

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



Más allá de la recta: estrategias de
Fórmula 1 con modelos lineales
generalizados y árboles de decisión

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADA EN CIENCIA DE DATOS

PRESENTA

AINÉ FERNÁNDEZ GREGORIO

ASESOR

DR. FELIPE JAVIER MEDINA AGUAYO

Resumen

La estrategia de carrera es uno de los principales desafíos que enfrentan los equipos de Fórmula 1 en el día de la competencia. Esta tesis tiene como objetivo determinar la estrategia óptima para cada equipo mediante la estimación del tiempo esperado de diversas alternativas estratégicas, utilizando modelos lineales generalizados. Además, se emplean modelos basados en árboles de decisión para comparar y seleccionar la estrategia más adecuada. Para el análisis, se emplean datos de las temporadas 2019-2023, los cuales se descargaron de la API oficial de la Fórmula 1, y fueron procesados para estimar los coeficientes de los modelos que permiten calcular el tiempo de vuelta, el tiempo perdido en *pits* y la probabilidad de aparición de un *Safety Car*. Finalmente, se comparan las estrategias óptimas obtenidas con las estrategias reales aplicadas en algunas carreras de la temporada 2024. En carreras sin intervenciones del *Safety Car*, se obtuvo una precisión aceptable, entre el 70 % y el 75 %. Sin embargo, en carreras con intervenciones del *Safety Car* o con estrategias arriesgadas por parte de los equipos, la precisión del modelo disminuye.

Summary

Race strategy is one of the main challenges faced by Formula 1 teams on race day. This thesis aims to determine the optimal strategy for each team by estimating the expected time for various strategic alternatives, using generalized linear models. Additionally, decision tree-based models are employed to compare and select the most suitable strategy. For the analysis, data from the 2019-2023 seasons was used, downloaded from the official Formula 1 API and processed to estimate the coefficients of the models that calculate lap times, time lost in the pits, and the probability of a *Safety Car* appearance. Finally, the optimal strategies obtained are compared with the actual strategies applied in some races of the 2024 season. In races without *Safety Car* interventions, the model achieved an acceptable accuracy of between 70 % and 75 %. However, in races with *Safety Car* interventions or with risky strategies implemented by teams, the model's accuracy decreases.

Índice general

| | |
|---|-----------|
| Introducción | 1 |
| 1. El mundo de la Fórmula 1 | 5 |
| 1.1. ¿Qué es la Fórmula 1? | 5 |
| 1.1.1. ¿Por qué nadie quiere parar en <i>pits</i> ? | 6 |
| 1.1.2. La estrategia de carrera en la Fórmula 1 | 8 |
| 1.1.3. La suerte también juega en la pista | 9 |
| 1.2. Trabajo relacionado | 11 |
| 1.2.1. Literatura | 11 |
| 1.2.2. En la Fórmula 1 | 13 |
| 2. El motor oculto de la F1: Datos | 15 |
| 2.1. Extracción de datos | 15 |
| 2.2. Creación de variables y limpieza de datos | 17 |
| 2.2.1. Tiempos de vuelta | 18 |
| 2.2.2. Costo de la parada en <i>pits</i> | 21 |
| 2.2.3. Coches de seguridad (<i>Safety Cars</i>) | 22 |
| 2.3. Base de datos final | 23 |
| 3. Vueltas de reconocimiento: Exploración de Datos | 26 |
| 3.1. Vueltas Rápidas en Seco y Neumáticos | 26 |

