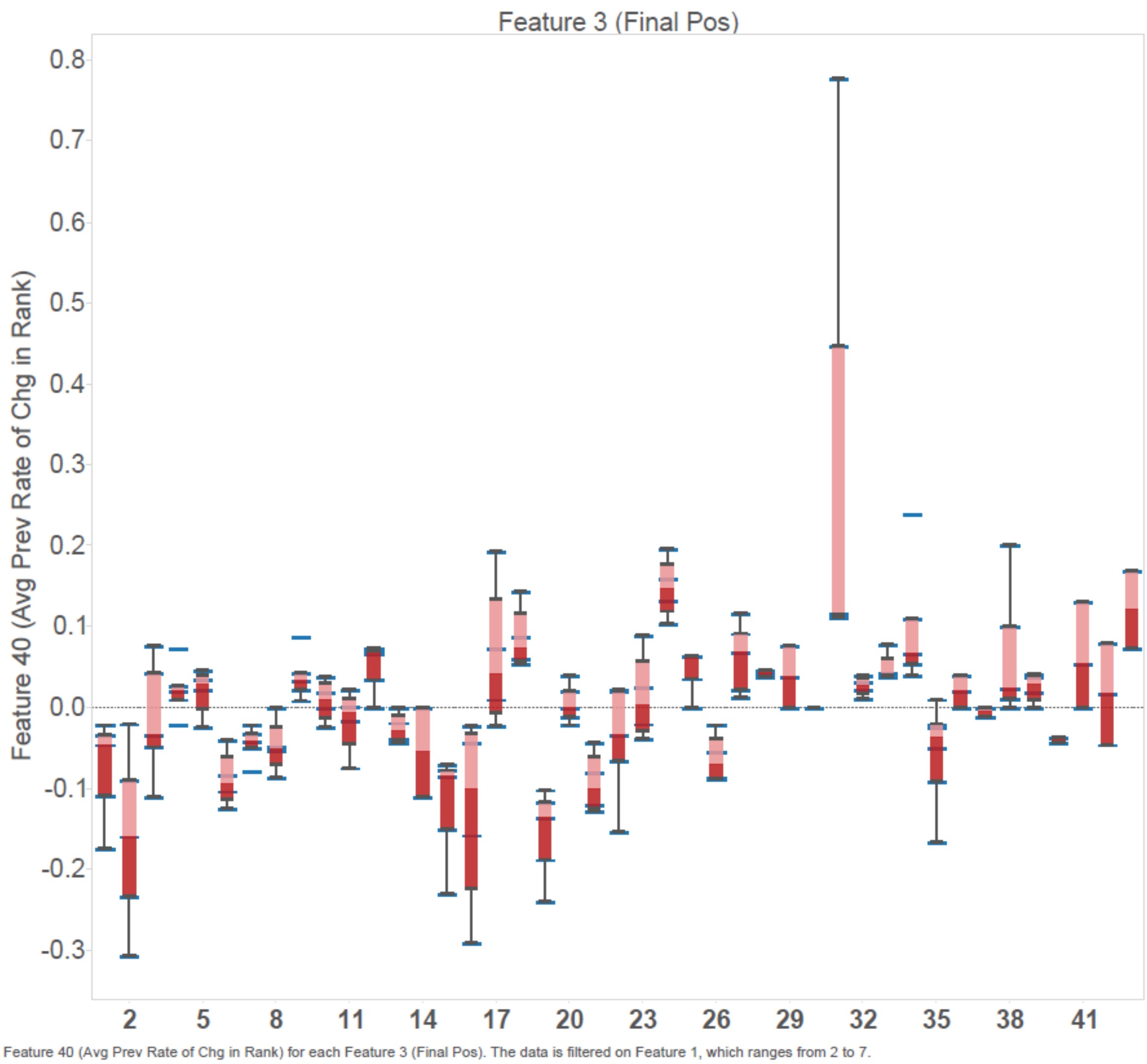


Figure 29. Box Plot of Momentum of Track Position for Southeast_F Race 1 2014 for Drivers based on Finishing Position

The Southeast_F Race 1 in 2014 shown in Figure 29 showed a similar pattern to the earlier box plots. More than three quarters of the drivers registered either positive or negative momentums based on the first to third quartiles. Driver momentum was calculated across 5 epochs of the 400-lap race for each driver. Taking reference from the first driver, driver momentum was captured at around laps 93, 154, 219, 289, and 360.

The Southeast_I race in 2014 is represented in Figure 30 and shows a similar pattern to Southeast_F Race 1. More than three quarters of the drivers had either positive or negative momentums based on the measurements from the first to third quartiles. Driver momentum was calculated across 7 epochs of the 335-lap race. The first driver's epochs were tabulated at laps 71, 109, 164, 199, 248, 285, and 318.

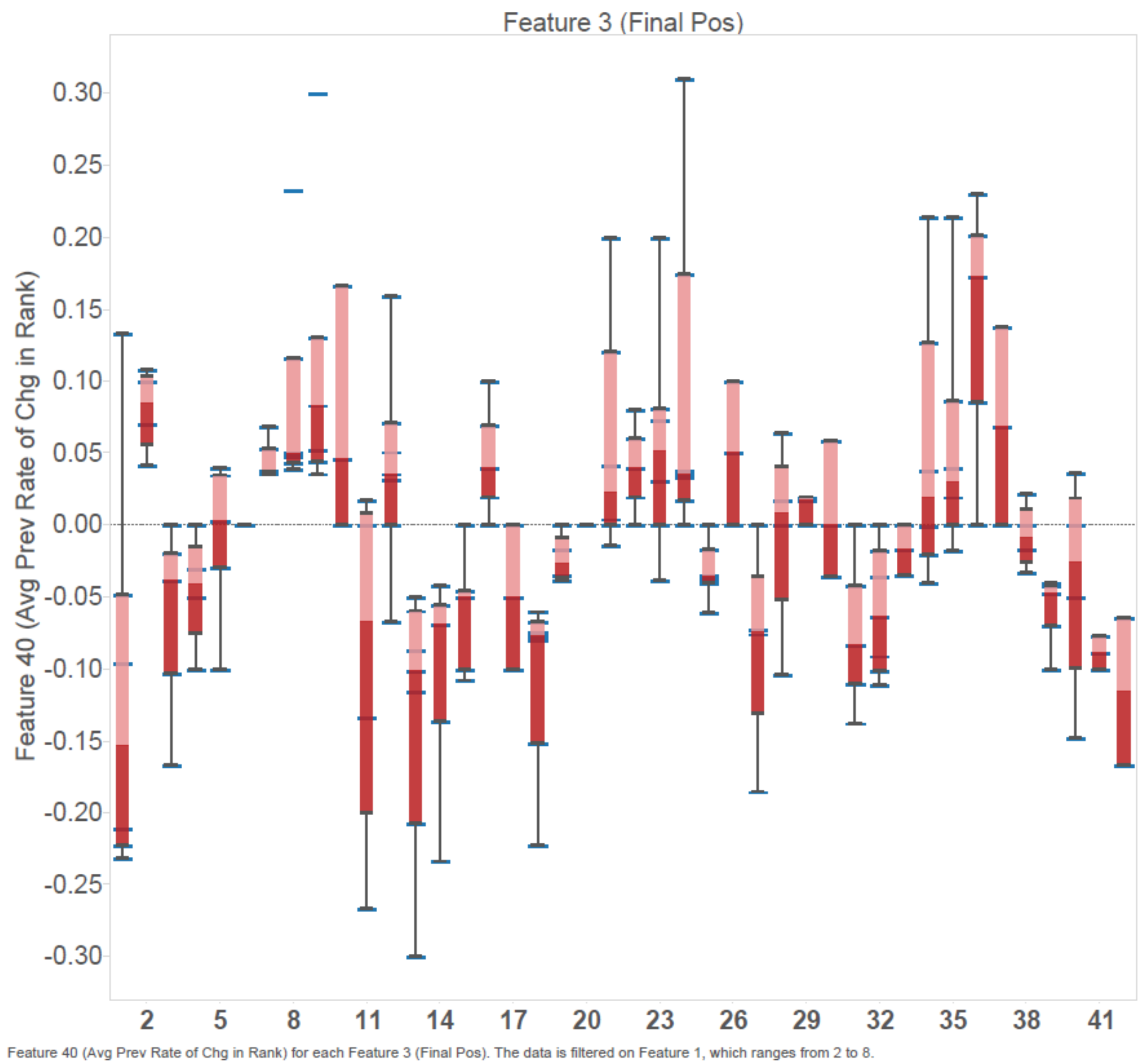


Figure 30. Box Plot of Momentum of Track Position for Southeast_I Race 2014 for Drivers based on Finishing Position

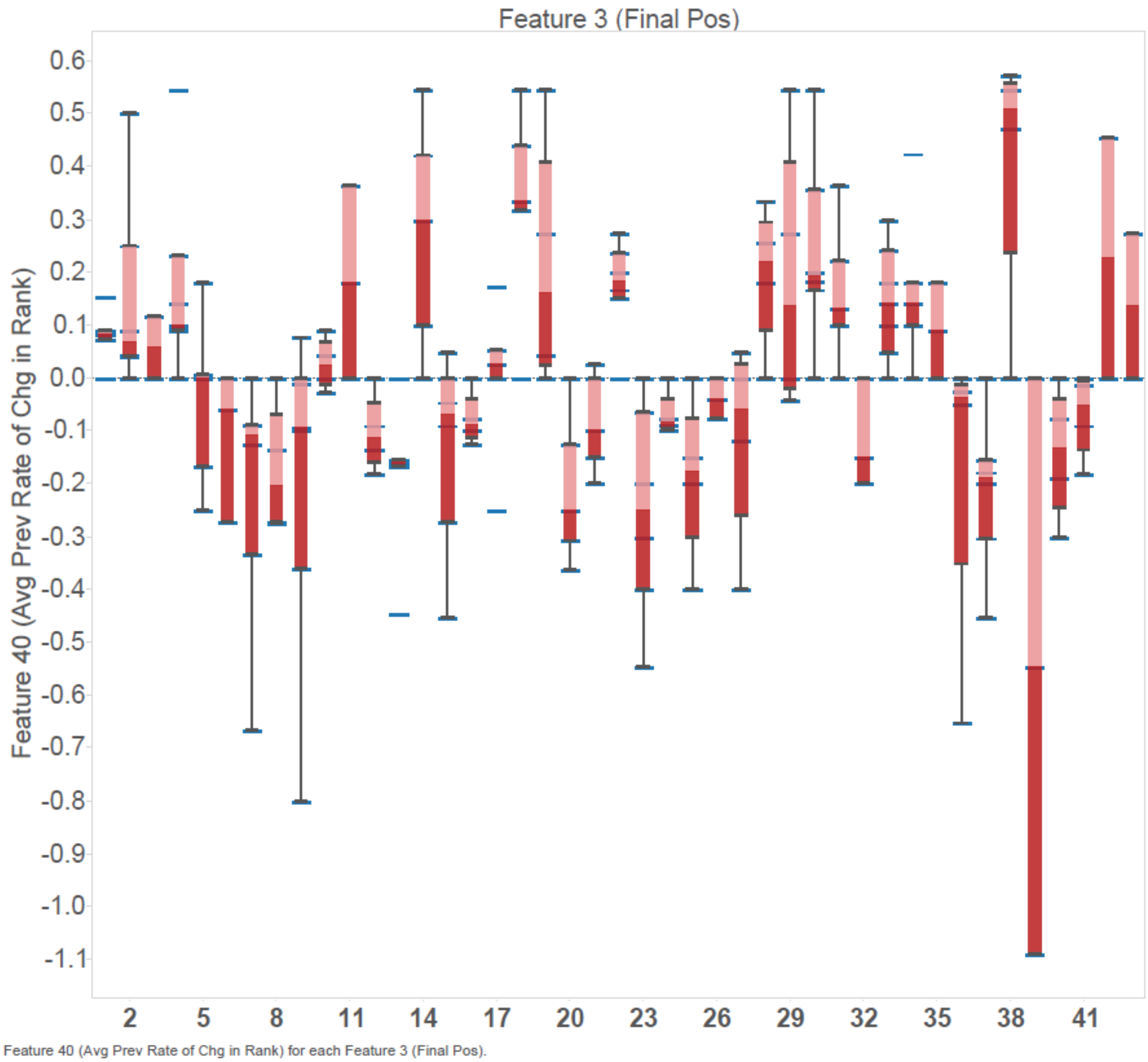


Figure 31. Box Plot of Momentum of Track Position for West_C Race 2014 for Drivers based on Finishing Position

Finally, we looked at the West_C race in 2014 shown in Figure 31. The results were again similar to what we saw for the other races. More than three quarters of the drivers had either positive or negative momentums based on the measurements from the first to third quartiles. Driver momentum was calculated across 6 epochs of the 206-lap race. The top driver’s epochs used in the calculation ended at laps 37, 85, 111, 135, 161, and 192.

As a result of our observations across several races, and taking into account both high tire wear and low tire wear tracks, we believe that driver momentum is generalizable across a majority of drivers in NASCAR Cup races. This has several possible implications:

1. The performance of each driver can be estimated in the early part of each race. Drivers that gain track position usually continue doing so throughout the race, while those that lose track position often continue to do so throughout the race.

2. If driver momentum can be predicted early in the race, it may be possible for drivers to strategize their overtaking and pit stop decisions on tracks with very unique race strategies such as Southeast_B and Southeast_A. According to our industry sponsor, in such races, some drivers deliberately stay behind the pack to conserve their car, and only start driving more aggressively during later stages of the race.

3. Driver momentum plots could be coupled with variables such as pit stop times, pit road times, and gap time between drivers to provide more fine-grained analysis on the impact of tire change decisions.

3.4 Combining Pit Crew Performance and Driver Momentum

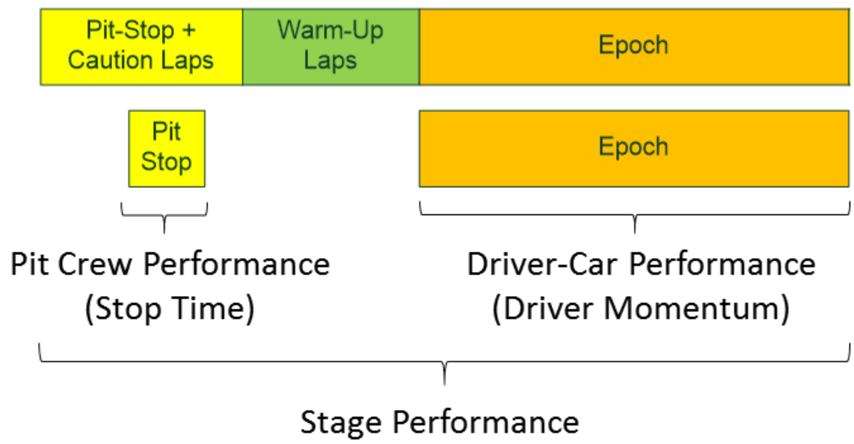


Figure 32. Using Stop Time and Driver Momentum to Estimate Stage Performance

Conceptually, stop times and driver momentum are metrics that measure performance of different parts of each stage of the race, as shown in Figure 32. Stop times determine pit crew performance while driver momentum measures driver-car performance.

Based on observations made regarding pit stop timing in relation to finish positions, it appears that good performance within epochs could cancel out poor pit crew performance and vice-versa. Hence, moving up one level to characterize the performance of each driver from the perspective of a stage of the race, we decided to combine both features to see if our findings could tell us more about overall race performance.

We selected a few races to plot driver momentum against stop times:

1. Southwest_B, Race 2, 2014
2. Northeast_D, Race 1, 2014
3. West_C, 2014
4. Northeast_C, Race 2, 2014
5. Midwest_A, Race 2, 2014

We developed scatter plots by selecting only the medians of stop times for 4-tire pit stops and driver momentum. 4-tire pit stop times were selected instead of 2-tire pit stop times because of the higher frequency of 4-tire pit stops. The median of driver momentum was taken as it was easier to compare the performance of different drivers when each of them was represented by a single data point in the plot. We also colored the data points based on the finish positions in groups of 5 drivers each, from 1st to 20th. This allowed us to cluster drivers visually into different performance brackets.

Drivers clustered at the bottom-left corner of the plot were those with the best driver momentum and 4-tire pit stop times. We thought that a correlation between finish position and aggregated driver momentum and stop time performance was very likely.

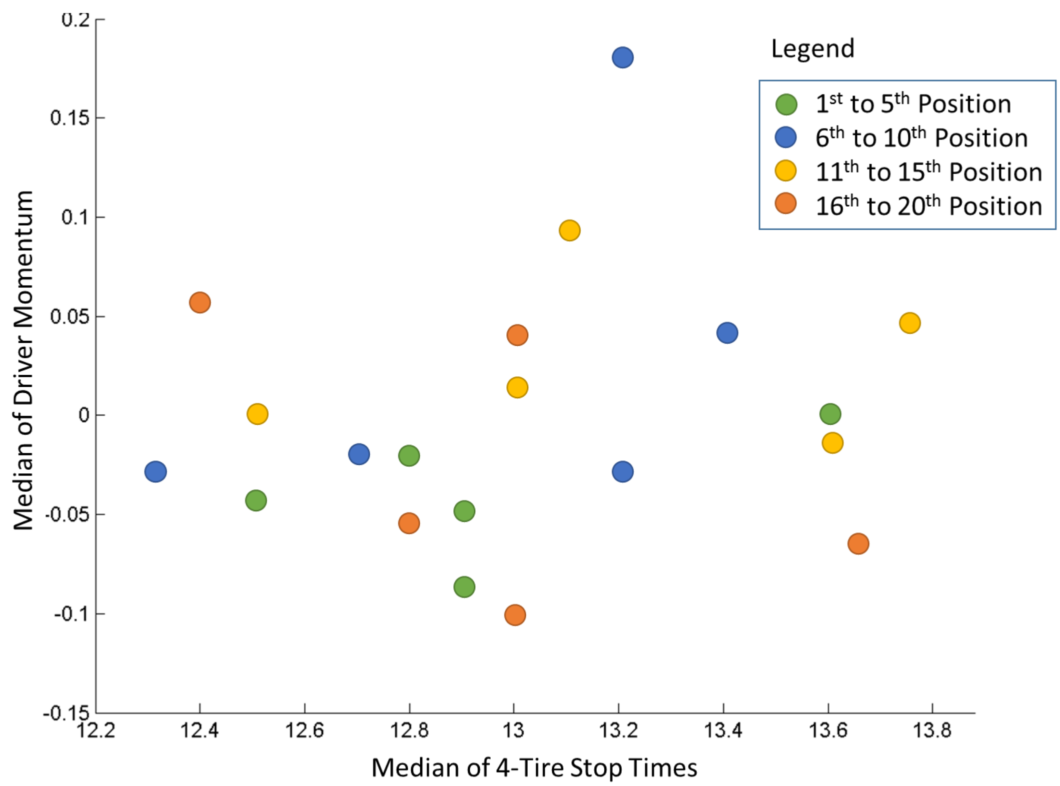


Figure 33. Scatter Plot of Median Stop Times against Median Driver Momentum based on Final Track Position for Southwest_B Race 2 2014

In the Southwest_B Race 2 in 2014 as shown in Figure 33, the top 5 drivers had the best driver momentum compared to the other groups of drivers. 4 out of the top 5 drivers were clustered at the bottom left corner of the graph, and the same pattern was seen for the 6th to 10th-placed drivers. The pattern was not as clearly visible when comparing the 11th to 15th drivers against the 16th to 20th.

Although there was a small indication that drivers with better 4-tire pit stop time and driver momentum tended to end the race with better finish positions, we wanted to see if the correlation between driver finish positions and this combination of features was clearer in other races.

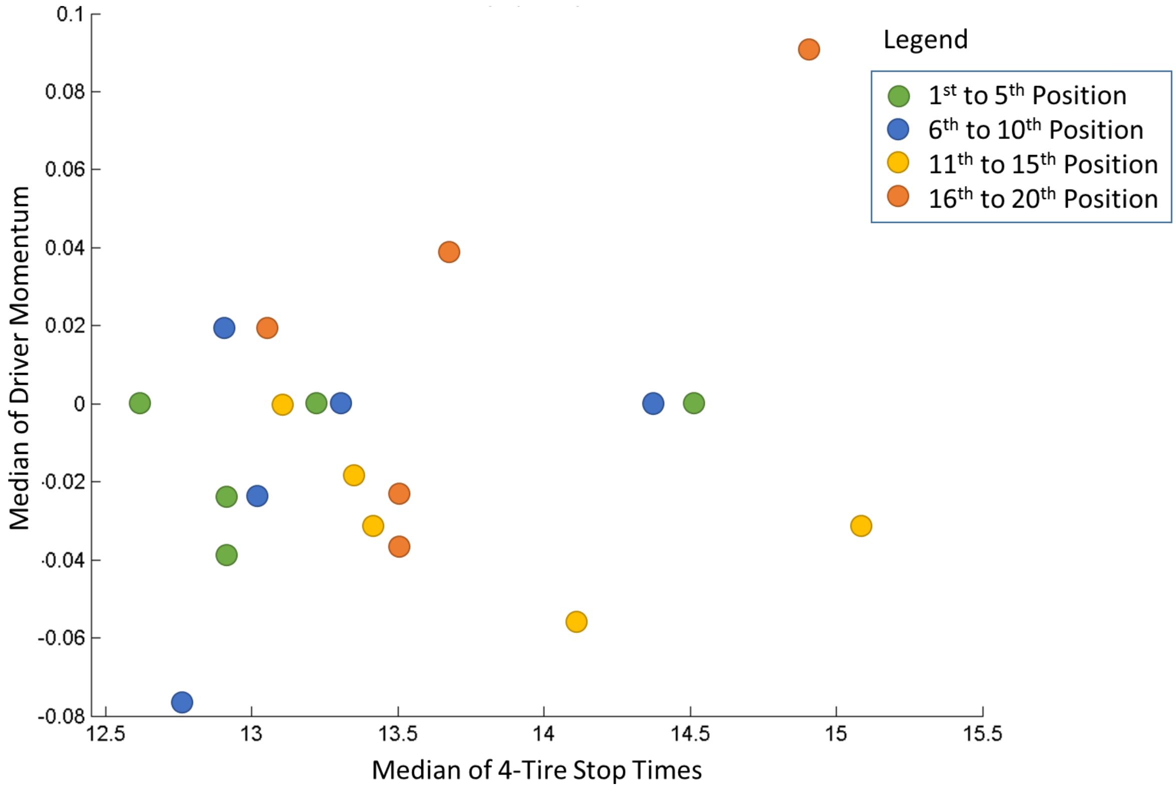


Figure 34. Scatter Plot of Median Stop Times against Median Driver Momentum based on Final Track Position for Northeast_D Race 1 2014

In the Northeast_D Race 1 in 2014 as shown in Figure 34, 4 out of the top 5 drivers were clustered at the bottom left corner of the graph, and the same pattern was seen for the 6th to 10th-placed drivers. In this particular race, the 11th to 15th drivers were clustered more closely to the bottom left corner of the graph compared to the 16th to 20th, indicating that the pattern of combined driver momentum and 4-tire pit stop times held true across all 4 clusters of drivers.

As for the West_C race in 2014 as shown in Figure 35, the top 5 drivers did not have the best driver momentum and pit stop time combination. Instead, it was the 6th to 10th-placed drivers that fell mostly within the bottom left corner of the scatter plot. There was no substantial difference between the 11th to 15th and 16th to 20th-placed drivers, but when the 1st to 10th and 11th to 20th drivers were considered as 2 groups instead of 4, the cluster of the top 10 drivers tended to be closer to the bottom left corner compared to the 11th to 20th.

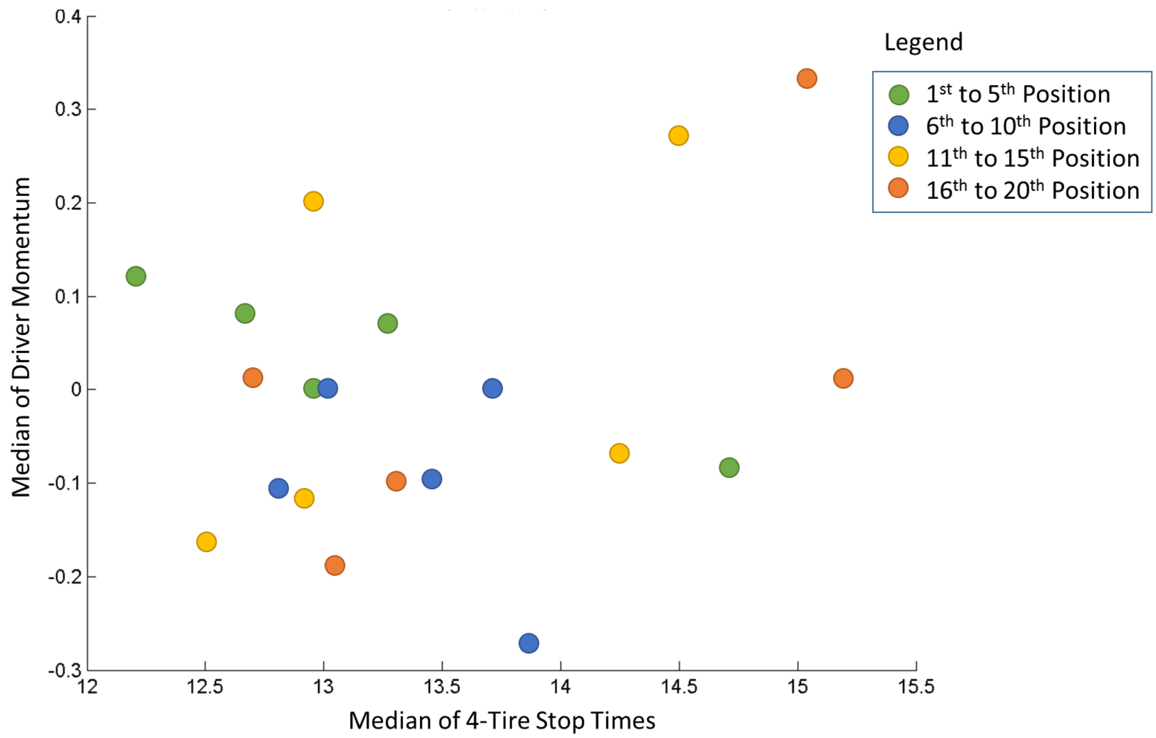


Figure 35. Scatter Plot of Median Stop Times against Median Driver Momentum based on Final Track Position for West_C Race 2014

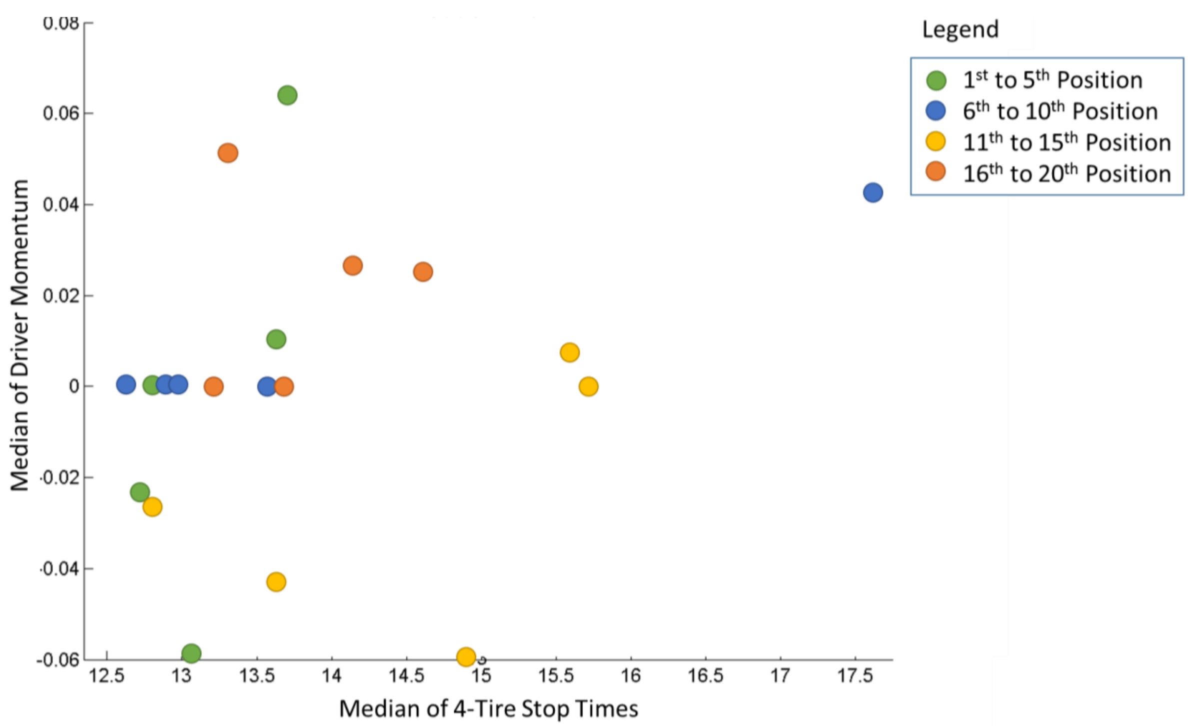


Figure 36. Scatter Plot of Median Stop Times against Median Driver Momentum based on Final Track Position for Northeast_C Race 2 2014

A relatively large number of drivers had a median driver momentum of 0 in the Northeast_C2 race in 2014 as shown in Figure 36. However, there was some differentiation when comparing their 4-tire pit stop times across the different clusters of drivers. 4 of the top 5 drivers were close to or within the bottom left corner of the plot, and 4 of the 6th to 10th drivers were the same as well. As for the next 10 drivers, the 11th to 15th-placed drivers performed comparatively better than the 16th to 20th-placed drivers.

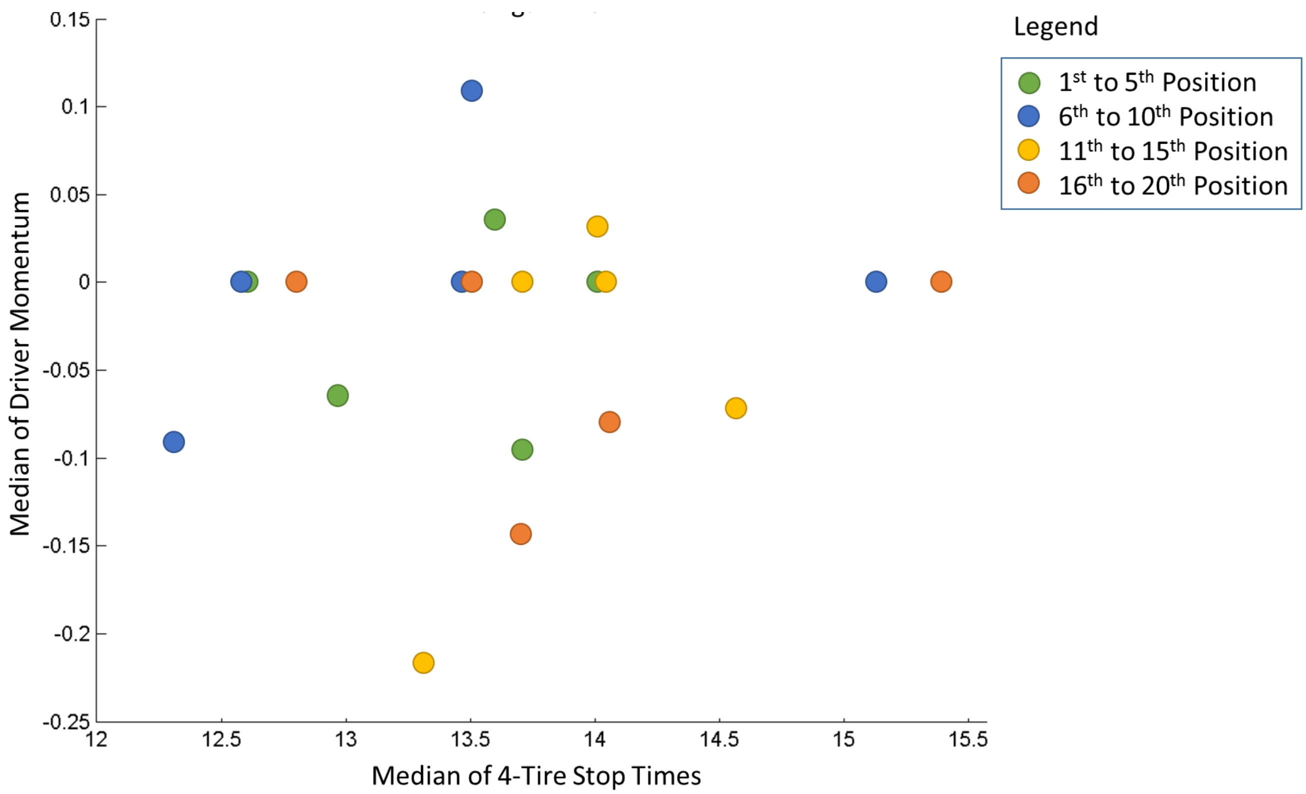


Figure 37. Scatter Plot of Median Stop Times against Median Driver Momentum based on Final Track Position for Midwest_A Race 2 2014

Lastly, we studied the Midwest_A Race 2 of 2014 in Figure 37 to see if the general patterns from the previous races were visible. In this instance, the top 5 drivers had better driver momentum and 4-tire pit stops than the 6th to the 10th drivers as their plots were comparatively closer to the left of the graph. However, the pattern was not visible when comparing the 11th to 15th and 16th to 20th-drivers.

Nonetheless, across the 5 races we studied, it could be possible to generalize the insight that drivers who finished closer to the top tended to have better combinations of driver momentum and 4-tire pit stop times. We believed that it would be useful to introduce the combination of median 4-tire stop times and driver momentum as an additional feature in the machine learning software.

4. Analysis of Race Characteristics

4.1 Tire Degradation

Although there are 36 races in the NASCAR season, some races run on the same track more than once. Circuits such as those in Southwest_A, Southwest_B, Southeast_D, Midwest_C, and Southeast_E, organize two NASCAR races each season. As a result, there are 23 tracks, mostly oval-shaped, but of different lengths, used for racing.

Characteristics of each race track have a large impact on tire change decisions. Depending on factors such as the speed of the car, the banking of the track, the mix of the asphalt used, and other factors, tires will degrade at different rates. In addition, racing tires are designed to wear, and as cars race along the track, they also lay down rubber, which improves the grip of the track over time.

Ultimately, these factors influence the top speeds that the cars are able to achieve each lap. And, as the tires begin to wear, the cars' top speeds tend to decrease. All this is measured based on the lap times that the cars clock for each lap. Without measuring the thickness of the tires directly to determine the actual rate of tire wear, we can infer the rate, or **% degradation**, by calculating the change in lap times from one lap to the next:

$$\frac{t_{lap} - t_{lap-1}}{t_{optimal}} \times 100\%$$

Calculated over the course of an outing, a linear regression graph can be generated to calculate the effect of the track on tire wear, with $t_{lap} - t_{lap\text{-}1}$ represented by the slope, and $t_{optimal}$ represented by the y-intercept. An example of the linear regression for the track in Southwest_A using data from the race in 2014 is seen in the following chart.

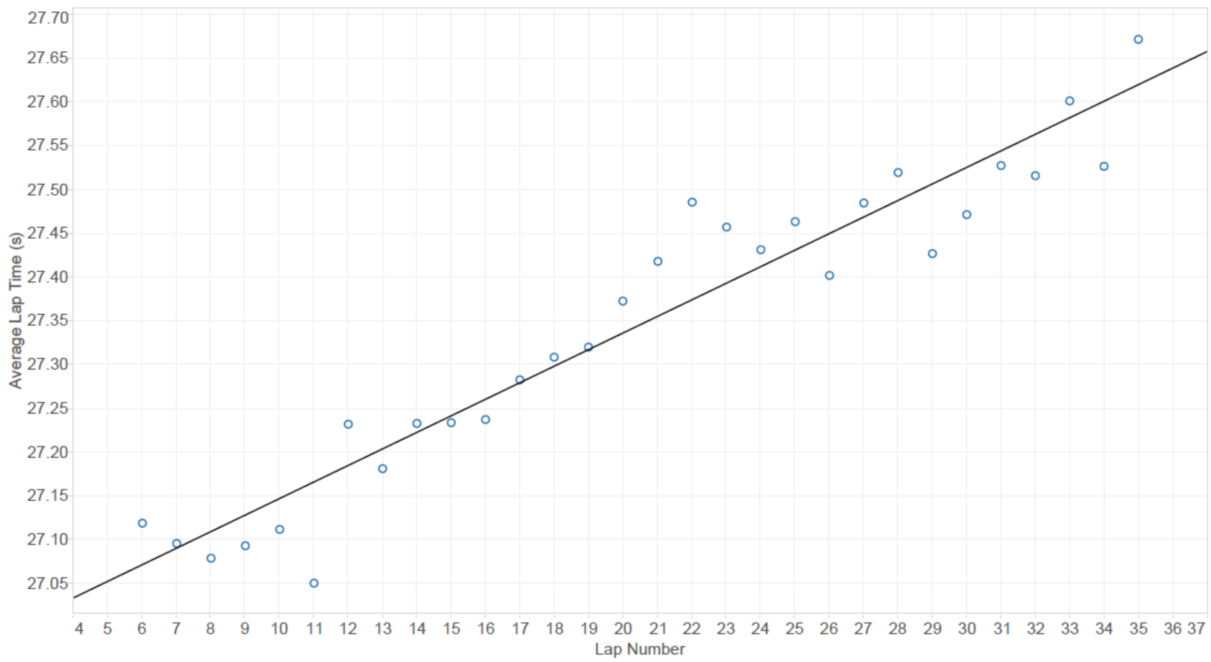


Figure 38. Linear Regression for Southwest_A Race 1 2014, based on the Average Lap Time of the Top 20 Finishers from Laps 6 through 35

There are several factors that need to be taken into account when generating the linear regression graph for each track. These result in a number of assumptions that we make in determining the track's tire degradation.

First, we measure tire wear based on the lap times from early in the race. This is done because the characteristic of the track changes over time as more rubber is laid down on the surface as the cars race around the track. In addition, all cars run on fresh tires at the beginning of the race, compared to later in the race when different tire change strategies come into play, resulting in some cars having fresher tires than others.

Second, we ignore the lap times for the bottom half of the field, and only select the average lap times of the top 20 drivers based on their finishing positions. Accidents and car failures tend to happen throughout the race, and it would not be appropriate to include the lap times of cars that are heavily damaged or who enter the pits frequently for repair work. Moreover, the top half of the field of drivers tend to have higher chances of winning the race as those at the bottom, so it is more useful to average the lap times for those which have good driver-car performance.

Third, at least the first 3 laps of the race are dropped from the regression graph. This is necessary as cars are heavily bunched up at the beginning of the race. As a result, lap times tend to be slower as cars jostle for track position in the initial laps. In addition, cars need to be running at higher speeds in order to heat up their tires to perform optimally.

Fourth, we remove caution laps when an accident occurs, debris is spotted on track, or a scheduled yellow flag is enforced. Drivers slow down substantially during these laps to follow the pace car and allow race officials to safely clear hazards from the track.

Fifth, lap time samples from near the next pit stop onwards are removed so that we can capture optimal performance of the current set of tires only. Lap times increase dramatically as cars slow down to enter the pits, change tires, and refuel. Similarly, very heavy tire wear will cause drivers to slow down substantially before entering the pits.

Sixth, outliers in the plot are manually removed. We do this because there may be times when cars lap the track more slowly than usual, such as when drivers are bunched together and lots of jostling for track position occurs. Drivers typically lap the track more slowly when they are defending their position or trying to overtake slower cars in front.

Lastly, samples of lap times from a subsequent leg of the race are used whenever the starting leg is littered with cautions or pit stops. We observed this on tracks such as West_A, Southeast_G, Southwest_B, Southeast_A, Midwest_A, and Northeast_A. In those instances, we calculated an adjusted $t_{optimal}$ based on the original y-intercept and slope using the following calculation:

$$Adjusted\ t_{optimal} = t_{optimal} + slope\ \times (lap - 4)$$

For this calculation, the next earliest leg of the race is used where most, if not all the top 20 cars are driving on 4 fresh tires after a caution is lifted. Similar to the calculation for $t_{optimal}$, the first 3 laps are ignored, resulting in the adjusted optimal lap time taken at lap 4. This allows us to compare the results of tire degradation across all 23 tracks.

Graphs generated for each of the 23 race tracks can be found in Section 9.3. Their respective tire degradations are shown in Figure 39, starting from the track with the lowest tire wear (Southeast_A) to the highest tire wear (Midwest_E).

Track	% Degradation	Intercept $T_{optimal}$	Adjusted $T_{optimal}$	Slope	Standard Deviation	p-value
Southeast_A	0.007%	48.6534	48.8471521	0.0036557	0.002978	0.228563
Northeast_A	0.033%	70.6622	71.3199116	0.0234897	0.005541	0.0008251
Southeast_B	0.045%	45.8644		0.0206142	0.00926	0.0429453
Southeast_C	0.053%	15.6563		0.0082956	0.000503	< 0.0001
Midwest_A	0.055%	37.8045	37.9920564	0.0208396	0.002097	< 0.0001
Midwest_B	0.055%	30.6209		0.0168518	0.001076	< 0.0001
Northeast_B	0.055%	52.7364		0.0290275	0.003139	< 0.0001
Midwest_C	0.063%	30.1902		0.0191001	0.002543	< 0.0001
Northeast_C	0.067%	29.2737		0.0195242	0.000593	< 0.0001
Southwest_A	0.070%	26.9578		0.0189037	0.001051	< 0.0001
West_A	0.079%	29.0689	30.1903388	0.0238604	0.001861	< 0.0001
Northeast_D	0.086%	23.5551		0.020243	0.000666	< 0.0001
Midwest_D	0.095%	30.5025		0.028877	0.001924	< 0.0001
Southeast_D	0.128%	29.4007		0.0376473	0.00123	< 0.0001
Southeast_E	0.145%	19.5991		0.0284421	0.001891	< 0.0001
Southeast_F	0.177%	22.222		0.0393687	0.001738	< 0.0001
Southeast_G	0.183%	31.3084	32.0099848	0.0584654	0.001418	< 0.0001
West_B	0.190%	76.733		0.145521	0.007296	< 0.0001
Southwest_B	0.197%	27.9639	29.1683256	0.0573536	0.001589	< 0.0001
West_C	0.206%	40.3092		0.0830462	0.006846	< 0.0001
Southeast_H	0.213%	29.0498		0.0618795	0.001946	< 0.0001
Southeast_I	0.232%	30.8645		0.0715526	0.002702	< 0.0001
Midwest_E	0.236%	50.3967		0.119024	0.004839	< 0.0001

Figure 39. List of Tracks based on Tire Degradation, with the T(optimal), the Slope of Time Degradation, and its p values

We produced the following graph in Figure 40 that visually shows the tire degradation for each track.

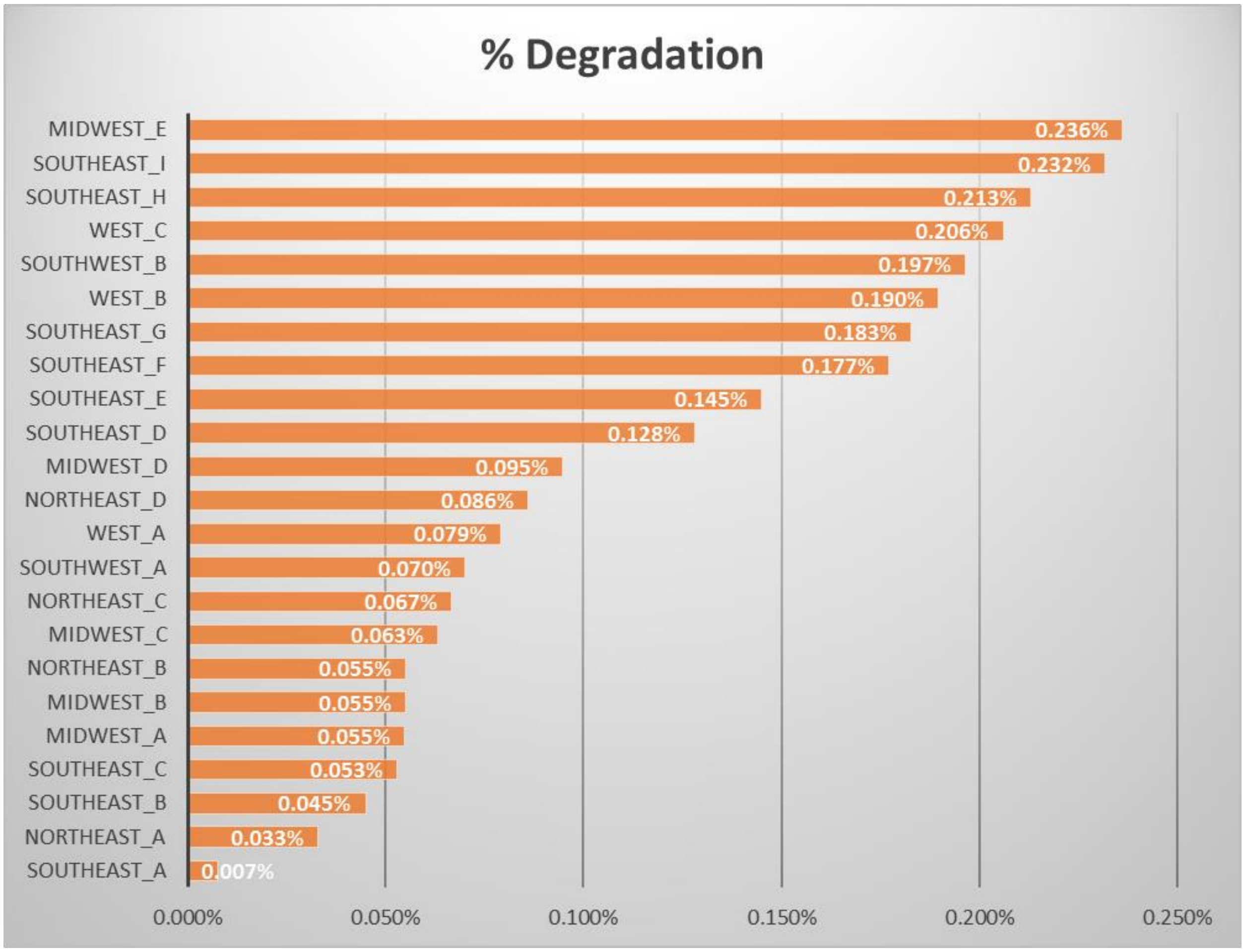


Figure 40. 2014 Tracks ordered by % Tire Degradation

We proceeded to classify tracks based on 4 categories: (1) Very low, (2) low, (3) medium, and (4) high tire wear. The classifications in Figure 41 are:

Very Low Tire Wear	Low Tire Wear	Medium Tire Wear	High Tire Wear
1. Southeast_A	1. Northeast_A	1. West_A	1. Southeast_F
	2. Southeast_B	2. Northeast_D	2. Southeast_G
	3. Southeast_C	3. Midwest_D	3. West_B
	4. Midwest_A	4. Southeast_D	4. Southwest_B
	5. Midwest_B	5. Southeast_E	5. West_C
	6. Northeast_B		6. Southeast_H
	7. Midwest_C		7. Southeast_I
	8. Northeast_C		8. Midwest_E
	9. Southwest_A		

Figure 41. Classification of 2014 Tracks based on Tire Degradation

4.2 Ratio of Tire Change Decisions

Another method to determine the tire strategy used for each race track is to study historical records of tire change decisions used for previous races on the same track. This is particularly useful when factors apart from tire wear degradation act in concert with other track or race characteristics. For instance, cars may decide to stay out of the pit under regular racing conditions because pitting may result in rivals lapping them due to the low lap times. This would be a bigger consideration for tracks such as Southeast_C and Southeast_E, and less so for tracks such as West_B and Northeast_A.