| | | |
|-----------|---|-----------|
| 3.1.1. | Tiempos de vuelta por kilómetro por Piloto y Equipo | 27 |
| 3.1.2. | Neumáticos | 29 |
| 3.2. | Circuitos y pitstops | 30 |
| 3.3. | Safety Cars | 32 |
| 4. | El plan perfecto para la victoria: Modelos | 36 |
| 4.1. | Modelos lineales generalizados | 36 |
| 4.1.1. | Estimación y ajuste del modelo | 38 |
| 4.1.2. | Evaluación del ajuste y selección de modelos | 39 |
| 4.2. | <i>LapTimePerKM</i> | 42 |
| 4.2.1. | Selección de variables y modelo | 42 |
| 4.2.2. | Resultados | 44 |
| 4.3. | <i>Pitstops</i> | 46 |
| 4.3.1. | Selección de variables y modelo | 46 |
| 4.3.2. | Resultados | 48 |
| 4.4. | <i>Inlaps</i> | 50 |
| 4.4.1. | Selección de variables y modelo | 50 |
| 4.4.2. | Resultados | 51 |
| 4.5. | <i>Outlaps</i> | 53 |
| 4.5.1. | Selección de variables y modelo | 53 |
| 4.5.2. | Resultados | 53 |
| 4.6. | <i>Safety Car</i> | 55 |
| 4.6.1. | Selección de variables y modelo | 55 |
| 4.6.2. | Resultados | 56 |
| 5. | Arranca la carrera: Implementación | 58 |
| 5.1. | Árboles de Decisión | 58 |
| 5.2. | Funciones comunes de los modelos | 61 |
| 5.2.1. | vidapromedio | 61 |

| | |
|--|-----------|
| 5.2.2. <i>tiempoStint</i> | 62 |
| 5.2.3. <i>pitstopcost</i> | 63 |
| 5.3. Modelo Determinista | 63 |
| 5.3.1. Segundo <i>stint</i> con el mismo compuesto | 64 |
| 5.3.2. Segundo <i>stint</i> con un compuesto diferente | 65 |
| 5.3.3. Resultados | 66 |
| 5.4. Modelo Ventana | 66 |
| 5.5. Modelo Rival | 67 |
| 6. La Línea de Meta: Resultados | 70 |
| 6.1. Gran Premio de Bahréin 2024 | 71 |
| 6.1.1. Modelo Determinista | 72 |
| 6.1.2. Modelo Ventana | 73 |
| 6.2. Gran Premio de Singapur 2024 | 74 |
| 6.2.1. Modelo Determinista | 75 |
| 6.2.2. Modelo Ventana | 76 |
| 6.3. Gran Premio de Monza 2024 | 76 |
| 6.3.1. Modelo Determinista | 77 |
| 6.3.2. Modelo Ventana | 78 |
| 6.4. Gran Premio de China 2024 | 78 |
| 6.4.1. Modelo Determinista | 79 |
| 6.4.2. Modelo Ventana | 81 |
| 6.5. <i>Safety Cars</i> | 81 |
| Debrief: Discusión | 85 |
| Glosario | 89 |
| Referencias | 92 |

Índice de cuadros

| | | |
|------|--|----|
| 2.1. | Tipos de compuestos de neumáticos para vueltas en seco en Fórmula 1 | 19 |
| 2.2. | Cambio de nombres de equipos en la Fórmula 1 | 20 |
| 2.3. | Descripción de las etiquetas de estatus de la pista | 23 |
| 2.4. | Descripción de variables en el conjunto de datos de vueltas rápidas | 24 |
| 2.5. | Descripción de variables del conjunto de datos de <i>pitstops</i> | 25 |
| 2.6. | Descripción de variables del conjunto de datos de <i>Safety Cars</i> | 25 |
| 3.1. | Longitud de los circuitos | 30 |
| 3.2. | Riesgo según el porcentaje de ocurrencia | 33 |
| 3.3. | Porcentaje de Ocurrencia y Categoría de Riesgo por Circuito (2019-2023) | 34 |
| 4.1. | Comparación de diferentes familias de distribución para el modelo <i>LapTimePerKM</i> | 43 |
| 4.2. | <i>LapTimePerKM</i> - Intercepto y otras variables | 44 |
| 4.3. | <i>LapTimePerKM</i> - Equipos | 45 |
| 4.4. | <i>LapTimePerKM</i> - Circuitos | 46 |
| 4.5. | <i>LapTimePerKM</i> - Pilotos | 47 |

| | | |
|-------|--|----|
| 4.6. | Comparación de diferentes familias de distribución para el modelo de <i>Pitstops</i> | 48 |
| 4.7. | <i>Pitstop Time</i> - Intercepto | 48 |
| 4.8. | <i>Pitstop Time</i> - Circuitos | 49 |
| 4.9. | Comparación de diferentes familias de distribución para el modelo de <i>Inlaps</i> | 50 |
| 4.10. | <i>LapTimePerKM-Inlaps</i> Intercepto y otras variables | 51 |
| 4.11. | <i>LapTimePerKM-Inlaps</i> (Circuitos) | 52 |
| 4.12. | Comparación de diferentes familias de distribución para el modelo de <i>Outlaps</i> | 53 |
| 4.13. | <i>Outlaps</i> - Intercepto y Compuestos | 54 |
| 4.14. | <i>Outlaps</i> - Circuitos | 54 |
| 4.15. | Comparación de modelos para la probabilidad de un <i>Safety Car</i> | 55 |
| 4.16. | Probabilidad de <i>Safety Car</i> - Intercepto y LapNumber | 56 |
| 4.17. | Probabilidad de <i>Safety Car</i> - Circuitos | 57 |
| 5.1. | Ejemplo del modelo rival | 69 |
| 6.1. | Estrategias óptimas propuestas por el modelo para Bahréin 2024 | 73 |
| 6.2. | Mejor Estrategia y Ventana de Parada | 74 |
| 6.3. | Estrategias óptimas y su precisión para Singapur 2024 | 75 |
| 6.4. | Mejor Estrategia y Ventana de Parada | 76 |
| 6.5. | Estrategias óptimas y su precisión según el modelo para Monza 2024 | 78 |
| 6.6. | Mejor Estrategia y Ventana de Parada | 78 |
| 6.7. | Estrategias óptimas y su precisión según el modelo determinista para China 2024 | 80 |
| 6.8. | Mejor Estrategia y Ventana de Parada | 81 |

Índice de figuras

| | |
|---|----|
| 3.1. Tiempos de vuelta por KM por piloto y por equipo | 28 |
| 3.2. Comportamiento del tipo de neumáticos | 29 |
| 3.3. Tiempos en <i>pits</i> por circuito | 31 |
| 3.4. Tiempos de vuelta por KM por circuito | 32 |
| 3.5. Safety Cars por vuelta | 33 |
| 5.1. Árbol de decisión que siguen los modelos | 60 |
| 5.2. Rama del Primer <i>stint</i> con <i>Hards</i> | 64 |
| 6.1. Estrategias de neumáticos para el Gran Premio de Bahréin 2024 | 72 |
| 6.2. Estrategias de neumáticos para el Gran Premio de Singapur 2024 | 74 |
| 6.3. Estrategias de neumáticos para el Gran Premio de Monza 2024 | 77 |
| 6.4. Estrategias de neumáticos para el Gran Premio de China 2024 | 79 |
| 6.5. Probabilidad de <i>Safety Car</i> : Circuitos de riesgo bajo . . | 82 |
| 6.6. Probabilidad de <i>Safety Car</i> : Circuitos de riesgo medio . | 83 |
| 6.7. Probabilidad de <i>Safety Car</i> : Circuitos de riesgo alto . . | 83 |
| 6.8. Probabilidad promedio de <i>Safety Car</i> por nivel de riesgo | 84 |

Introducción

¿Cuánto importa un segundo en el apasionante mundo de la Fórmula 1? La respuesta es simple: todo. Un segundo puede definir la *pole position* en una clasificación donde las diferencias se miden en milésimas, o puede convertirse en la línea divisoria entre la gloria y la decepción en una batalla por el podio. Recordemos el Gran Premio de Brasil de 2023, donde Fernando Alonso y Checo Pérez protagonizaron una batalla por el podio que se tuvo que definir por *photofinish* que dejó a los aficionados al borde de sus asientos.

En la Fórmula 1, cada segundo cuenta. Por esta razón, los mecánicos dedican la mayor parte del año a perfeccionar la técnica de cambiar neumáticos en tan solo 2 segundos o menos. De hecho, el récord de la parada más rápida en *pits*¹ pertenece a McLaren, con un impresionante tiempo de 1.80 segundos. Increíble, ¿verdad?