Figure 42 lists all races in 2014, the track they raced on, and the number of 0, 2, and 4 tire-change pit stops made by all drivers participating in each race. When 2014 data was not available, data from the corresponding race in 2013 was used instead. The ratio of 2 versus 4 tire-change pit stops was calculated, with results from the previous section on tire degradation included in the last column for reference. The table is arranged in descending order by the ratio of 2 versus 4 tire-change pit stops, which allowed us to identify tracks where different tire change strategies were a real possibility.

Race Name	0-Tire Stops	2-Tire Stops	4-Tire Stops	Ratio of 2 versus 4 Stops	Tire Wear
Southeast_B, Race 1	74	96	123	0.78	Low
Southeast_A, Race 2	15	65	87	0.75	Very Low
Northeast_C, Race 1	2	56	78	0.72	Low
Southeast_A, Race 1	82	88	144	0.61	Very Low
Southeast_B, Race 2	51	59	99	0.60	Low
Southwest_A, Race 1	26	76	153	0.50	Low
Midwest_A, Race 1	52	69	140	0.49	Low
Southwest_A, Race 2	25	95	196	0.48	Low
Northeast_C, Race 2	49	69	202	0.34	Low
Northeast_B, Race 2	32	64	192	0.33	Low
Midwest_D	9	18	58	0.31	Medium
Midwest_A, Race 2	47	58	189	0.31	Low
Midwest_C, Race 1	246	57	197	0.29	Low
Southeast_C, Race 2	18	54	208	0.26	Low
Midwest_E	10	41	185	0.22	High
Northeast_B, Race 1	29	43	217	0.20	Low
Midwest_C, Race 2	18	41	217	0.19	Low
West_A	16	26	154	0.17	Medium
Northeast_D, Race 1	21	36	246	0.15	Medium
Southeast_E, Race 2	40	24	218	0.11	Medium
Southeast_E, Race 1	36	28	298	0.09	Medium

Midwest_B	9	24	274	0.09	Low
Southeast_H	14	28	386	0.07	High
Northeast_A	33	8	123	0.07	Low
Southeast_G	10	22	360	0.06	High
Southeast_C, Race 1	19	10	169	0.06	Low
Southeast_I	20	19	327	0.06	High
West_B	18	9	164	0.05	High
West_C	20	19	355	0.05	High
Northeast_D, Race 2	17	11	221	0.05	Medium
Southeast_D, Race 1	52	17	342	0.05	Medium
Southeast_D, Race 2	8	11	227	0.05	Medium
Southwest_B, Race 2	8	15	341	0.04	High
Southeast_F, Race 1	12	11	269	0.04	High
Southwest_B, Race 1	34	10	298	0.03	High
Southeast_F, Race 2	3	4	197	0.02	High

* Figures marked in red refer to statistics taken from the corresponding races in 2013 as 2014 data was not available.

Figure 42. List of 2014 Races based on Ratio of 2 versus 4 Tire Pit Stops and Associated Tire Degradation.

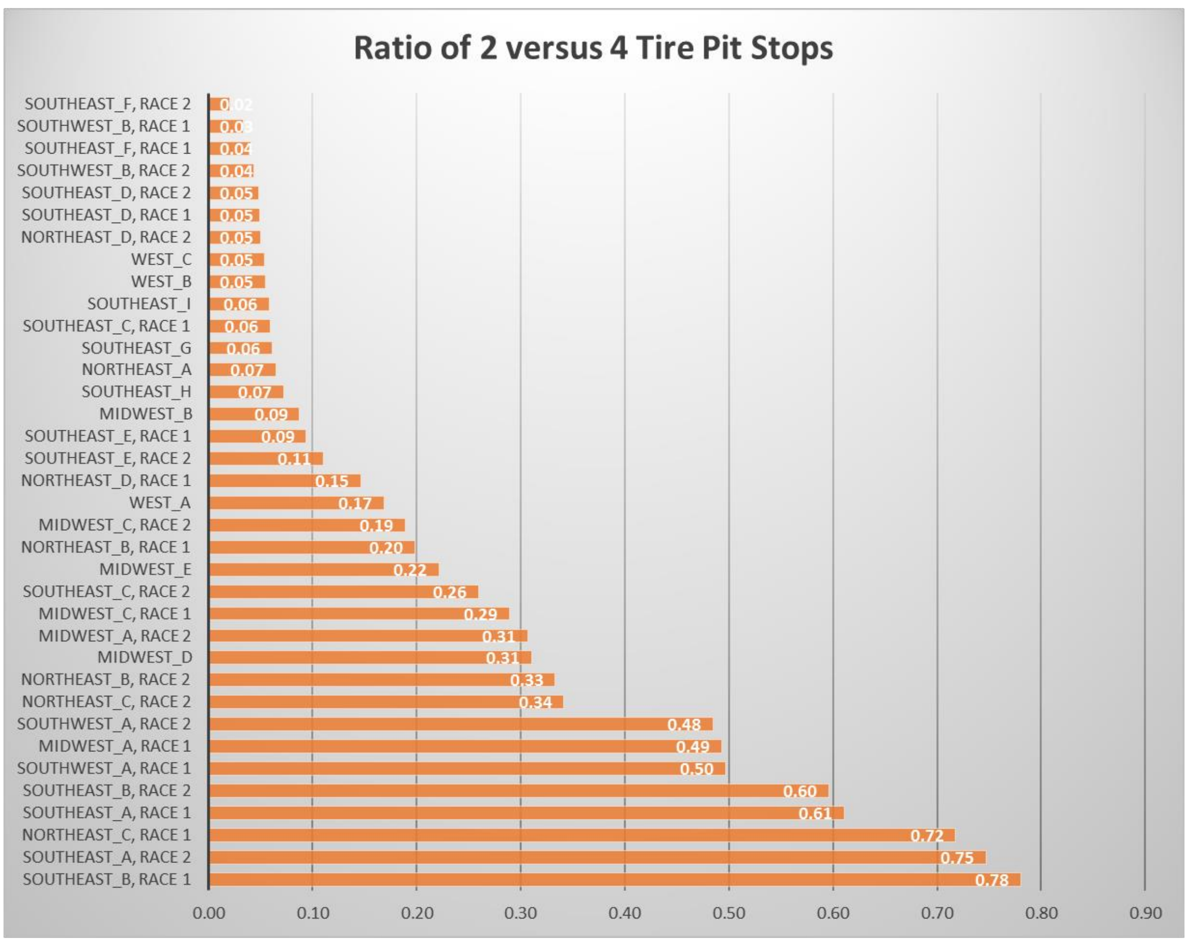


Figure 43. 2014 Races ordered by Ratio of 2 versus 4 Tire Pit Stops

In general, tracks with lower tire wear had higher ratios of 2 to 4-tire change pit stops, indicating that 2-tire changes were a reasonable option when tire degradation was lower. This was evident for tracks such as Northeast_C, Southeast_A, Southeast_B, Southwest_A and Midwest_A. On the other end of the scale, tracks with high tire wear tended to have drivers gravitate towards 4-tire change pit stops. This was obvious for Southeast_F, Southwest_B, Southeast_I, West_B, Southeast_H, and West_C. Drivers who pitted on those tracks very rarely chose to change 2 tires or fewer.

However, there was some variability in the tire change decisions for selected tracks. For instance, the second Northeast_D race had a relatively low ratio compared to the first. In addition, the Midwest_E race was clustered with low tire wear tracks although it was a high tire wear track itself. Such observations suggest that a selected number of tracks require detailed fine-tuning, as predicted tire wear and historical tire change decisions do not correspond with each other.

Nonetheless, given the fact that several tracks exhibit strong attributes where tire wear and tire-change decisions are strongly correlated, appropriate weights should be applied to the following tracks when considering the predictive model, based on the following classifications.

Bias towards 4-tire Pit Stops	No Bias Applied	Bias towards Mixed Strategy
1. Southwest_B	1. Northeast_D	1. Southeast_B
2. Southeast_F	2. Southeast_C	2. Midwest_E
3. West_C	3. Midwest_B	3. Midwest_D
4. Southeast_I	4. Southeast_E	4. Northeast_C
5. Southeast_G	5. West_A	5. Southwest_A
6. Southeast_H		6. Midwest_A
7. West_B		7. Southeast_A
8. Northeast_A		8. Northeast_B
9. Southeast_D		9. Midwest_C

Figure 44. Classification of Tracks based on Ratio of 2 versus 4 Tire Changes

4.3 Changing Fewer Tires or Staying Out when Tires are Changed Recently

It is also particularly important to model behavior based on recent tire change decisions. Drivers may opt to stay out on track or to change two instead of four tires if they had recently made a pit stop. The following chart in Figure 45 illustrates the tire change strategy employed by teams at different stages of the race, based on their track position 4 laps prior to pitting. The horizontal axis represents the pre-pit track position of a driver, and the vertical axis shows the lap number of the 2014 Midwest_B race, which is classified as a track with low tire wear.

Each point on the chart represents a pit stop made by a driver. Green circles are pit stops where 4 tires were changed, orange circles are 2-tire pit stops, and red circles are 0-tire pit stops. Shaded horizontal bands represent caution laps of the race, and the car number and lap that it pitted are labelled next to each circle.

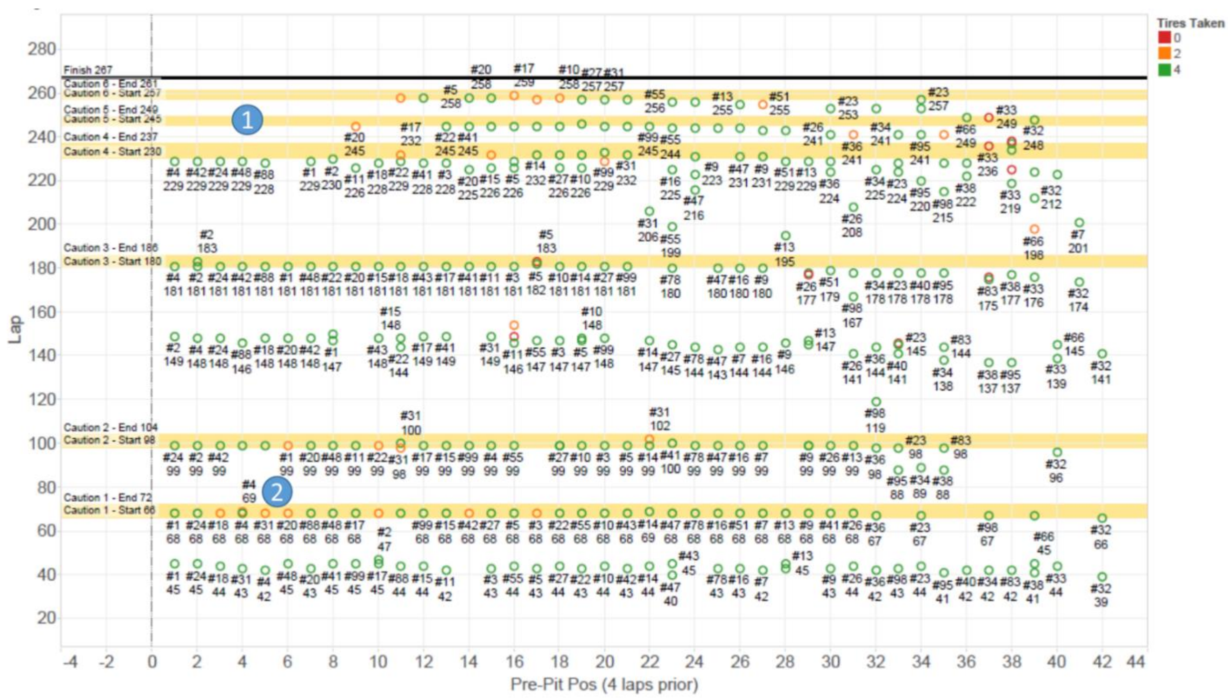


Figure 45. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Midwest_B Race 2014

There are two important observations made for the race. The first is represented by the blue circle labelled "1", which shows that the drivers at the front of the pack refrained from pitting in the final stages of the race. Presumably, this was to maintain their track position because pitting would push them further down the order. Drivers beyond the 10th position probably took chances to change their tires in the hope that the loss of track position due to pitting could be overcome by the increased speeds they obtained whilst running on fresh tires.

The second observation occurs in the blue circle labelled “2” at the second series of pit stops during the first caution flag. Cars #18, #31, and #20 decided to pit for 2 tires instead of 4. This is likely to have been due to the recent series of pit stops around lap 45. Assuming that a fresh set of tires works optimally for 45 laps, they would have been less than halfway through their tires during the pit stop on lap 68.

Next, we studied whether these two phenomenon occurred in the 2014 Southwest_A race as shown in Figure 46. Similar to Midwest_B, the Southwest_A track was characterized as one with low tire-wear. Assuming that our observations for Midwest_B were accurate, drivers at the front of the field should not have pitted in the last leg of the race. Also, if tire wear on the track was sufficiently low, drivers could choose to change 2 or fewer tires during a pit stop, provided that the previous pit stop was done recently.

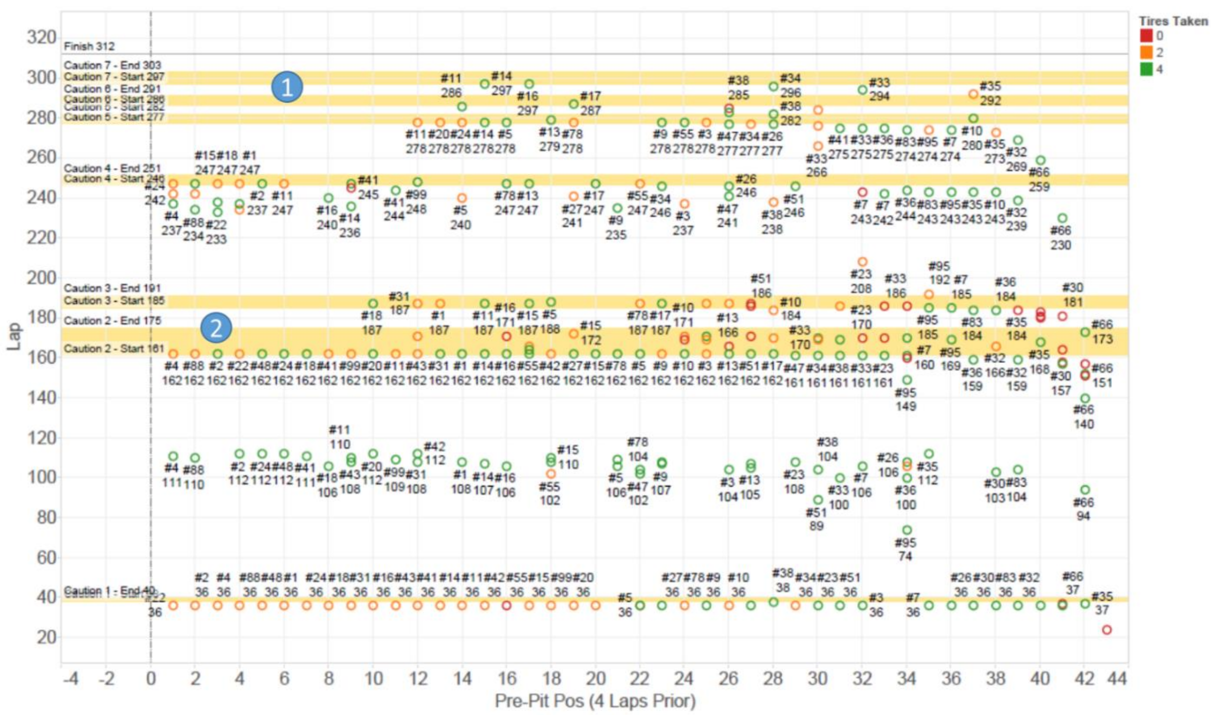


Figure 46. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southwest_A1 Race 2014

Indeed, cars at the front of the field did not pit during the last stage of the race, even when a caution occurred. This area is indicated by the blue circle labelled “1”. The blue circle labelled “2” also confirms our observation that as long as pit stops were made recently, there would be an option for drivers to change just 2 tires instead of 4.

We needed to study whether these two observations emerged in other circumstances as well. For example, if a track with low tire wear did not have cautions near the end of the race, would drivers at the front of the pack still insist on staying out? Another sample, this time from the 2014 Midwest_C Race 1 was reviewed in Figure 47.

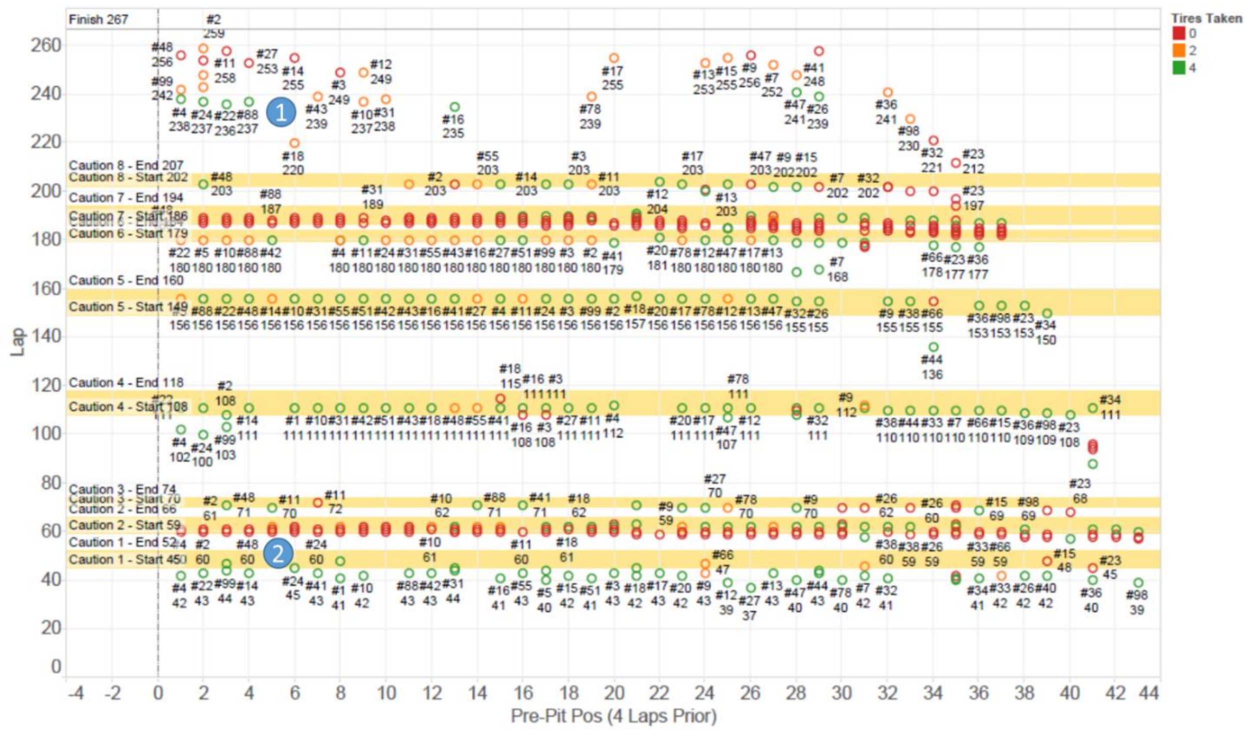


Figure 47. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Midwest_C Race 1 2014

Based on observations in the Midwest_C race, it was clear that drivers had the flexibility to change fewer than 4 tires when a pit stop was made recently. This was indicated by the pit stops made around the second and third cautions indicated by the blue circle labelled “2”. On the other hand, for the last stage of the race, drivers at the front pitted despite the lack of caution flags. The reasons could be traced to the decision not to stop during the 8th caution period of the race. Firstly, fresh tires and a full tank of fuel would only have given drivers enough to race for 40 laps, which would not have brought them to the end of the race. Secondly, drivers had probably expected one last caution before the race was over, and so they decided not to pit. Because the caution did not happen, several cars had to pit under regular racing conditions for fuel and some fresh tires.

We can draw a couple of conclusions from the observations over these 3 races with **low tire wear**:

1. Drivers at the front of the field will refuse to pit during the last stage of the race as long as their cars have the capability to run until the very end.
2. Drivers who pitted relatively recently had the flexibility to change two or four tires during a caution period.

The next step was to observe whether these conclusions applied to tracks with high tire wear. Based on our findings about the classification of tracks, races in Southeast_I and Southwest_B are considered good comparisons. In this first comparison, we relied on the 2014 Southeast_I race as shown in Figure 48. We studied the tire change strategy during the last stage of the race, and also looked at instances where pit stops were relatively close to each other to see if similar patterns were visible.

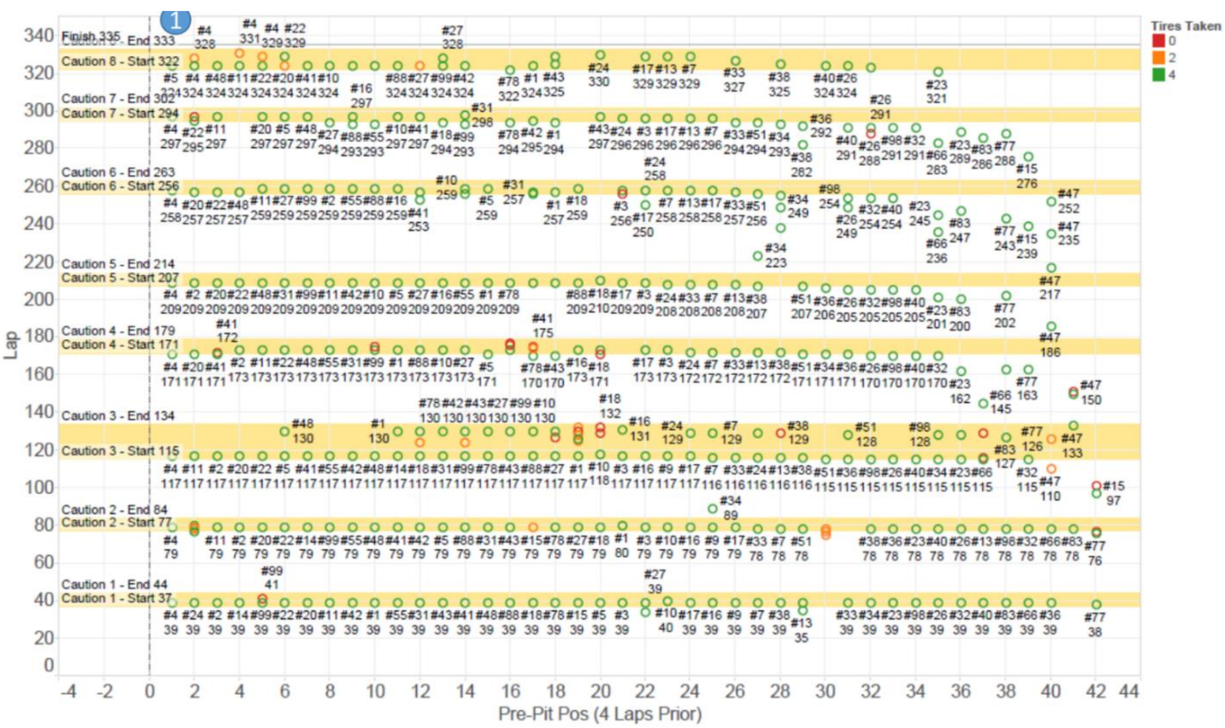


Figure 48. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southeast_I Race 2014

Tracks with very high tire wear usually had very few instances of 2-tire changes. Consistent with that observation, the Southeast_I race predominantly exhibited pit stops with 4-tire changes. Even when cautions occurred close to the end of the race, drivers at the front had to pit for fresh tires, as seen in the vicinity of the blue circle labelled “1”. In addition, even when drivers had pitted recently, they usually opted for 4-tires instead of 2, except when pit stops were made almost back-to-back.

It was important for us to verify these against another track with high tire-wear. As such, we chose the 2014 Southwest_B Race 1 as seen in Figure 49 to check whether the observations made for Southeast_I were consistent.

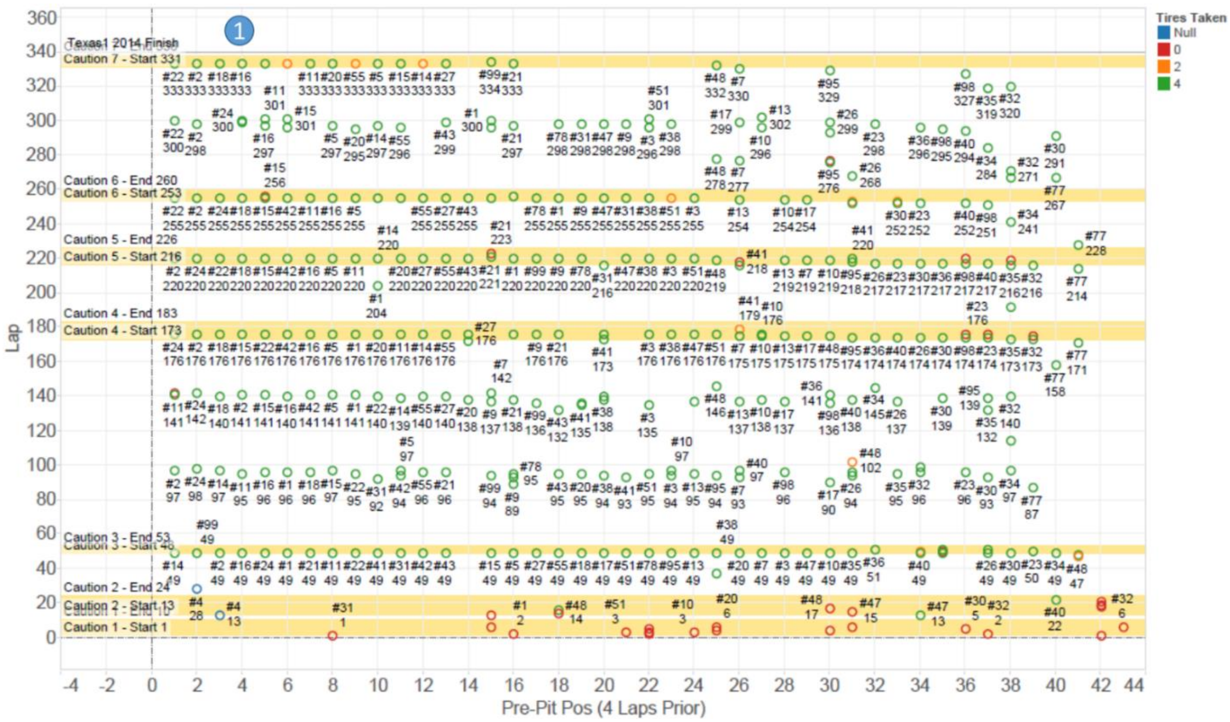


Figure 49. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southwest_B Race 1 2014

The pattern in the Southeast_I race repeated itself in Southwest_B. In the vicinity of the blue circle labelled "1", drivers decided to pit for fresh tires as late as during the caution flag on Lap 333, even though the race ended just 7 laps later. Pit stops also mostly involved 4-tire changes, except in the very early stages of the race, which was run under caution flag conditions.

We reached the following conclusions by studying these 2 races with **high tire wear**:

1. Drivers are very likely to pit during a caution regardless of when it occurs, even if that caution is close to the end of the race.
2. Drivers have very little flexibility to change 2-tires, and usually opt for 4-tires unless they have pitted extremely recently.

Collectively, our conclusions point us towards adopting different tire change strategies for races with high tire wear, compared to those with low tire wear.

5. Dataset Construction and Effects on Predictive Capability

5.1 Reconstructing the Dataset and Testing Methodology

In the original machine learning software, all races from 2012 and selected races from 2013 were aggregated to produce a consolidated dataset for training, testing, and validation. When running a machine learning model, it is critical that the dataset used is representative of the actual racing conditions, and that the features selected are useful in providing insights that will assist the machine learning algorithm to produce meaningful results.

For example, if a race was held in Southwest_B, we thought that it could be more useful to only use historical races from Southwest_B and similar tracks to build the dataset. This is because a number of race characteristics, such as road times for pit stops, the number of laps of a race, and other features that may be specific to the Southwest_B race may not have been comprehensively captured in the original dataset and accounted for in the selected features.

Rather than exhaustively account for those differences in track type and race characteristics by building additional features into the dataset, we thought about building a smaller dataset from scratch to include just a few races that we believed had very similar characteristics.

In order to understand whether our track classifications based on tire wear, tire change decisions, and the ratio of 2 versus 4 tire changes during pit stops had an impact on the predictive capability of the machine learning software, we needed to run a few tests against our hypotheses, which were:

1. Past low tire wear races with higher 2 versus 4 tire change ratios are good predictors for future low tire wear races with higher 2 versus 4 tire change ratios.
2. Past high tire wear races with lower 2 versus 4 tire change ratios are good predictors for future high tire wear races with lower 2 versus 4 tire change ratios.

We decided to run 5 types of tests by constructing datasets differently and tested the machine learning software against a selected race from within each dataset. The tests that we designed were:

1. **Test 1: Same Track**

   The race is tested against all races held on the same track only.

2. **Test 2: Similar Races**

   The race is tested against all races held on the same and similar tracks.

3. **Test 3: Different Races**

   The race is tested against all races held on different tracks.

4. **Test 4: All Races**

   The race is tested against all races.

5. **Test 5: Track Position Maintained**

   No dataset was built in this case. This test assumes that the track position of drivers at the start of each epoch remains the same at the end.

To compare outcomes, we calculated the root mean square error of the machine learning software to measure the differences in predictive capability across different test scenarios. This was calculated by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where n = number of predictions, y_i is the actual track position of the driver, an $\hat{y}_i$ is the track position predicted by the machine learning software.

We sought additional data from our industry sponsor, and obtained historical race data that was not previously available to us. This allowed us to build a more comprehensive dataset specific to the high and low tire wear tracks, and which also had different ratios of 2 versus 4 tire changes. The races were selected and classified in Figure 50.

Low Tear Wear, **Higher Ratio of 2 versus 4 Tire Changes**	**High Tear Wear,** **Lower Ratio of 2 versus 4 Tire Changes**
1. Midwest_C, Race 1, 2012	1. Southwest_B, Race 2, 2011
2. Midwest_C, Race 2, 2012	2. Southwest_B, Race 1, 2012
3. Midwest_C, Race 2, 2013	3. Southwest_B, Race 2, 2013
4. Midwest_C, Race 1, 2014	4. Southwest_B, Race 2, 2014
5. Midwest_C, Race 2, 2014	5. Southeast_I, 2012
6. Midwest_B, 2011	6. Southeast_I, 2013
7. Midwest_B, 2012	7. Southeast_I, 2014
8. Midwest_B, 2013	
9. Midwest_B, 2014	

Figure 50. Races selected to test effect of dataset construction on predictive capability

The races are arranged by the ratio of 2 versus 4 tire changes in increasing order as shown in Figure 51. In general, high tire wear tracks had lower ratios of 2 versus 4 tire changes, as was the case in the Southwest_B and Southeast_I races. With the exception of the Midwest_C Race 1 in 2012, all the low tire wear tracks had relatively higher ratios of 2 versus 4 tire changes. Nonetheless, we classified Midwest_C Race 1 2012 together with the Low Tire Wear, Higher Ratio of 2 versus 4 Tire Changes, as the track had an identical track length and shape compared to Midwest_B.

Race	2-Tires	4-Tires	2 versus 4 Ratio	Tire Wear	Track Length	Track Shape
Southwest_B, Race 1, 2012	5	257	0.019455	High	1.5 mi	Quad-Oval
Southeast_I, 2012	7	307	0.022801	High	1.54 mi	Quad-Oval
Southwest_B, Race 2, 2013	6	239	0.025105	High	1.5 mi	Quad-Oval
Southeast_I, 2013	15	350	0.042857	High	1.54 mi	Quad-Oval
Southwest_B, Race 2, 2014	15	341	0.043988	High	1.5 mi	Quad-Oval
Midwest_C, Race 1, 2012	11	216	0.050926	Low	1.5 mi	D-Shaped Oval
Southeast_I, 2014	19	327	0.058104	High	1.54 mi	Quad-Oval
Southwest_B, Race 2, 2011	23	277	0.083032	High	1.5 mi	Quad-Oval
Midwest_B, 2014	24	274	0.087591	Low	1.5 mi	D-Shaped Oval
Midwest_B, 2011	27	218	0.123853	Low	1.5 mi	D-Shaped Oval
Midwest_B, 2012	30	180	0.166667	Low	1.5 mi	D-Shaped Oval
Midwest_C, Race 2, 2014	41	217	0.18894	Low	1.5 mi	D-Shaped Oval
Midwest_C, Race 1, 2014	57	197	0.28934	Low	1.5 mi	D-Shaped Oval
Midwest_B, 2013	66	208	0.317308	Low	1.5 mi	D-Shaped Oval
Midwest_C, Race 2, 2013	94	176	0.534091	Low	1.5 mi	D-Shaped Oval
Midwest_C, Race 2, 2012	59	104	0.567308	Low	1.5 mi	D-Shaped Oval

Figure 51. Race and Track Statistics, and Details for Races Selected for Test Effect of Dataset Construction

Other factors, such as the length and shape of each track, were also considered when grouping tracks together. The Southwest_B and Southeast_I tracks had very similar track lengths of 1.5 and 1.54 miles respectively, and were of the quad-oval shape. The Midwest_C and Midwest_B tracks were both 1.5 miles each, and both were D-shaped oval tracks. In this regard, we built the dataset using tracks that were as common to each other as feasible, but had different tire wear and 2 versus 4 tire change characteristics so that we could isolate those characteristics and determine whether our hypotheses held or not.

5.2 Test Case: Low Tire Wear Track and Higher 2 versus 4 Tire Change Ratio

The results for the low tire wear tracks are recorded in Figure 52 with their relative order from "1" to "5" in brackets, with "1" having the lowest root mean square error.

	Midwest_C Race 2, 2014 (RMSE of Test)	Midwest_B, 2014 (RMSE of Test)
Test 1: Same Track	2.2242 (4)	2.8169 (4)
Test 2: Similar Races	2.2075 (3)	2.6746 (1)
Test 3: Different Races	2.192 (2)	2.7624 (3)
Test 4: All Races	2.1609 (1)	2.6967 (2)
Test 5: Track Position Maintained	3.422 (5)	3.8113 (5)

Figure 52. Root Mean Square Error of Machine Learning Software Test Cases for Midwest_C Race 2 2014 and Midwest_B Race 2014

Although our hypothesis stated that races on tracks with similar tire wear and 2 versus 4 tire change ratios were good predictors of future races with similar characteristics, it appeared that using different races could actually return better results, as seen when comparing the Midwest_C Race 2's Test 3 against Test 2 and Test 1. In fact, when the Midwest_C Race 2 was predicted using a dataset comprising different races (Test 4), the results were the best amongst all tests. This suggested that the effects of aggregation across a larger sample of races could outstrip differences in track characteristics for predictions.

For the Midwest_B 2014 race, the test against different races (Test 3) also performed better than historical races on the same track (Test 1). Although our hypothesis appeared to be correct in this case because the test for similar races (Test 2) performed the best among the lot, the results were just marginally different when compared to aggregating all races (Test 4). In addition, Test 4 produced better results than Test 1, which contradicted our hypothesis. Another finding was that in all cases, the machine learning software performed better than the naïve algorithm (Test 5), which returned a root mean square error that was about 1 more than the rest. This tallied with findings in earlier research done, which concluded that the machine learning software was better in predicting changes in track position compared to naïve predictions.

Lastly, aggregating races across tracks of different tire wear and tire change ratios (Test 4) tended to perform better than simply looking at past races on the same track (Test 1). This suggested that it may be pointless to focus on selecting races for the construction of the dataset, as the predictive capability of the software would not be adversely affected by including races of different track characteristics. However, our caveat is that throughout our dataset, only oval tracks with very similar track lengths were used.

5.3 Test Case: High Tire Wear Track and Lower 2 versus 4 Tire Change Ratio

The results for the high tire wear tracks are recorded in Figure 53 with their relative order from “1” to “5” in brackets, with “1” having the lowest root mean square error.

	Southwest_B, Race 2, 2014 (RMSE of Test)	Southeast_I, 2014 (RMSE of Test)
Test 1: Same Track	2.5007 (4)	2.6129 (4)
Test 2: Similar Races	2.192 (2)	2.5708 (1)
Test 3: Different Races	2.2075 (3)	2.5924 (2)
Test 4: All Races	2.1609 (1)	2.5983 (3)
Test 5: Track Position Maintained	3.348 (5)	3.5999 (5)

Figure 53. Root Mean Square Error of Machine Learning Software Test Cases for Southwest_B Race 2 2014 and Southeast_I Race 2014

The test which returned the best result for the Southwest_B Race 2 was the one that aggregated all races (Test4), followed by that of similar races (Test 2), different races (Test 3), and the same track (Test 1). Similar to the result with the low tire wear tracks, it appeared that aggregating races regardless of their tire wear and tire change ratio characteristics did not have a detrimental effect on the predictive capability of the machine learning software. Instead, it performed the best among the lot.

The results for Southeast_I showed that making predictions against similar races (Test 2) performed the best, followed by different races (Test 3), all races (Test 4), and the same track (Test 1). Even though our hypothesis turned out to be true in this scenario, the difference in the root mean square error for the tests was razor thin, at around 0.04 from the first to the fourth test. This indicated that the way the dataset was constructed was not important in determining the predictive capability of the software for this race.

As with the test on the low tire wear track, Test 5 turned in the poorest result, showing once more that the machine learning software performed better than the naïve algorithm.

In conclusion, we believe that our hypotheses regarding the importance of tire wear and tire change ratios for the machine learning software does not hold. There appears to be no discernable benefit from constructing the dataset according to tire wear and tire change ratio characteristics, when using the current set of features in the machine learning software, and selecting tracks that have similar lengths and shapes.

6. Overall Findings and Relation to Previous Research

We looked at the research done in this thesis and compared it to insights discovered from the paper (Tulabandhula & Rudin, 2014) to understand whether our results tallied and see if there were additional insights that could be gleaned for further work.

6.1 New Insights and Reinforcement of Previous Insights

1. Strong Correlation between Tire Wear and Tire Change Decisions

There is a strong correlation between the tire wear characteristics of a track and tire change decisions. Tracks with low tire wear have higher ratios of 2 versus 4 tire change decisions, while those with high tire wear have much lower ratios of 2 versus 4 tire change decisions. This finding could be incorporated into the machine learning model to offer better prescriptions of tire change decisions for each pit stop.

2. Correlation between Finishing Position and Combination of Driver Momentum and Pit Crew Performance

Building on features implemented in the machine learning software, we found that there was some correlation between a driver's finishing position and the combination of driver momentum and pit crew performance. The combination of driver momentum and pit crew performance was useful in determining a driver's performance across a stage of the race, which comprised an epoch, warm-up laps, and caution periods.

3. Aggregation of Races for Dataset

It should be possible to aggregate races of different track characteristics for the construction of the dataset for the machine learning software even if tire change ratios and tire wear are fundamentally different. However, it is important to note that the high and low tire wear tracks selected for this test were oval tracks of similar lengths. We did not test the software against tracks of different lengths or those that were road courses, which might return different results.

4. Importance of Later Stages of a Race

Our studies indicate that tire change decisions at the later stages of the race take on relatively greater importance because poor performance at the beginning can be made up through good driver and pit crew performance as the race progresses. The prevalence of caution periods appears to be the main reason for this, because cars can catch up with the rest of the pack and recover lost laps during such periods. Poor decisions made near the end have no way of being corrected unless the performance of the driver in that last stage is exceptional.

6.2 Responses to Primary Research Questions

Returning to the original questions posed in this thesis, our responses based on the research conducted are:

1. If a driver pits and makes a certain tire change decision, what is the predicted change in track position before the next pit stop?

The machine learning software has generally been able to predict the change of track position of a driver with an accuracy better than the naïve approach, where the initial track position of the driver is retained throughout the duration of the epoch. New features based on our improved understanding of how end stages of a race influence outcomes, and the correlation of tire wear to tire change decisions may enhance the software's predictive quality for the Prediction Interval.

2. When should a driver pit during the course of a race?

Ideally, drivers should pit just before cautions are declared during a race because they would end up at the front of the field when others pit during the caution. However, because predicting caution periods is difficult, we suggest that pit stops should be done whenever caution flags are waved, unless we are near the end of a race on a low tire wear track.

3. How many tires should be changed during that pit stop?

4 tires should be changed at every pit stop, particularly on high tire wear tracks. For tracks which have low tire wear characteristics, 2-tire pit stops or 0-tire stops are recommended when tire changes have been done very recently, and the tires still exhibit very good performance based on the driver's lap time.

4. What are the likely pit stop decisions made by the driver's rivals?

We have not yet predicted the pit stop decisions made by a driver's rivals, but there are a couple of ways that this can be done. The first is to study what the optimal tire change strategy would be by looking at the relative age of the rival's tires and the tire wear characteristic of the track. The second is to look at the decisions of drivers ahead of the rival. This characteristic of protected drivers (Tulabandhula & Rudin, 2014) could be a good indicator on how to decide what rivals would do.

7. Future Work

7.1 Adding New Features into the Machine Learning Software

Following our additional findings in this thesis, new features could be implemented in the machine learning software to determine if these insights would result in better predictive outcomes:

1. **Combination of Driver Momentum and Pit Crew Performance**

   This feature could be useful in providing higher-level insights into performance for each stage of the race to help inform the tire-change decision-making process when making pit stops.

2. **Percentage of Race Complete**

   The stage of the race is presently contained in the feature matrix by recording the number of laps completed and the leg number for each driver. In order to tie the stage of the race to specific tire change decisions, it may be useful to capture the percentage of the race completed by dividing the lap number of the pit stop by the total number of laps for the race.

3. **Inclusion of More Features Surrounding Pit Stops**

   In order to improve the predictions made for the prediction interval, more features surrounding the nature of pit stops could be included to provide further insight into portions of race stages that are not captured in the feature matrix. For instance, the change of track position in between epochs, gap timings of drivers, tire age in relation to track tire wear profile, and pit times as a fraction of lap times could be calculated and tested.

7.2 Building Track Profiles from Practice and Qualifying Sessions

Previous research suggested classifying tracks into different groups based on practice and qualifying times prior to the race proper (Tulabandhula & Rudin, 2014). Our independent discussions with engineers involved in Formula One revealed that this was commonly done because track characteristics tended to change from year to year, and the most accurate data was captured by studying track conditions during the race weekend. Although computer simulations done independently in Formula One tended to yield fairly accurate results beforehand, fine-tuning of car set-ups and predictive analytics for the race weekend were normally done based on driver and car performance from the later practice sessions. Feeding on-site track characteristics to the machine learning software during the race weekend could be useful in future.

7.3 Heuristics

Based on the analysis performed in the thesis, there are clear patterns for tire change decisions for races on different tracks. Appropriate weights could be introduced into the machine learning software to skew tire change decisions using the following rules:

Tracks with High Tire Wear and Low 2 versus 4 Tire Change Ratios

Applies to the following tracks: Southwest_B, Southeast_F, West_C, Southeast_I, Southeast_G, Southeast_H, and West_B

1. Pit whenever there is a caution, and change 4 tires whenever possible. The savings of about 6-7 seconds for a 2-tire stop is worth less than the better lap times clocked when on 4 fresh tires for the next leg of the race.

Tracks with Low Tire Wear and High 2 versus 4 Tire Change Ratios

Applies to the following tracks: Southeast_B, Northeast_C, Southwest_A, Midwest_A, Southeast_A, Northeast_B, and Midwest_C

1. If it is early in the race and a caution flag is waved, a driver may change 2-tires instead of 4 provided that the car's tires are relatively fresh and are expected to last until the next caution period. The savings of about 6-7 seconds is usually worth more than the slower lap times clocked for the next leg of the race compared to having 4 fresh tires.

2. If it is late in the race and drivers are within the top 10, do not pit unless absolutely necessary, such as when the tires are completely worn out, or if fuel has run out. Do not pit even if there is a caution, as maintaining position on track is more important than putting on fresh tires.

All Other Tracks

Applies to the following tracks: Northeast_A, Southeast_C, Midwest_E, West_A, Northeast_D, Southeast_E, Midwest_B, Southeast_D, Midwest_D

1. No clear pattern at this point.

7.4 Statistical Probability of Cautions during a Race

Drivers face a variety of hazards throughout the duration of a race, especially since up to 43 cars lap the track at high speeds, and regularly adjust their racing line to take advantage of changing track conditions and rivals ahead of, or behind them. This continual adjustment requires drivers to maintain high levels of concentration to avoid getting involved in incidents on the track. Accidents are prone to happen when drivers experience a momentary lapse of concentration.

Some tracks may present higher levels of difficulty to drivers, and as a result accidents may happen more frequently. These incidents will cause caution flags to be waved to slow down the race, and bring it under control such that race officials are able to clear hazards off the track. During these periods, drivers usually take the opportunity to pit for fresh tires and refuel.

Hence, it may be useful for us to calculate the probability of cautions during different stages of each race. Given that the incidence of caution flags are inherently probabilistic, it may be useful to rely on as many past races as possible to generate predictions for each track. Doing so would give NASCAR teams a better understanding of whether they should pit their drivers under regular racing conditions, or whether they should stay out on track to wait for a caution to occur. This ties in with one of our main research questions about predicting caution periods in advance, in order to call drivers into the pit immediately before caution flags are waved.

7.5 Adjustment of Epoch Characteristics

Finally, the machine learning software currently trims laps from the start and the end of each leg of the race to generate epochs. As a result, epochs are not included into the feature matrix when they are too short, such as when caution periods occur in rapid succession of each other. This may be detrimental to the predictive capability of the machine learning software, such as if cautions occur near the end of the race, when tire change decisions take on increasing importance. It may be useful to retain epochs regardless of their lengths near the end of the race or adjust the minimum allowable epoch length for each track.

8. References

Bi, F. (2014, Nov 13). *How Formula One Teams Are Using Big Data To Get The Inside Edge*. Retrieved from Forbes: http://www.forbes.com/sites/frankbi/2014/11/13/how-formula-one-teams-are-using-big-data-to-get-the-inside-edge/

Gartner. (2014). Major Myths About Big Data's Impact on Analytics. *Gartner*.

Hartigan, M. (2014, May 23). *Why This NASCAR Team Is Putting RFID Sensors On Every Person In The Pit*. Retrieved from Fast Company: http://www.fastcolabs.com/3031050/why-nascar-is-putting-rfid-sensors-on-every-person-in-the-pit

Jones, T. M. (2013, September 12). *Data Science and Open Source*. Retrieved from IBM developerWorks: http://www.ibm.com/developerworks/jp/opensource/library/os-datascience/

MeKenna, B. (2013, May 22). *McLaren uses high-speed data analytics to gain Formula 1 edge*. Retrieved from ComputerWeekly.com: http://www.computerweekly.com/news/2240184539/McLaren-uses-high-speed-data-analytics-to-gain-Formula-1-edge

Spencer, L. (2014, August 5). *Could tire sensors stop air pressure shenanigans?* Retrieved from Motorsport.com: http://www.motorsport.com/nascar-cup/news/could-tire-sensors-stop-air-pressure-shenanigans/

Stevens, T. (2013, Mar 13). *The racing line: Exploring NASCAR's technological dichotomy*. Retrieved from Engadget: http://www.engadget.com/2013/03/13/exploring-nascars-technological-dichotomy/

Tulabandhula, T., & Rudin, C. (2014). Tire Changes, Fresh Air, and Yellow Flags: Challenges in Predictive Analytics for Professional Racing. *Big Data*.

9. Appendix

9.1 Selected Races for the 2014 Dataset

The races for 2014 in alphabetical order, according to region and race number are:

Race Name (Abbreviated)
Midwest_A Race 1
Midwest_A Race 2
Midwest_B
Midwest_C Race 1
Midwest_C Race 2
Midwest_D
Midwest_E
Northeast_A
Northeast_B Race 1
Northeast_B Race 2
Northeast_C Race 1
Northeast_C Race 2
Northeast_D Race 1
Northeast_D Race 2
Southeast_A Race 1
Southeast_A Race 2
Southeast_B Race 1
Southeast_B Race 2
Southeast_C Race 1
Southeast_C Race 2
Southeast_D Race 1
Southeast_D Race 2
Southeast_E Race 1
Southeast_E Race 2
Southeast_F Race 1
Southeast_F Race 2
Southeast_G
Southeast_H
Southeast_I
Southwest_A Race 1

Southwest_A Race 2
Southwest_B Race 1
Southwest_B Race 2
West_A
West_B
West_C

Figure 54. List of Races for 2014 NASCAR Cup Series

9.2 Complete Feature Listing

ID.	Classification	Description	Dep. Var	Ind. Var
33	current outing	delrank (leg end - pre pit)'	X	
9	before pit	Feature ID 9		X
13	before pit	Feature ID 13		X
26	before pit	Feature ID 26		X
28	before pit	Feature ID 28		X
34	before pit	Feature ID 34		X
35	before pit	Feature ID 35		X
37	before pit	Feature ID 37		X
40	before pit	Feature ID 40		X
41	before pit	Feature ID 41		X
43	before pit	Feature ID 43		X
44	before pit	Feature ID 44		X
45	before pit	Feature ID 45		X
56	before pit	Feature ID 56		X
59	before pit	Feature ID 59		X
60	before pit	Feature ID 60		X
61	before pit	Feature ID 61		X
69	before pit	Feature ID 69		X
75	before pit	Feature ID 75		X
76	after pit	Feature ID 76		X
77	after pit	Feature ID 77		X
82	before pit	Feature ID 82		X
85	after pit	Feature ID 85		X
89	before pit	Feature ID 89		X
127	before pit	Feature ID 127		X
132	current outing	Feature ID 132		X
136	current outing	Feature ID 136		X
139	after pit	Feature ID 139		X
140	after pit	Feature ID 140		X
146	before pit	Feature ID 146		X
149	before pit	Feature ID 149		X
150	before pit	Feature ID 150		X
151	before pit	Feature ID 151		X
153	before pit	Feature ID 153		X
157	before pit	Feature ID 157		X
158	before pit	Feature ID 158		X
160	before pit	Feature ID 160		X
161	before pit	Feature ID 161		X

163	before pit	Feature ID 163		X
164	before pit	Feature ID 164		X
165	before pit	Feature ID 165		X
1	nonfeature	Feature ID 1		
2	nonfeature	Feature ID 2		
3	nonfeature	Feature ID 3		
4	current outing	Feature ID 4		
5	current outing	Feature ID 5		
6	current outing	Feature ID 6		
7	current outing	Feature ID 7		
8	current outing	Feature ID 8		
10	before pit	Feature ID 10		
11	before pit	Feature ID 11		
12	before pit	Feature ID 12		
14	before pit	Feature ID 14		
15	current outing	Feature ID 15		
16	current outing	Feature ID 16		
17	nonfeature	Feature ID 17		
18	nonfeature	Feature ID 18		
19	after pit	Feature ID 19		
20	after pit	Feature ID 20		
21	after pit	Feature ID 21		
22	current outing	Feature ID 22		
23	current outing	Feature ID 23		
24	nonfeature	Feature ID 24		
25	nonfeature	Feature ID 25		
27	after pit	Feature ID 27		
29	nonfeature	Feature ID 29		
30	after pit	Feature ID 30		
31	current outing	Feature ID 31		
32	current outing	Feature ID 32		
36	before pit	Feature ID 36		
38	before pit	Feature ID 38		
39	before pit	Feature ID 39		
42	before pit	Feature ID 42		
46	before pit	Feature ID 46		
47	before pit	Feature ID 47		
48	after pit	Feature ID 48		
49	after pit	Feature ID 49		
50	after pit	Feature ID 50		
51	after pit	Feature ID 51		

52	after pit	Feature ID 52		
53	after pit	Feature ID 53		
54	after pit	Feature ID 54		
55	after pit	Feature ID 55		
57	before pit	Feature ID 57		
58	before pit	Feature ID 58		
62	current outing	Feature ID 62		
63	current outing	Feature ID 63		
64	nonfeature	Feature ID 64		
65	after pit	Feature ID 65		
66	after pit	Feature ID 66		
67	current outing	Feature ID 67		
68	current outing	Feature ID 68		
70	current outing	Feature ID 70		
71	before pit	Feature ID 71		
72	before pit	Feature ID 72		
73	current outing	Feature ID 73		
74	after pit	Feature ID 74		
78	after pit	Feature ID 78		
79	after pit	Feature ID 79		
80	after pit	Feature ID 80		
81	after pit	Feature ID 81		
83	before pit	Feature ID 83		
84	after pit	Feature ID 84		
86	before pit	Feature ID 86		
87	before pit	Feature ID 87		
88	before pit	Feature ID 88		
90	after pit	Feature ID 90		
91	current outing	Feature ID 91		
92	current outing	Feature ID 92		
93	current outing	Feature ID 93		
94	after pit	Feature ID 94		
95	current outing	Feature ID 95		
96	current outing	Feature ID 96		
97	current outing	Feature ID 97		
98	before pit	Feature ID 98		
99	nonfeature	Feature ID 99		
100	before pit	Feature ID 100		
101	before pit	Feature ID 101		
102	current outing	Feature ID 102		
103	current outing	Feature ID 103		

104	current outing	Feature ID 104		
105	current outing	Feature ID 105		
106	current outing	Feature ID 106		
107	current outing	Feature ID 107		
108	current outing	Feature ID 108		
109	current outing	Feature ID 109		
110	current outing	Feature ID 110		
111	current outing	Feature ID 111		
112	current outing	Feature ID 112		
113	current outing	Feature ID 113		
114	current outing	Feature ID 114		
115	current outing	Feature ID 115		
116	current outing	Feature ID 116		
117	current outing	Feature ID 117		
118	current outing	Feature ID 118		
119	current outing	Feature ID 119		
120	current outing	Feature ID 120		
121	current outing	Feature ID 121		
122	current outing	Feature ID 122		
123	current outing	Feature ID 123		
124	current outing	Feature ID 124		
125	current outing	Feature ID 125		
126	current outing	Feature ID 126		
128	before pit	Feature ID 128		
129	before pit	Feature ID 129		
130	before pit	Feature ID 130		
131	after pit	Feature ID 131		
133	before pit	Feature ID 133		
134	before pit	Feature ID 134		
135	current outing	Feature ID 135		
137	before pit	Feature ID 137		
138	before pit	Feature ID 138		
141	before pit	Feature ID 141		
142	before pit	Feature ID 142		
143	before pit	Feature ID 143		
144	before pit	Feature ID 144		
145	before pit	Feature ID 145		
147	before pit	Feature ID 147		
148	before pit	Feature ID 148		
152	before pit	Feature ID 152		
154	before pit	Feature ID 154		

155	after pit	Feature ID 155		
156	after pit	Feature ID 156		
159	before pit	Feature ID 159		
162	before pit	Feature ID 162		

Figure 55. Complete Feature Listing, including Dependent and Independent Variables in Machine Learning Software

9.3 Slopes by Tire Wear for Track Classification

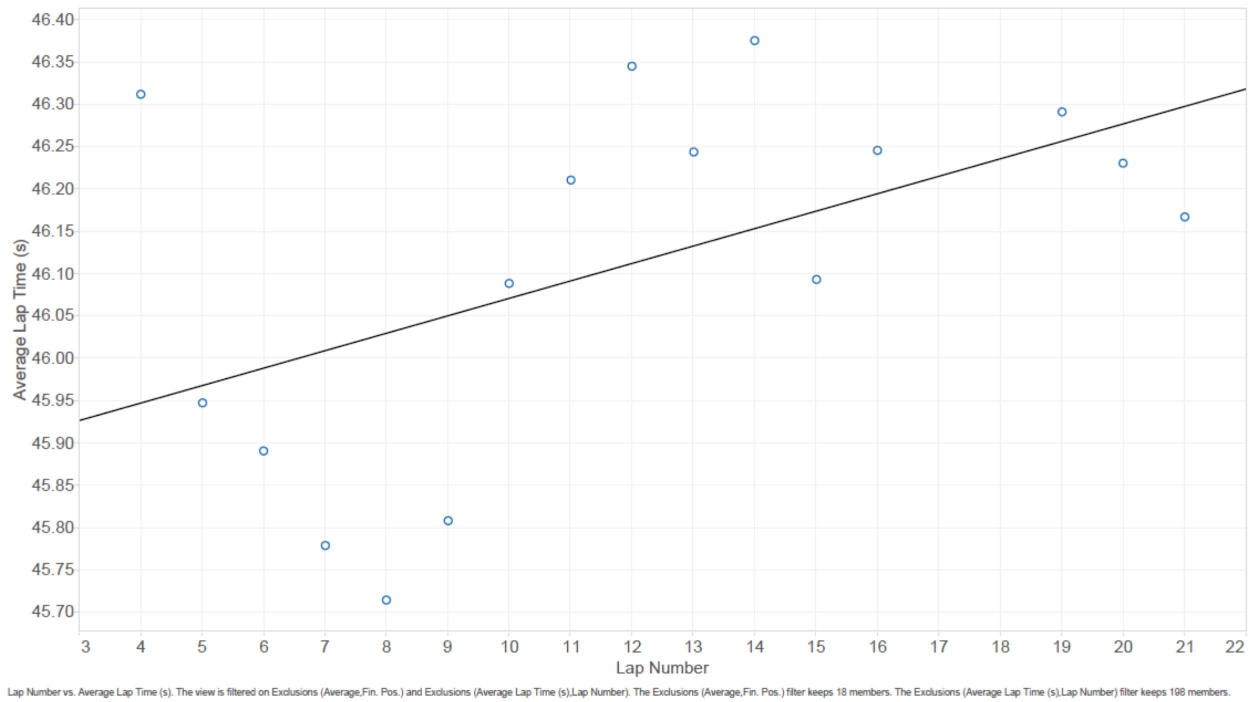


Figure 56. Slope of Tire Degradation for Southeast_B Track based on Southeast_B Race 2014

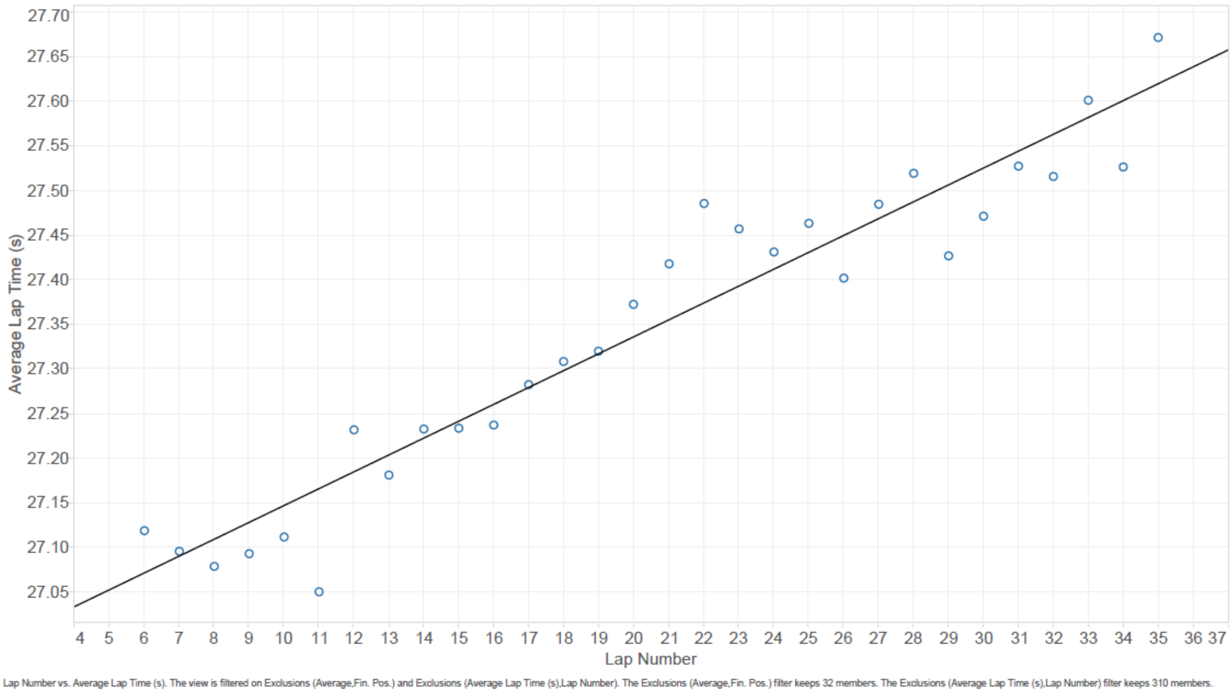


Figure 57. Slope of Tire Degradation for Southwest_A Track based on Southwest_A Race 1 2014

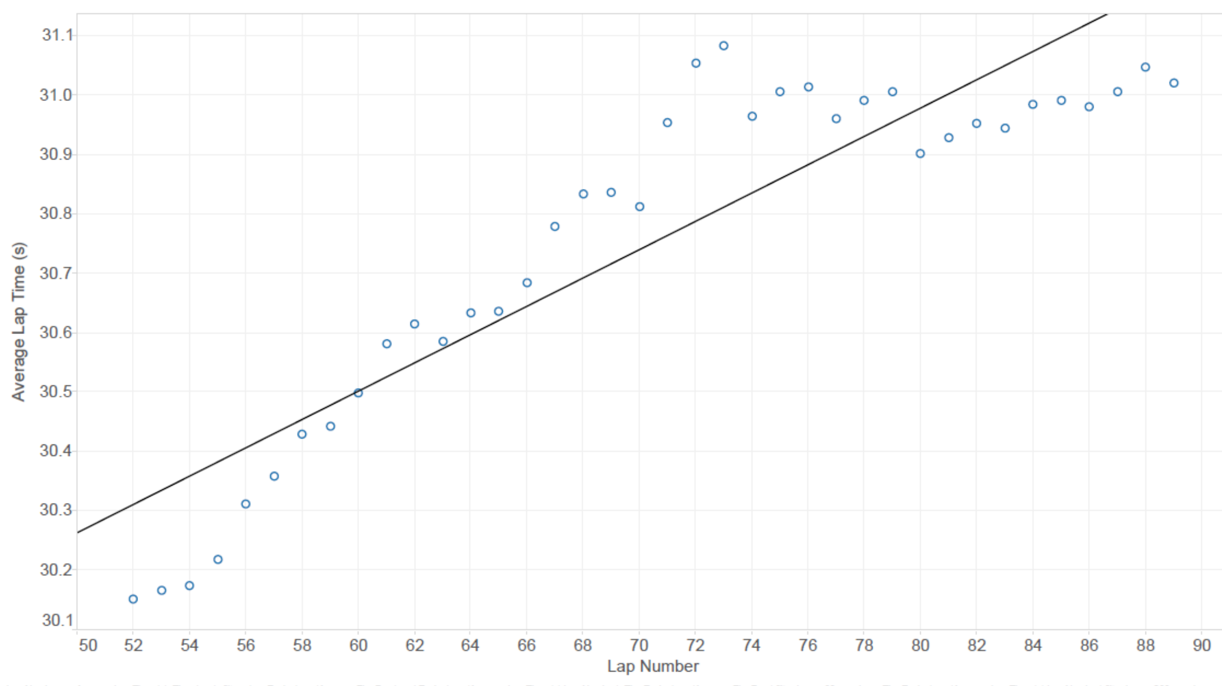


Figure 58. Slope of Tire Degradation for West_A Track based on West_A Race 2014

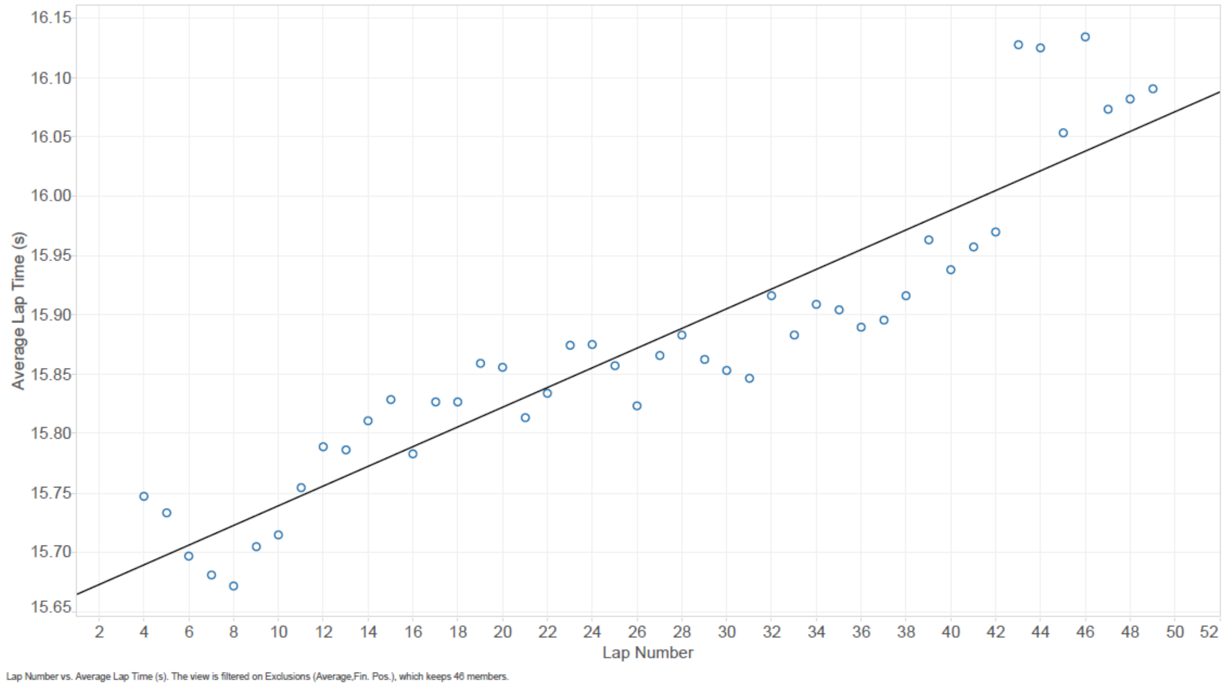


Figure 59. Slope of Tire Degradation for Southeast_C Track based on Southeast_C Race 1 2014

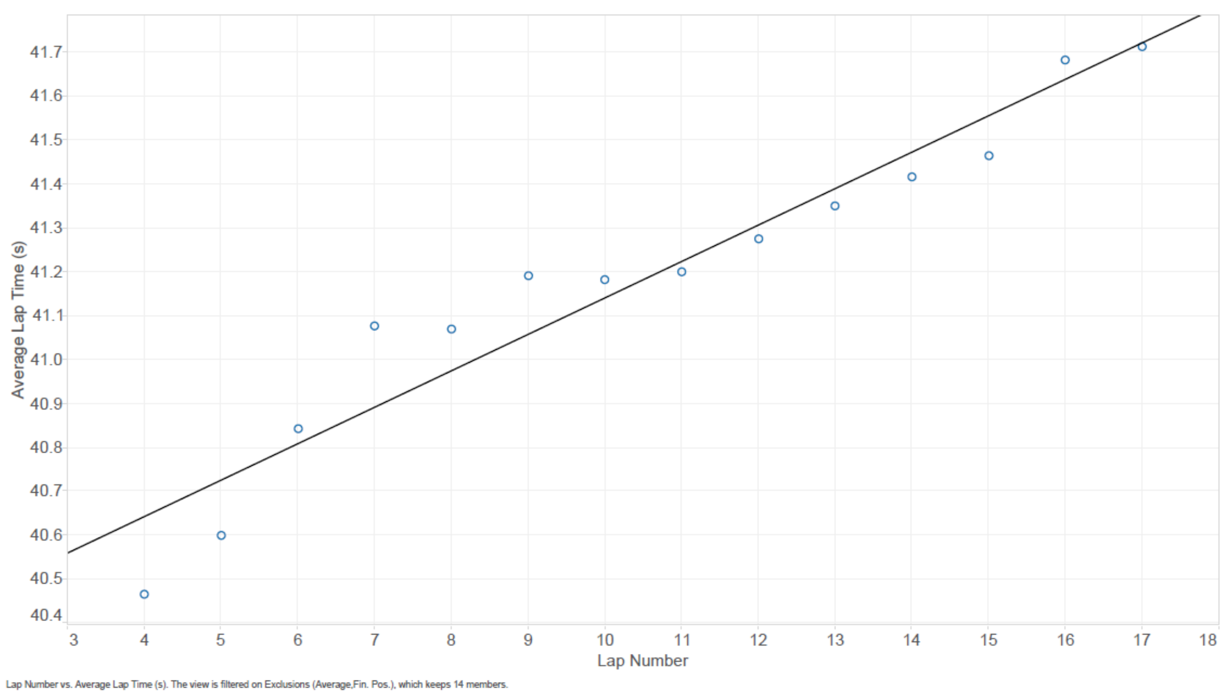
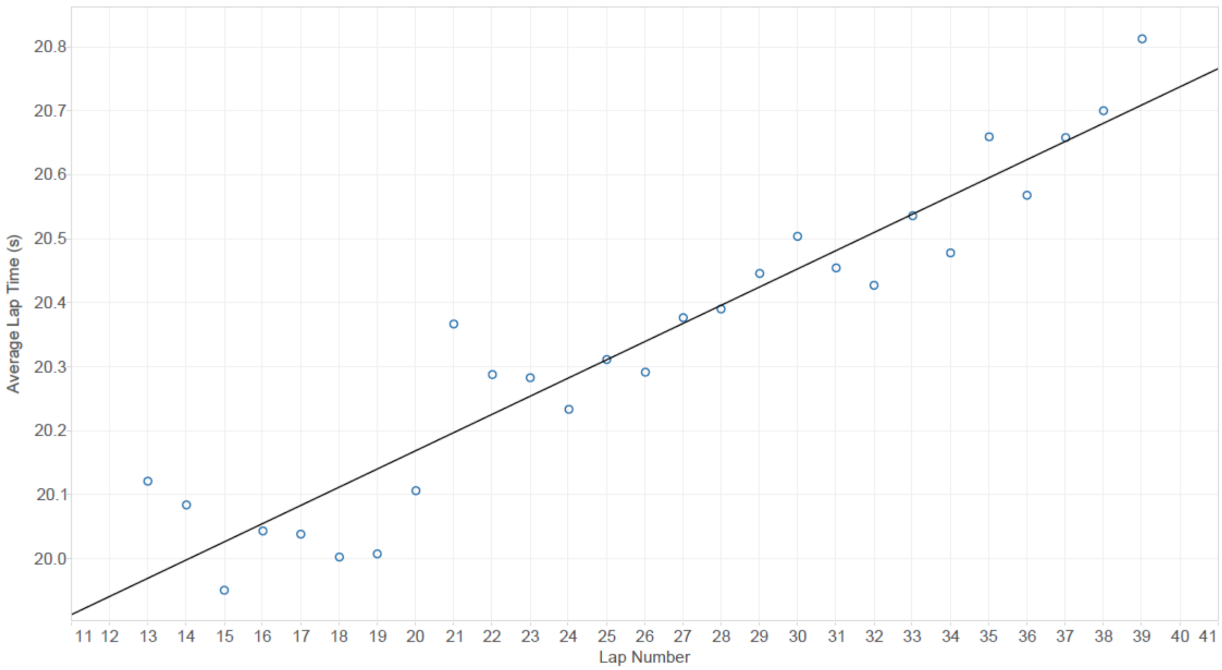
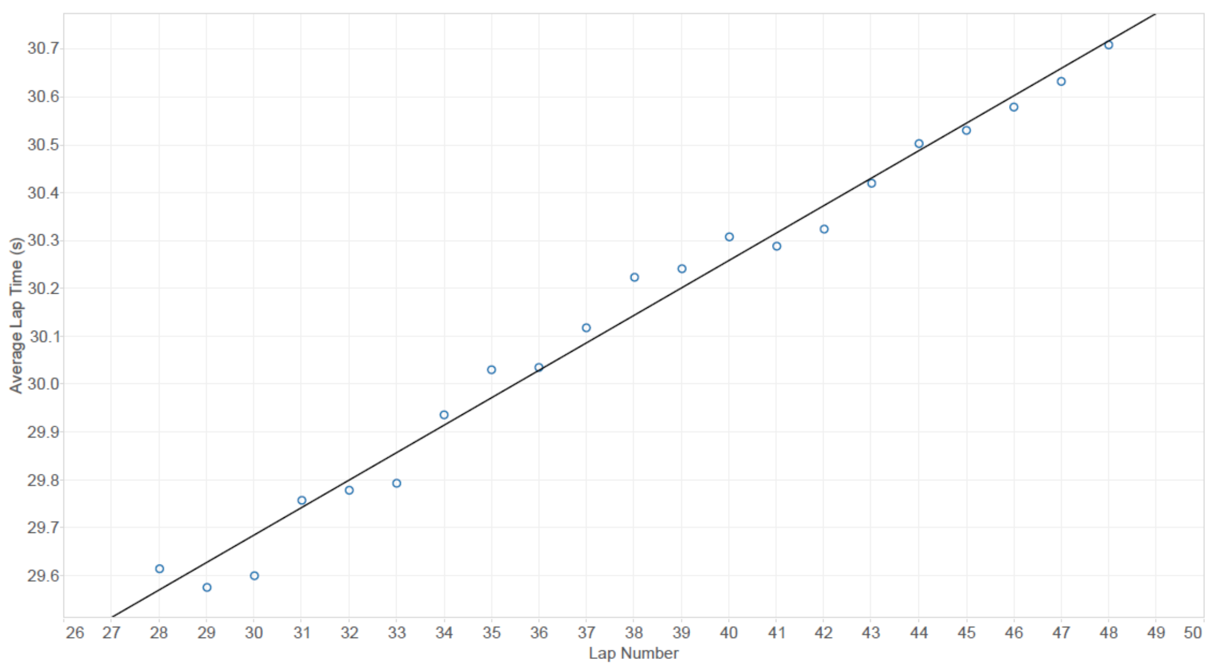


Figure 60. Slope of Tire Degradation for West_C Track based on West_C Race 2014



Lap Number vs. Average Lap Time (s). The view is filtered on Exclusions (Average,Fin. Pos.) and Exclusions (Average Lap Time (s),Lap Number). The Exclusions (Average,Fin. Pos.) filter keeps 30 members. The Exclusions (Average Lap Time (s),Lap Number) filter keeps 497 members.

Figure 61. Slope of Tire Degradation for Southeast_E Track based on Southeast_E Race 1 2014



Lap Number vs. Average Lap Time (s). The view is filtered on Exclusions (Average,Fin. Pos.) and Exclusions (Average Lap Time (s),Lap Number). The Exclusions (Average,Fin. Pos.) filter keeps 24 members. The Exclusions (Average Lap Time (s),Lap Number) filter keeps 337 members.

Figure 62. Slope of Tire Degradation for Southwest_B Track based on Southwest_B Race 1 2014

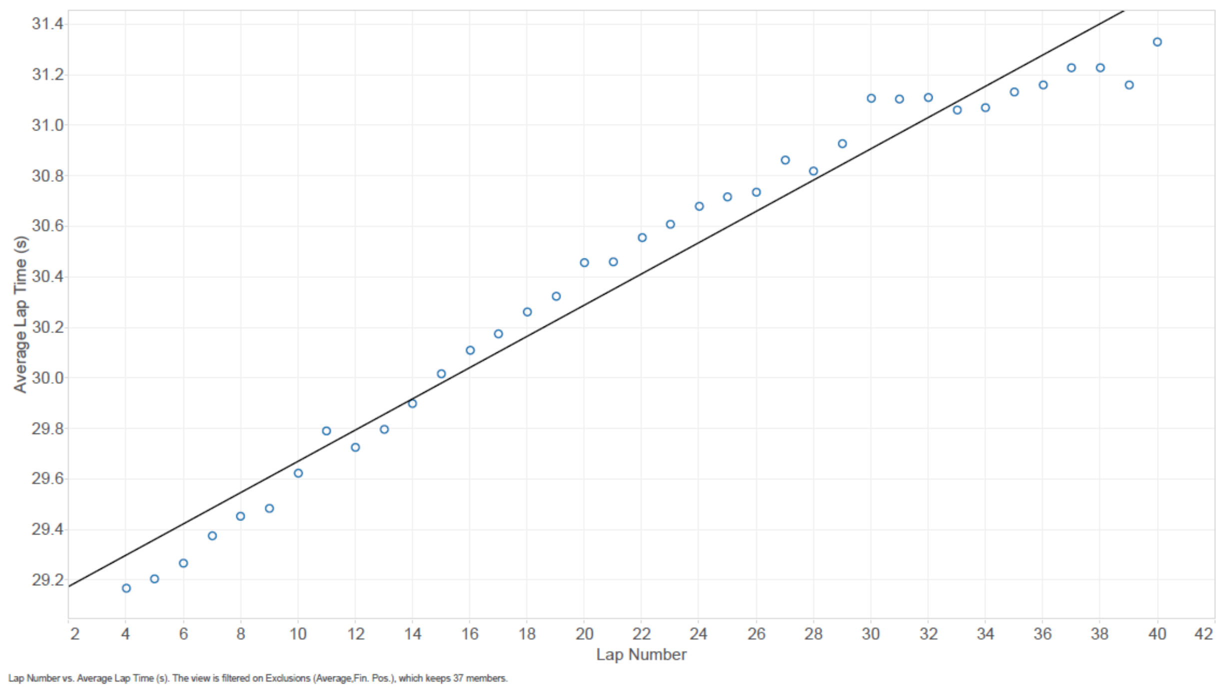


Figure 63. Slope of Tire Degradation for Southeast_H Track based on Southeast_H Race 2014

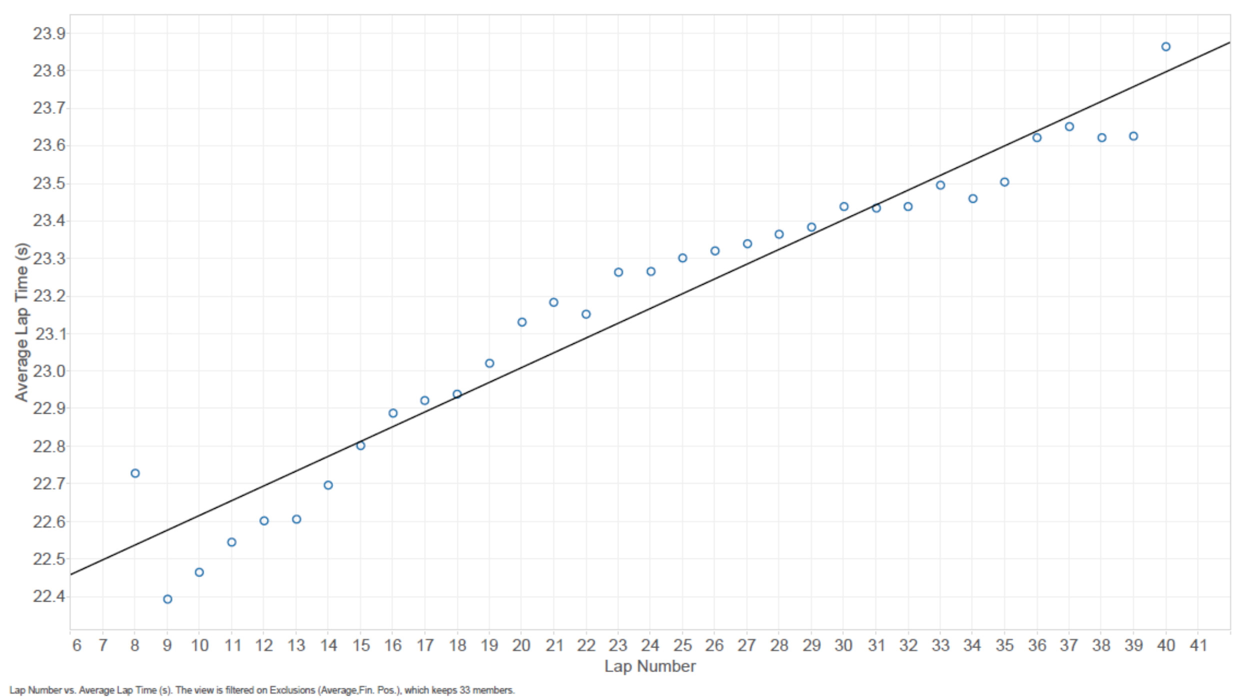


Figure 64. Slope of Tire Degradation for Southeast_F Track based on Southeast_F Race 1 2014

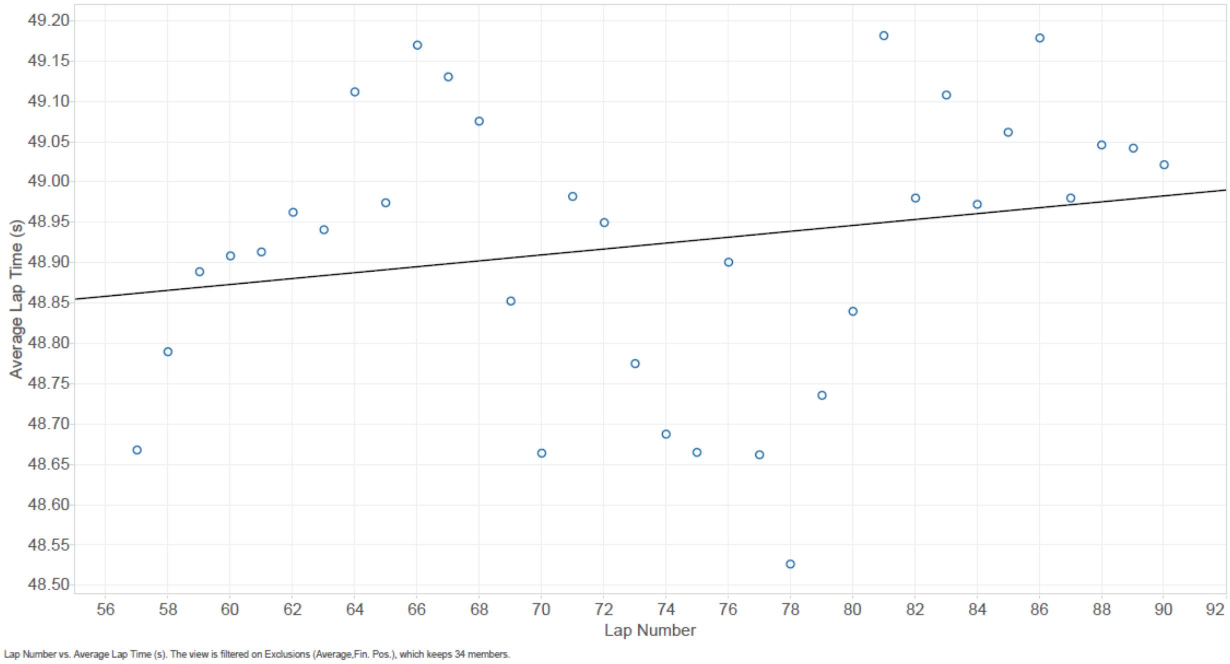


Figure 65. Slope of Tire Degradation for Southeast_A Track based on Southeast_A Race 1 2014

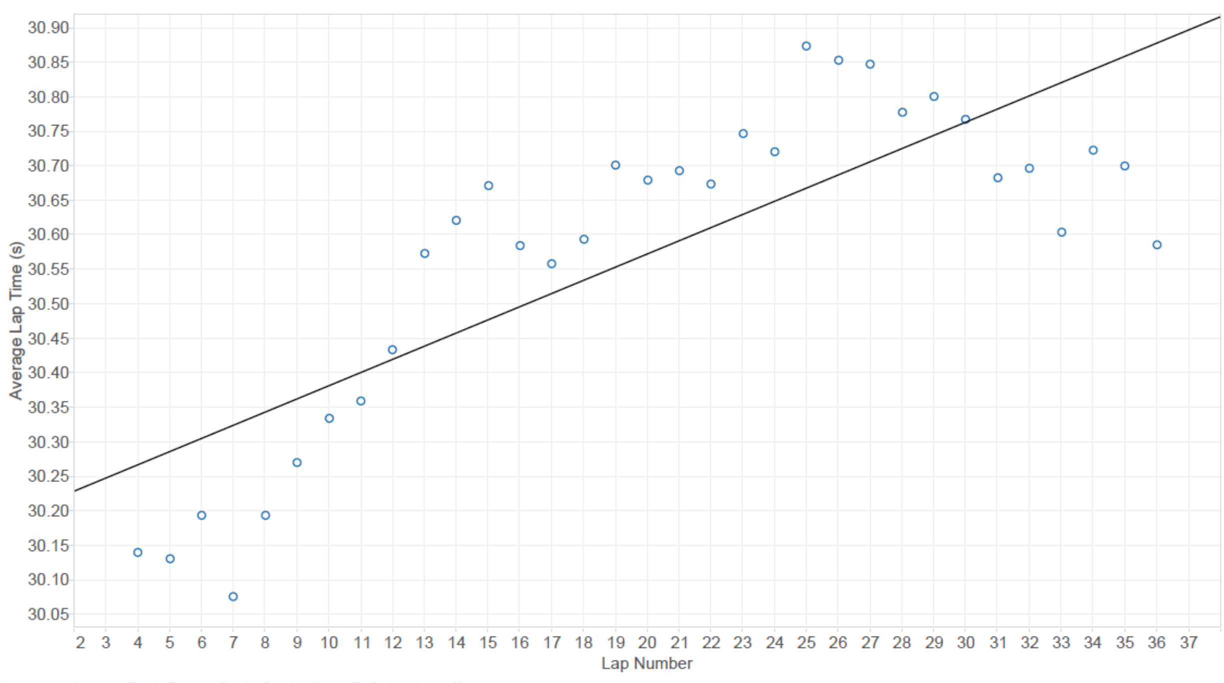


Figure 66. Slope of Tire Degradation for Midwest_C Track based on Midwest_C Race 1 2014

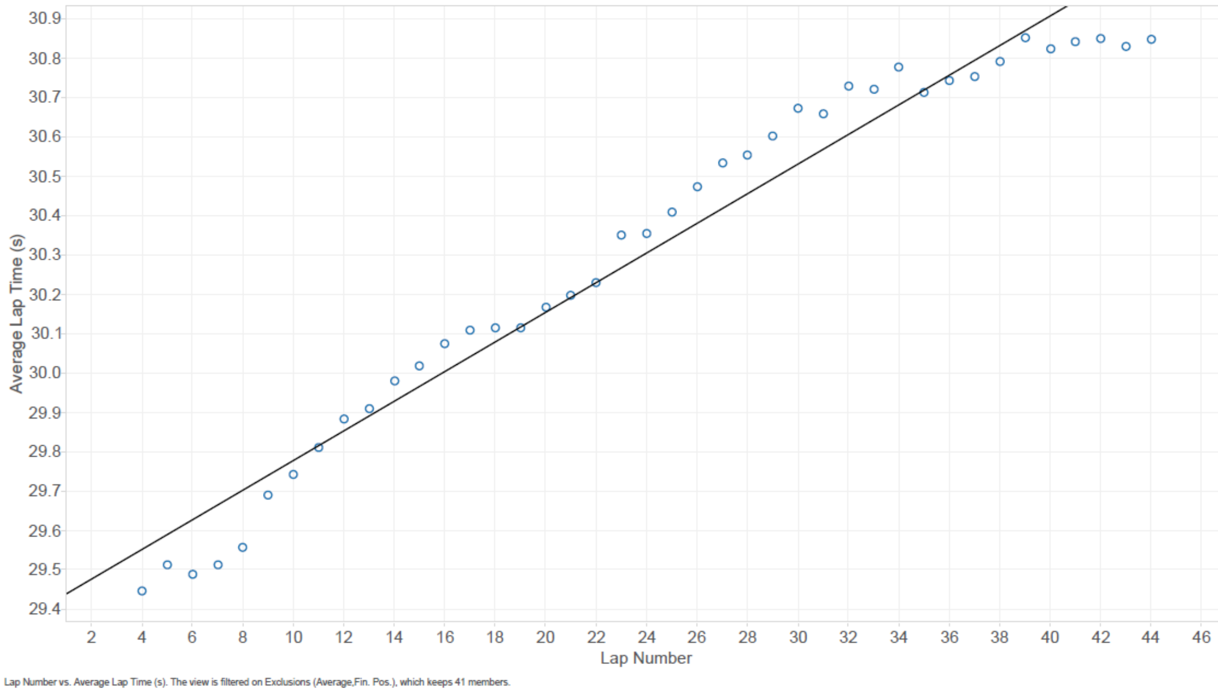


Figure 67. Slope of Tire Degradation for Southeast_D Track based on Southeast_D Race 1 2014

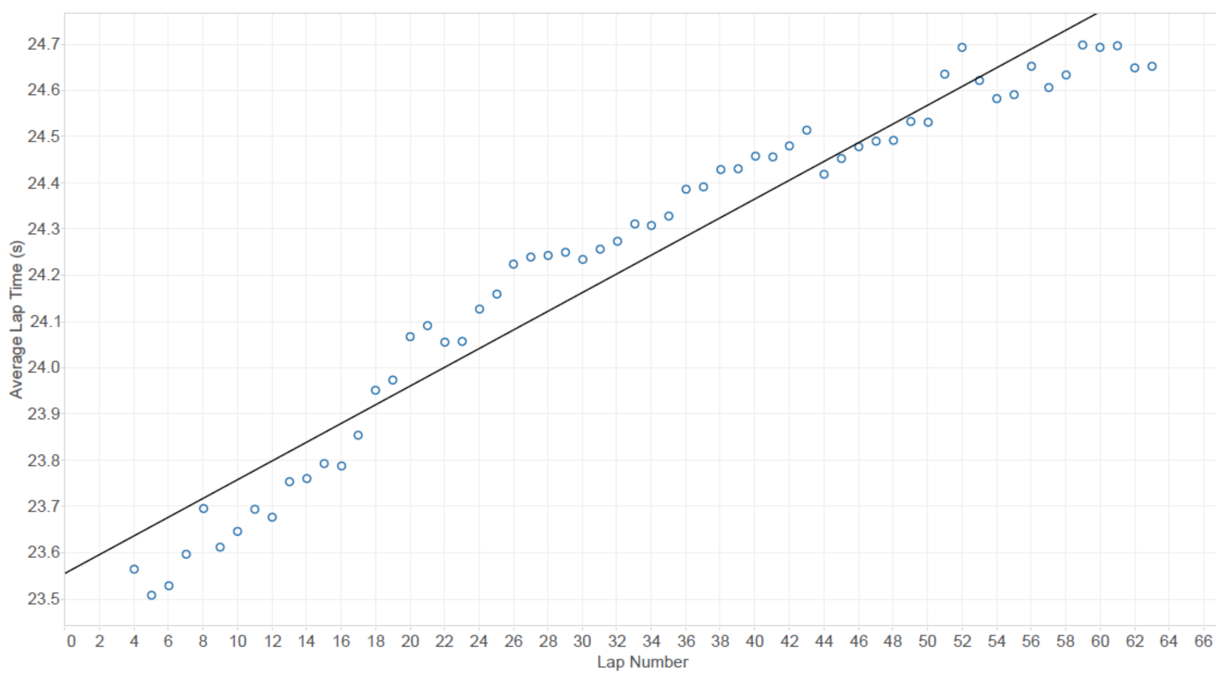


Figure 68. Slope of Tire Degradation for Northeast_D Track based on Northeast_D Race 1 2014

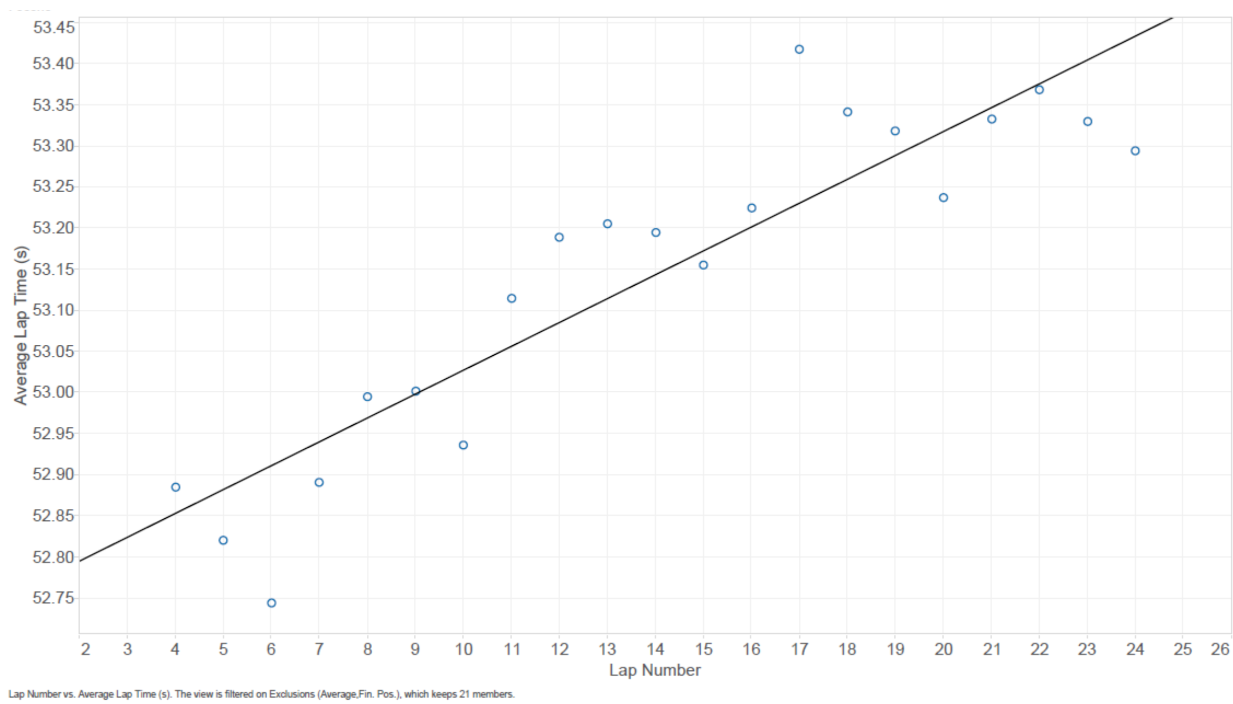


Figure 69. Slope of Tire Degradation for Northeast_B Track based on Northeast_B Race 1 2014

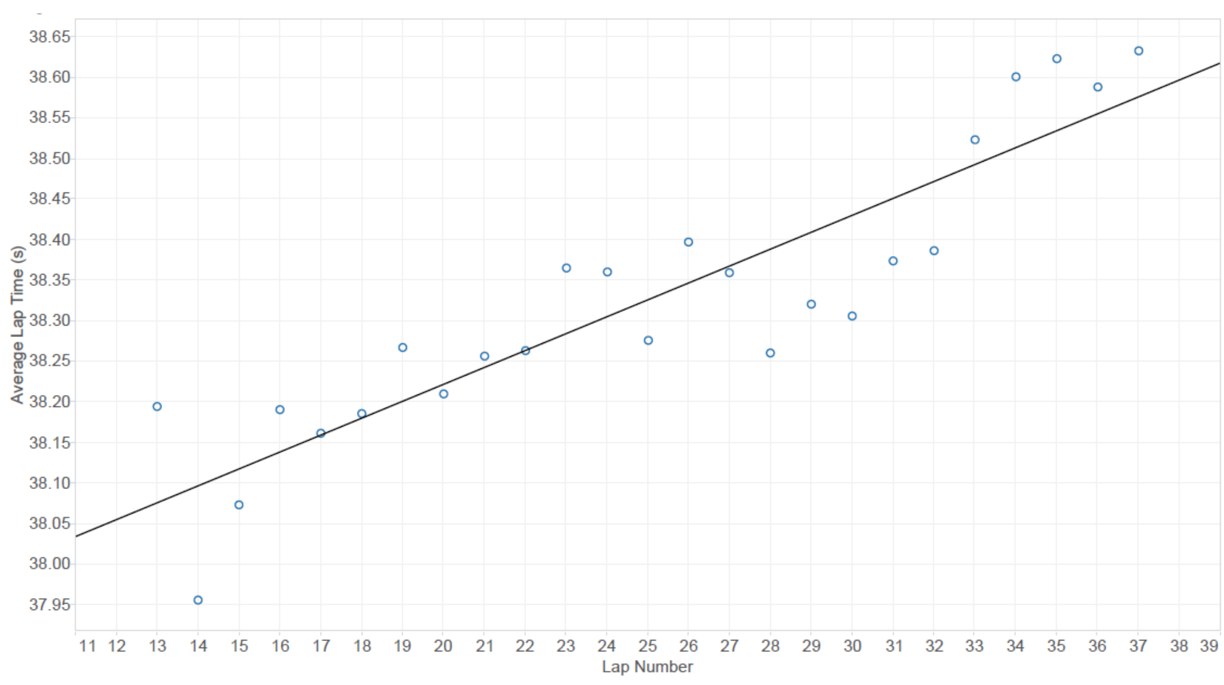


Figure 70. Slope of Tire Degradation for Midwest_A Track based on Midwest_A Race 1 2014

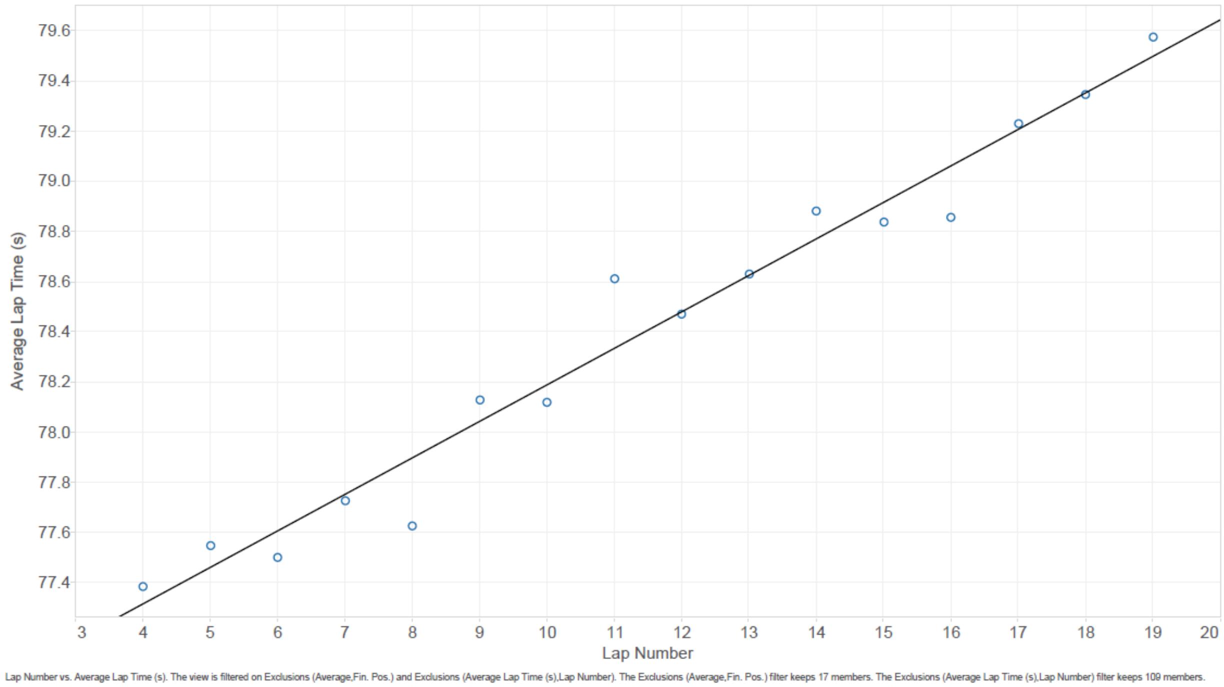


Figure 71. Slope of Tire Degradation for West_B Track based on West_B Race 2014

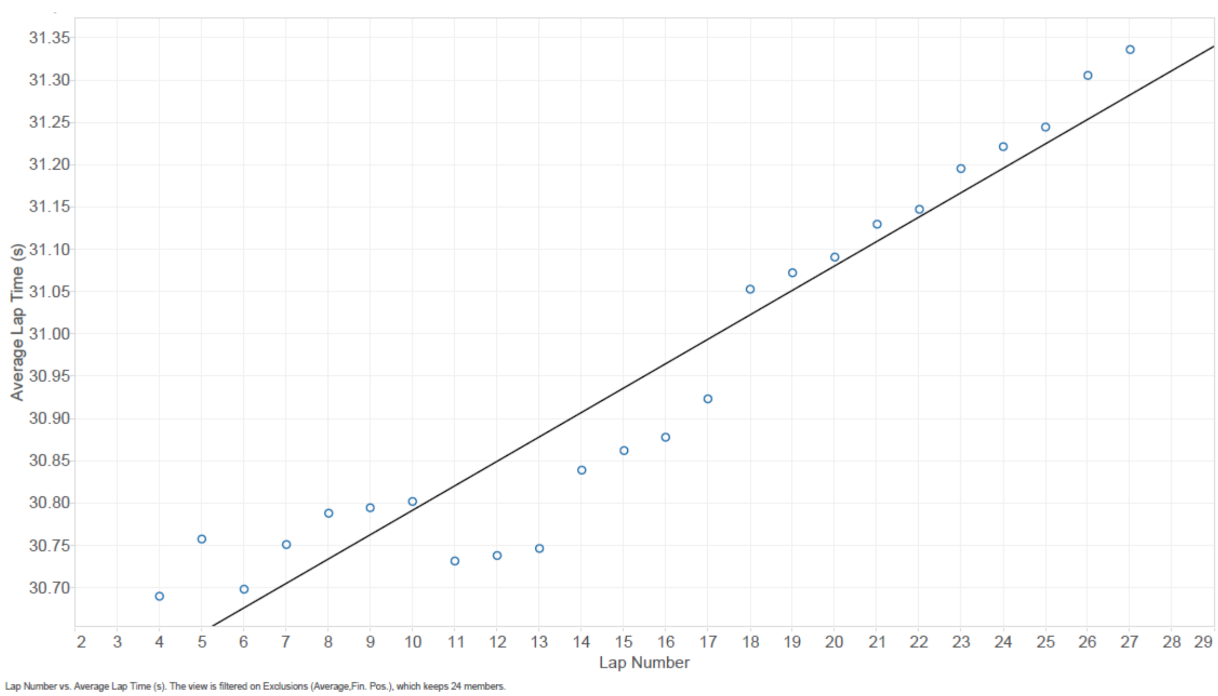


Figure 72. Slope of Tire Degradation for Midwest_D Track based on Midwest_D Race 2014

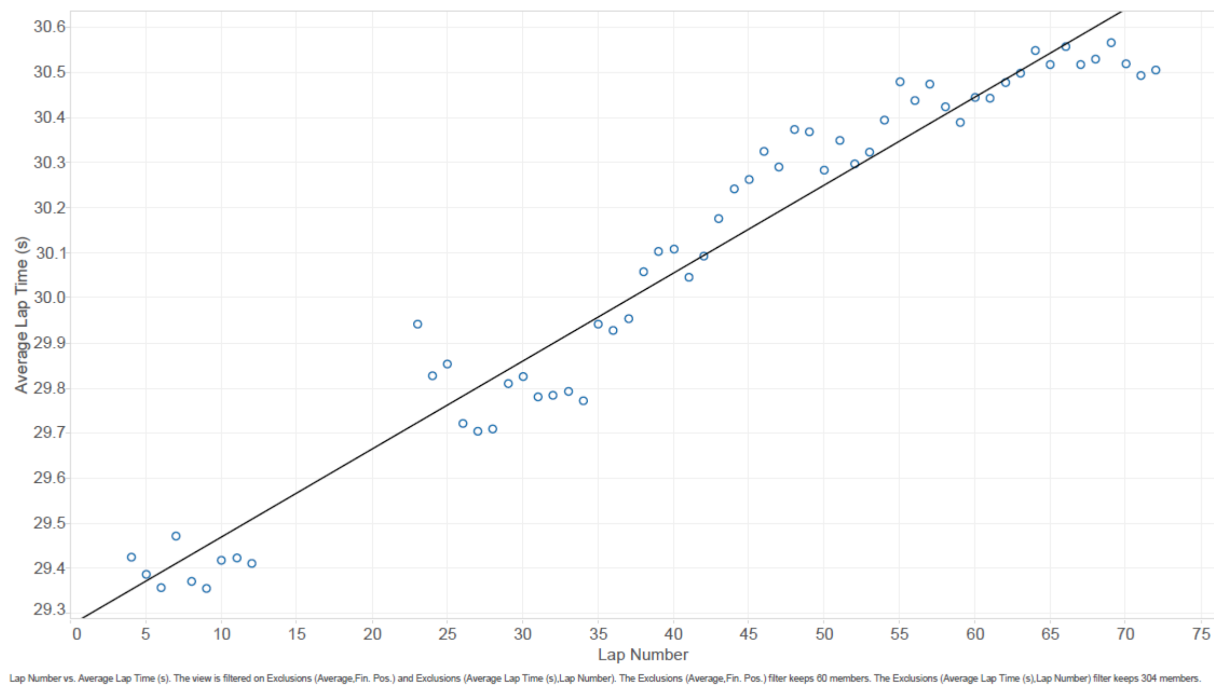


Figure 73. Slope of Tire Degradation for Northeast_C Track based on Northeast_C Race 1 2014

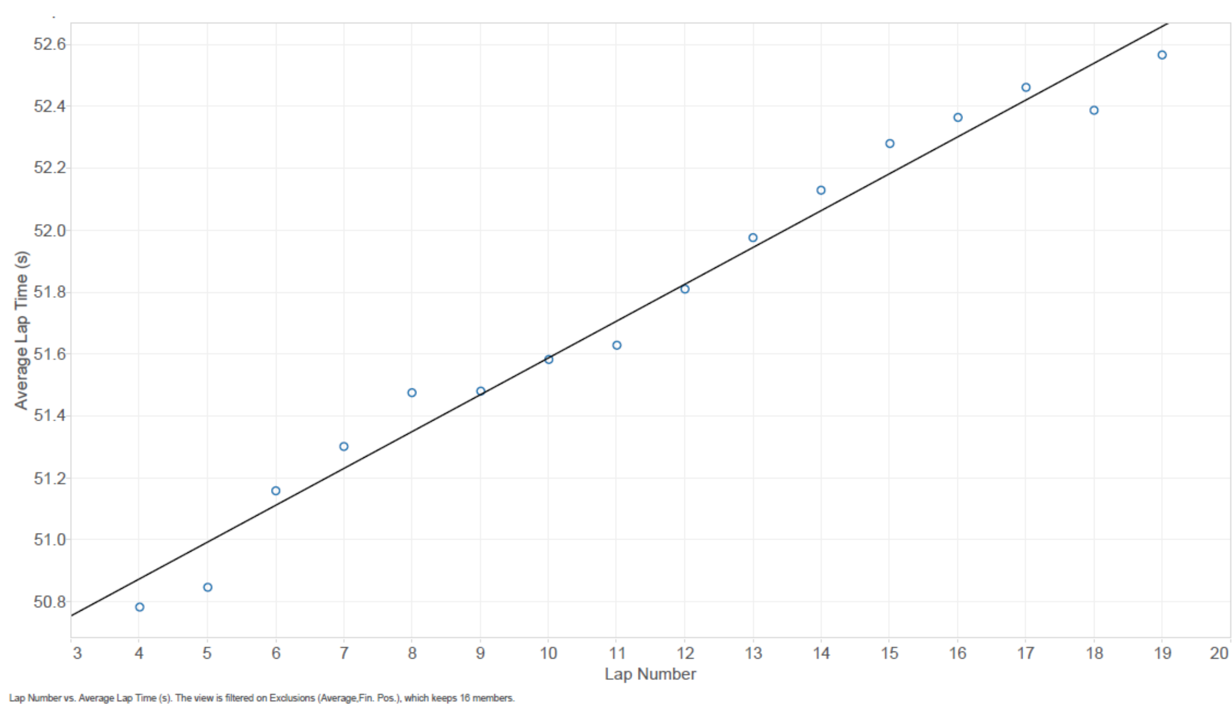


Figure 74. Slope of Tire Degradation for Midwest_E Track based on Midwest_E Race 2014

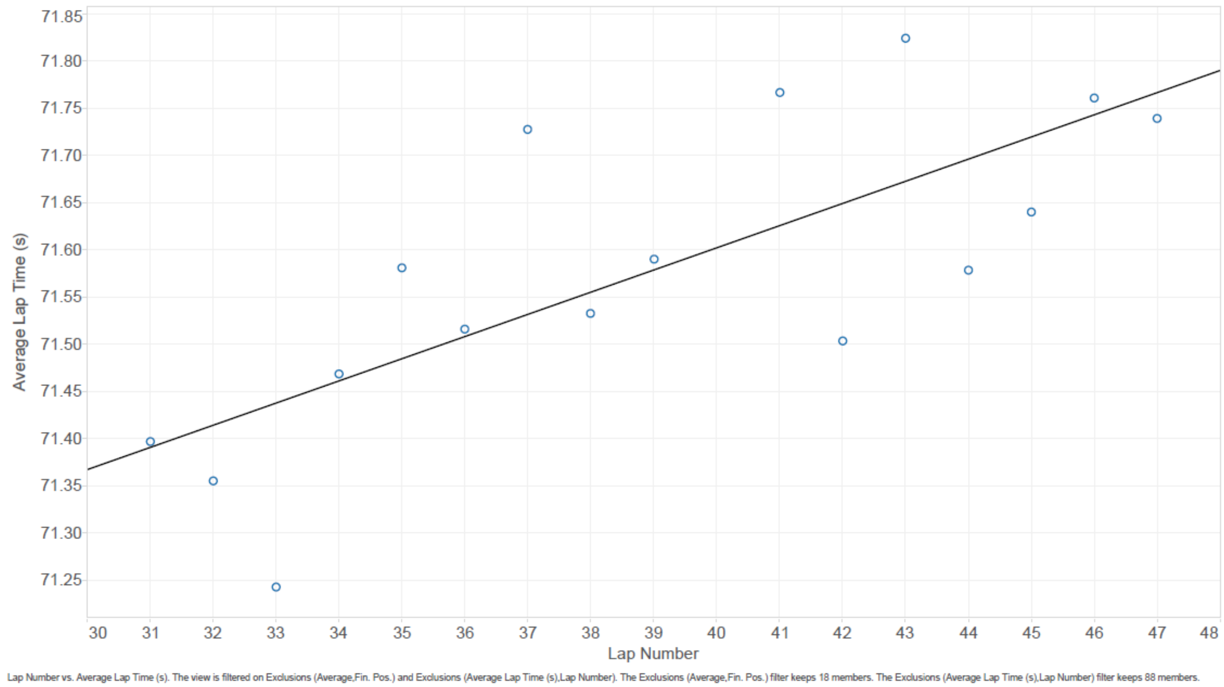


Figure 75. Slope of Tire Degradation for Northeast_A Track based on Northeast_A Race 2014

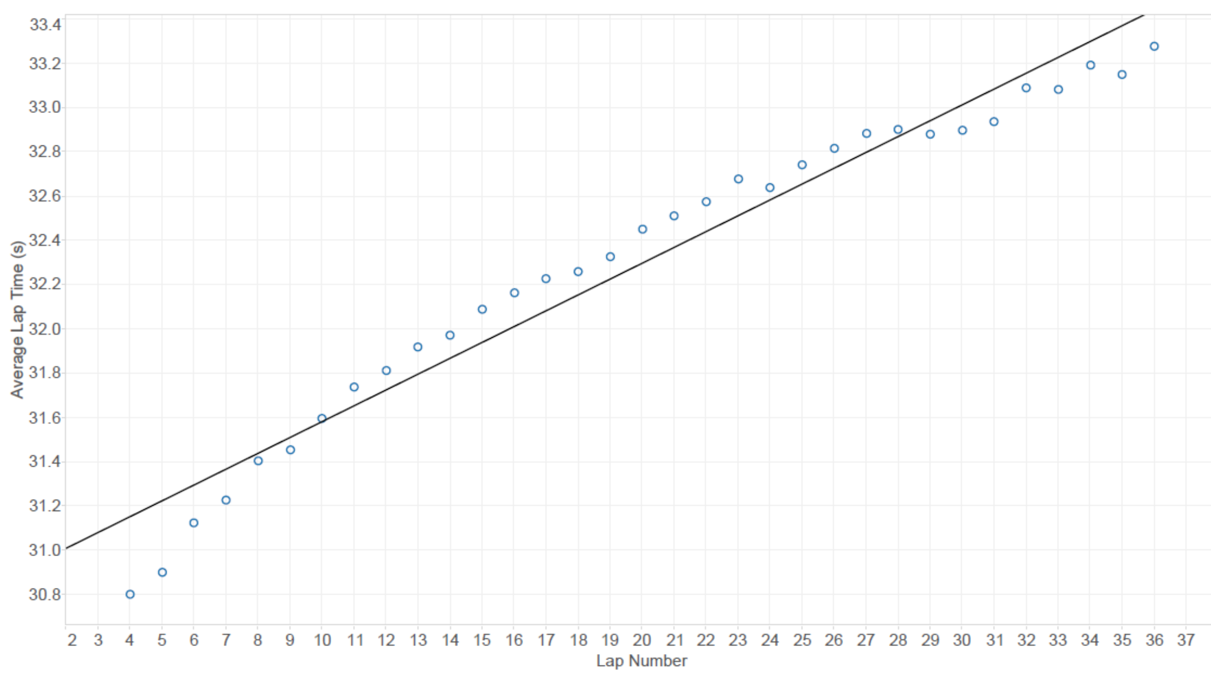


Figure 76. Slope of Tire Degradation for Southeast_I Track based on Southeast_I Race 2014

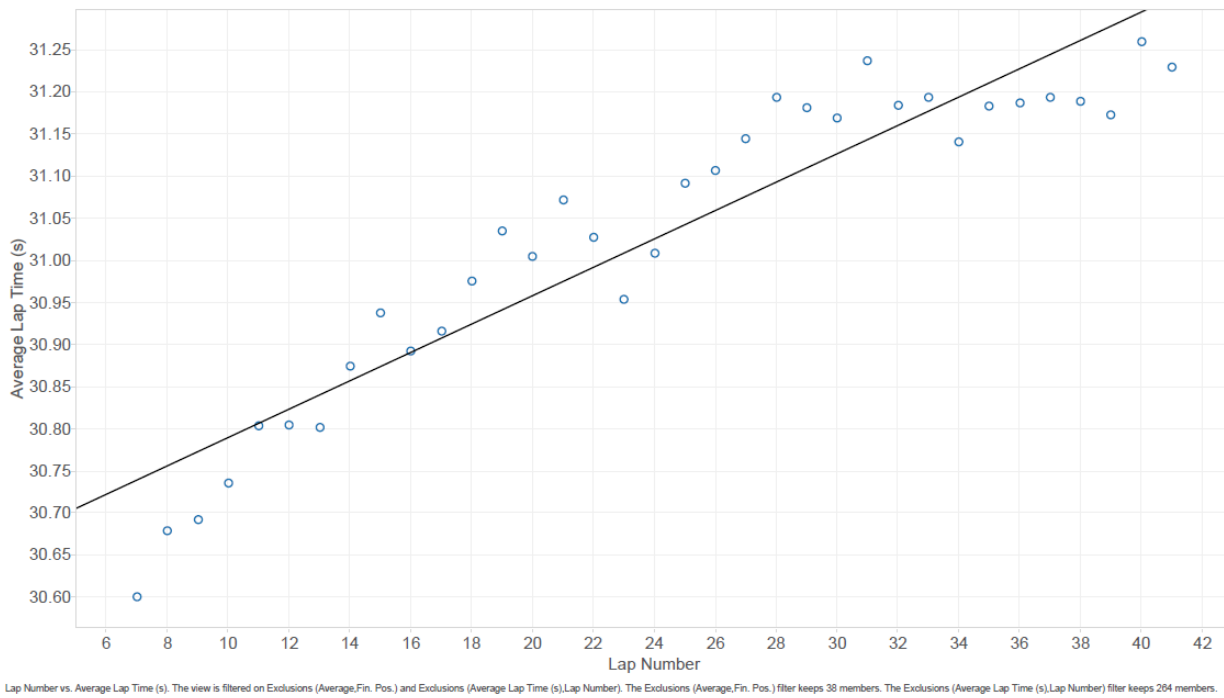


Figure 77. Slope of Tire Degradation for Midwest_B Track based on Midwest_B Race 2014

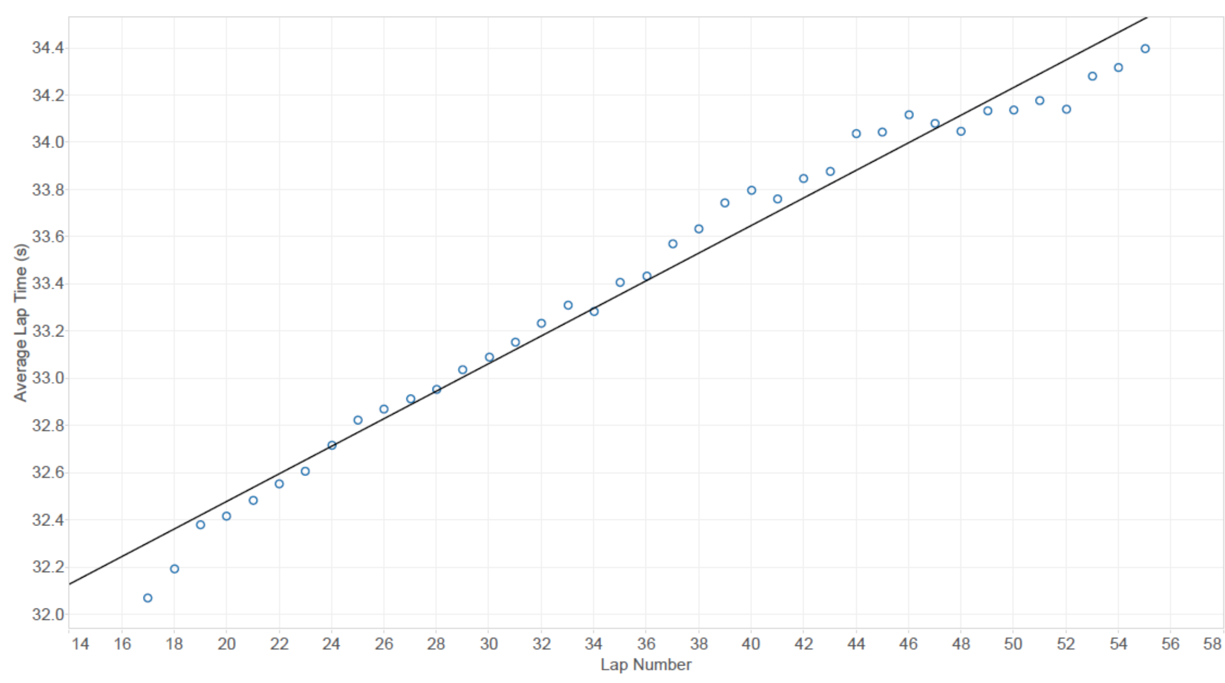
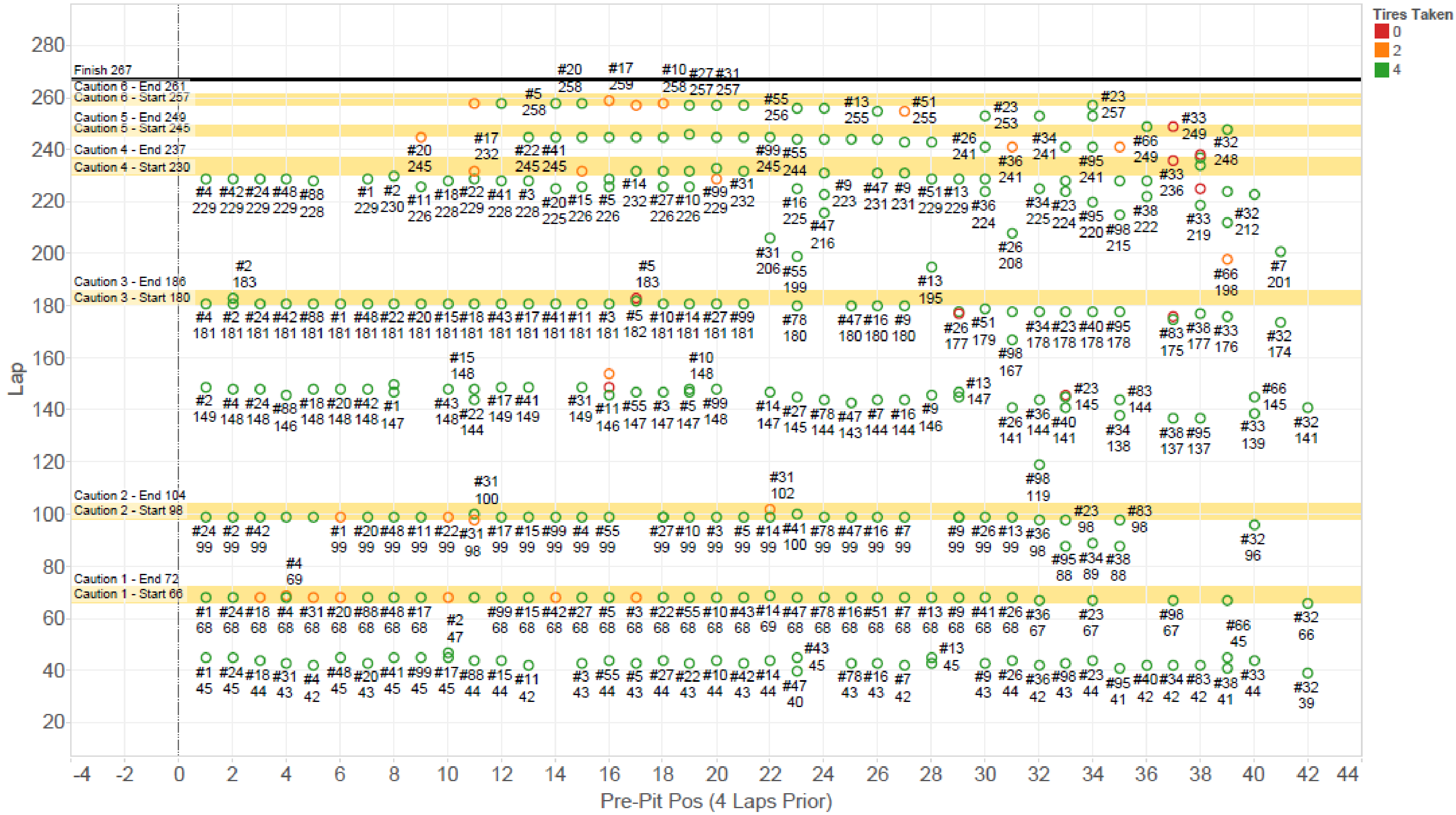


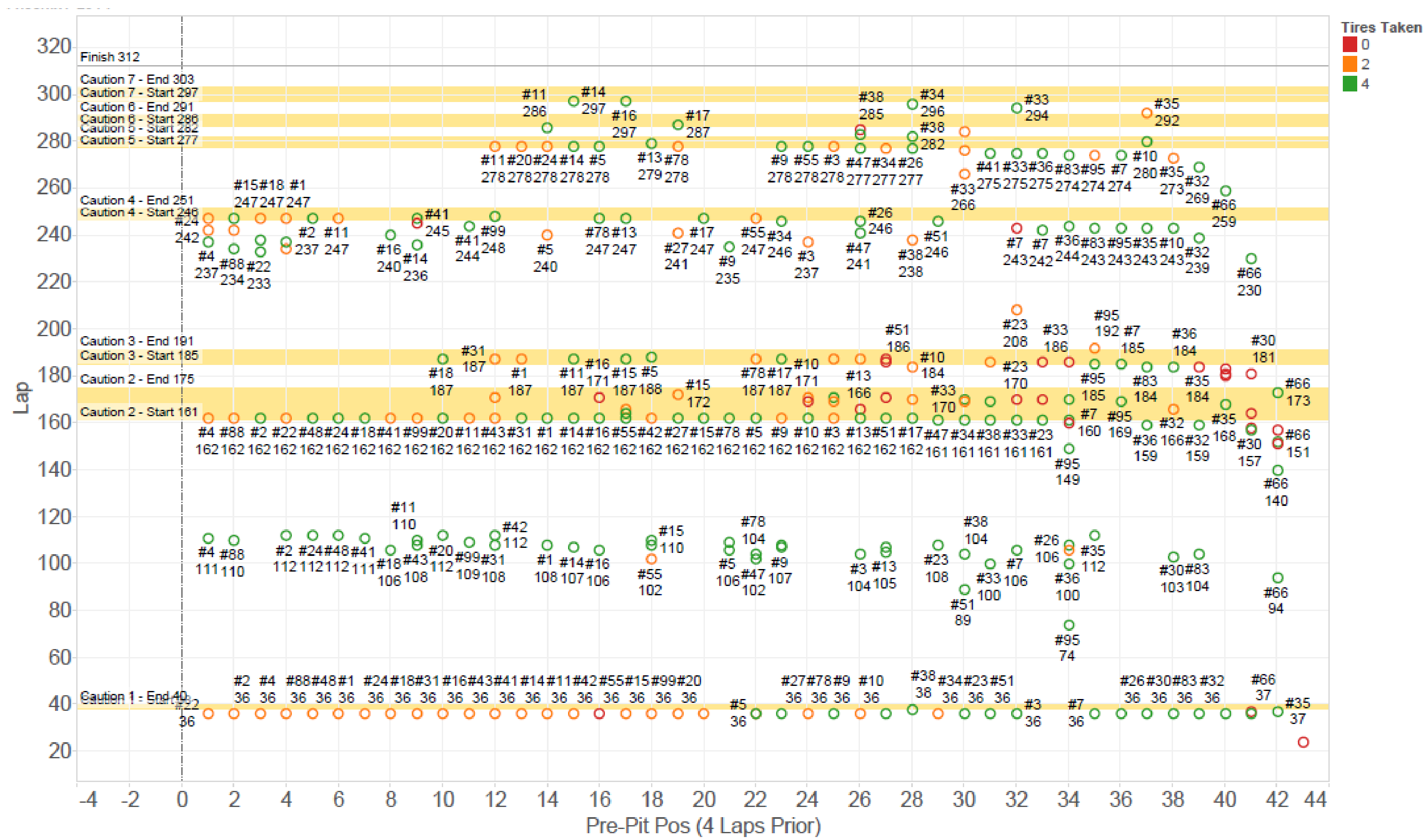
Figure 78. Slope of Tire Degradation for Southeast_G Track based on Southeast_G Race 2014

9.4 Tire Change Strategy according to Pre-Pit Position and Lap Number



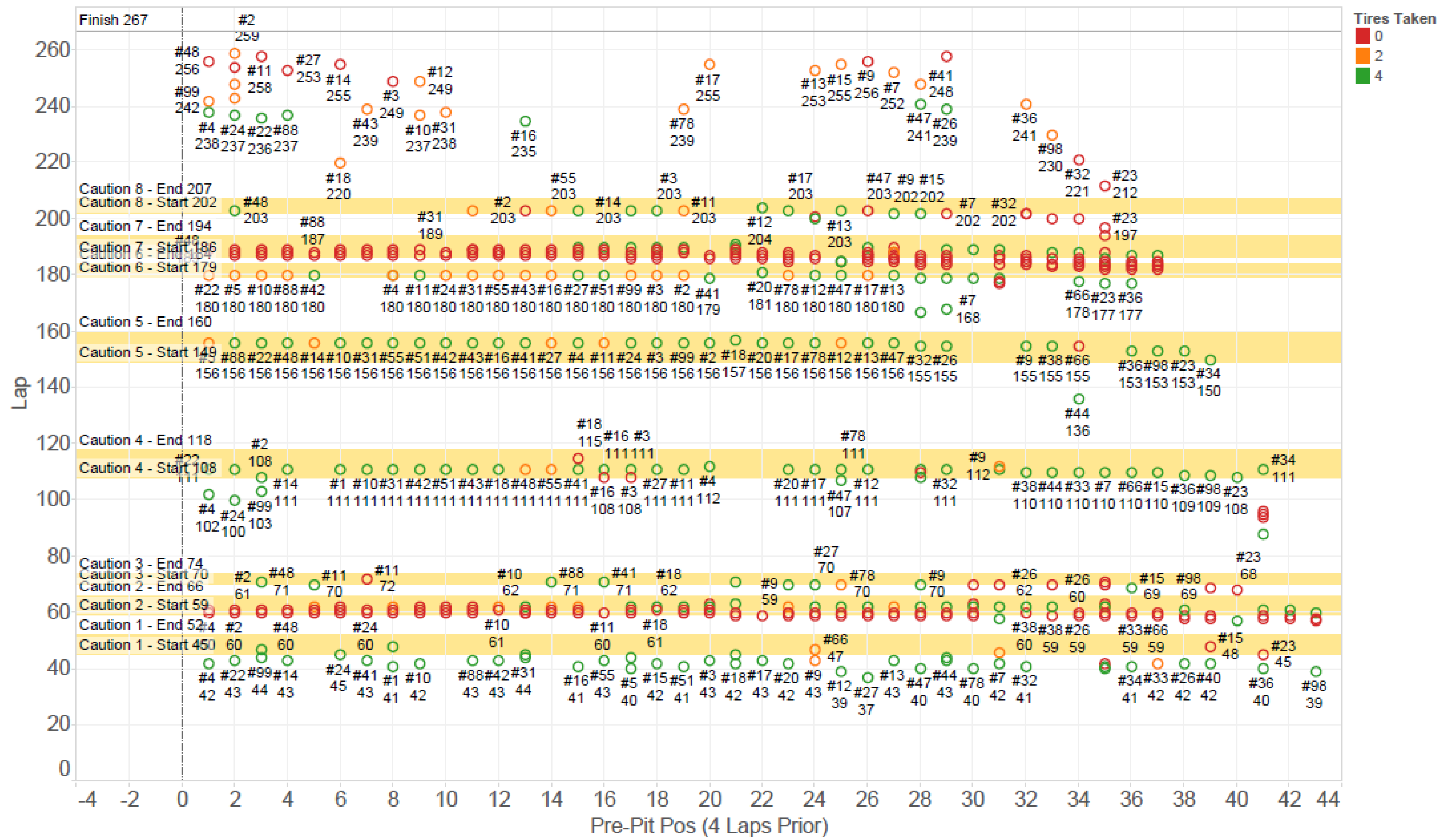
Pre-Pit Pos (4 Laps Prior) vs. Lap. Color shows details about sum of Tires Taken. The marks are labeled by Car and sum of Lap.

Figure 79. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Midwest_B Race 2014



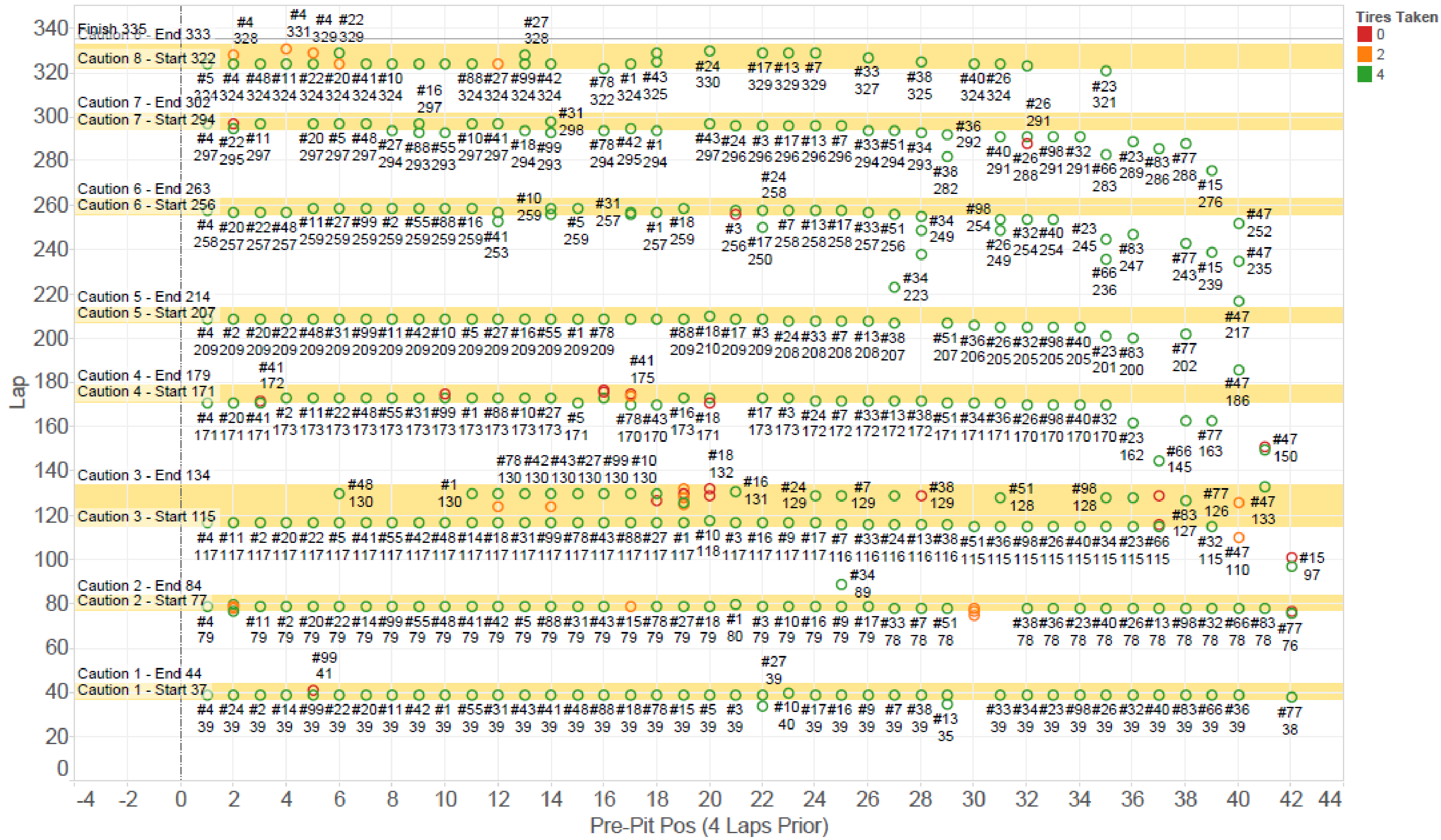
Pre-Pit Pos (4 Laps Prior) vs. Lap. Color shows details about sum of Tires Taken. The marks are labeled by Car and sum of Lap.

Figure 80. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southwest_A Race 1 2014



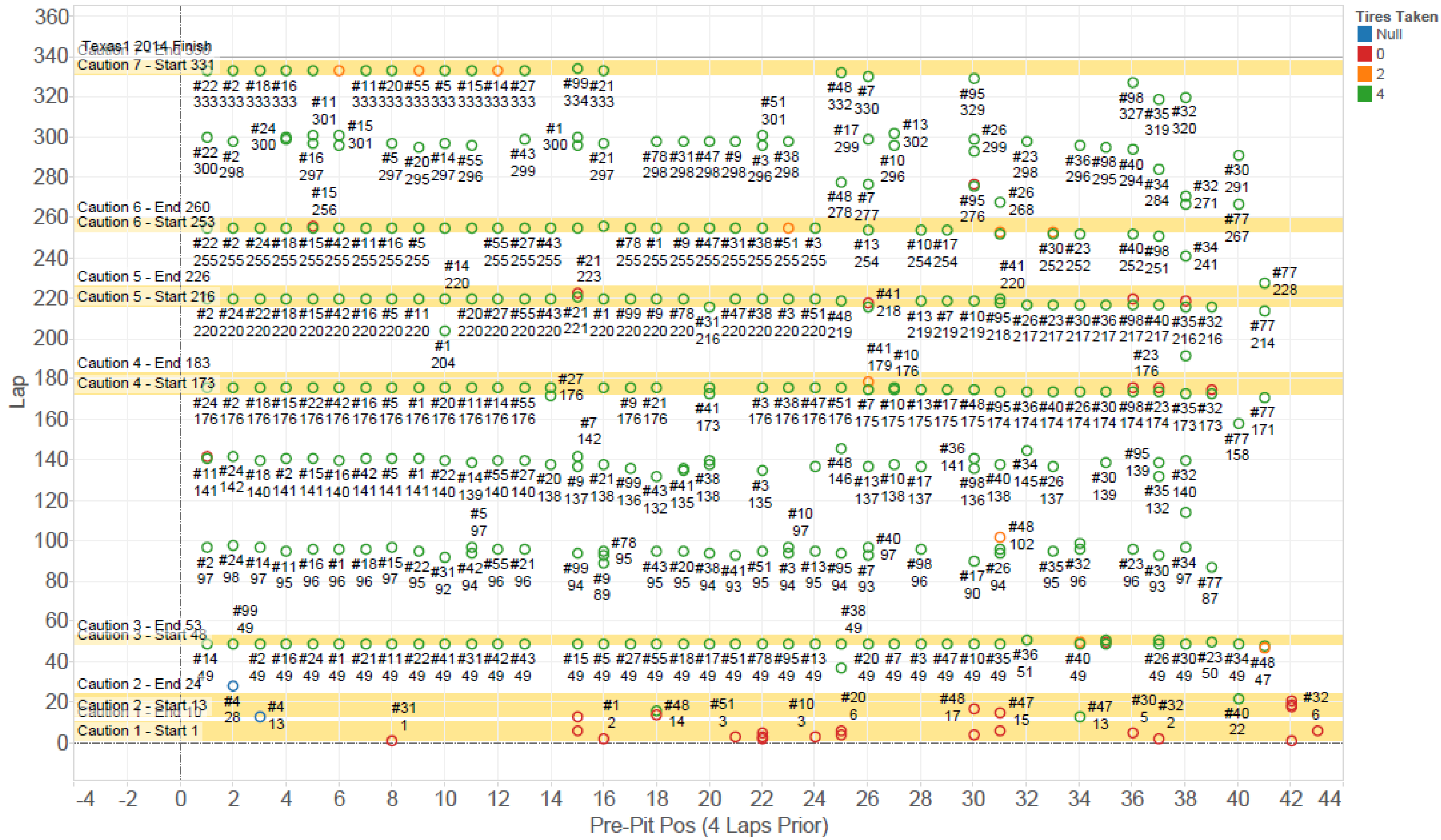
Pre-Pit Pos (4 Laps Prior) vs. Lap. Color shows details about sum of Tires Taken. The marks are labeled by Car and sum of Lap.

Figure 81. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Midwest_C Race 1 2014



Pre-Pit Pos (4 Laps Prior) vs. Lap. Color shows details about sum of Tires Taken. The marks are labeled by Car and sum of Lap.

Figure 82. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southeast_I Race 2014



Pre-Pit Pos (4 Laps Prior) vs. Lap. Color shows details about sum of Tires Taken. The marks are labeled by Car and sum of Lap.

Figure 83. Tire Change Decisions according to Pre-Pit Track Position and Lap Number for Southwest_B Race 1 2014