Muchos creen erróneamente que tener el coche más rápido y un piloto excepcional son las claves para conquistar el campeonato. Sin embargo, la realidad es que los equipos de Fórmula 1 invierten millones de dólares en *software* y departamentos de estrategia porque la victoria no solo depende de la velocidad, sino también de contar con

¹Para la definición de términos propios de la Fórmula 1 usados en esta tesis, consultar el glosario al final.

una estrategia ganadora. Si bien es más fácil idear la estrategia perfecta con un coche rápido, un error puede hacer que incluso el vehículo más dominante parezca lento. Una estrategia arriesgada, en este contexto, puede convertir un coche con rendimiento mediocre en uno que compita con los más veloces de la parrilla, todo gracias a una adecuada gestión de neumáticos.

Los neumáticos nuevos se traducen en un mejor rendimiento, lo que hace que decidir qué compuesto usar y cuándo realizar las paradas sea crucial. Un error puede costar valiosas posiciones en la pista, mientras que una decisión acertada puede catapultar a un piloto a la victoria.

El objetivo primordial de cada equipo al iniciar una carrera es completar la distancia en el menor tiempo posible. Sin embargo, este objetivo se complica por las regulaciones que exigen el uso de al menos dos compuestos diferentes de neumáticos durante la carrera. Esto crea un dilema: para hacer sus paradas en los *pits*, los coches deben cruzar el *pitlane*, que tiene un límite de velocidad, usualmente de 80 km/h, lo cual resulta increíblemente lento para un coche de Fórmula 1. Aunque las paradas sean rápidas, el tiempo perdido en *pits* se traduce en la pérdida de posiciones, a menos que los pilotos logren salir con neumáticos frescos que les ofrezcan una ventaja sobre sus rivales que aún no han parado. Además, factores impredecibles como la aparición del *Safety Car* o cambios climáticos pueden alterar el balance entre el tiempo perdido en *pits* y los coches que continúan girando en la pista.

Entonces, ¿qué pasaría si pudiéramos estimar el tiempo esperado de una estrategia determinada y decidir cuál es la mejor para cada equipo? Este es el objetivo central de esta tesis. Estimar el tiempo esperado de una estrategia es un desafío complejo, ya que implica considerar múltiples factores que afectan el tiempo de vuelta. Básicamente, se deben estimar dos componentes clave: el tiempo

perdido en *pits* y el tiempo de cada *stint*, que depende de los tiempos de vuelta estimados durante la duración del *stint*. Sin embargo, este proceso no es sencillo, dado que el rendimiento de los neumáticos varía a lo largo de la carrera.

Dado que el objetivo es estimar el tiempo total esperado de una estrategia de carrera, resulta crucial calcular con precisión estos dos componentes clave, los cuales están influenciados por múltiples factores. Los modelos lineales generalizados son una herramienta adecuada para este propósito, ya que permiten estimar el tiempo de vuelta por kilómetro, lo que posibilita calcular el tiempo total esperado de los *stints*. Además, estos modelos pueden utilizarse para estimar el tiempo perdido en *pits* y la probabilidad de eventos inciertos como la intervención del *Safety Car*, lo que permite adaptar la estrategia de carrera ante estas intervenciones.

Mediante modelos lineales generalizados, se busca identificar los factores que influyen en los tiempos de vuelta, así como aquellos que determinan el tiempo perdido en *pits*, lo que nos permitirá estimar el tiempo de cada estrategia con diferentes compuestos de neumáticos.

Una vez que tenemos los tiempos esperados, la decisión sobre qué estrategia implementar no es fácil. Para abordar esta incertidumbre, también utilizamos árboles de decisión, que crean diversas estrategias y nos ayudan a seleccionar la opción con el mínimo tiempo esperado.

Esta tesis se propone contribuir a la literatura existente sobre la determinación de la estrategia óptima en Fórmula 1 al proporcionar un nuevo enfoque para optimizar estrategias a través de modelos lineales generalizados y árboles de decisión. La estructura de esta tesis es la siguiente:

En el Capítulo 1, se ofrece un contexto sobre el mundo de la Fórmula 1 y se revisan trabajos relacionados. El Capítulo 2 detalla el

proceso de extracción de datos y la transformación de variables previas al análisis. En el Capítulo 3, se lleva a cabo un análisis exploratorio de datos, fundamental en toda investigación. El Capítulo 4 presenta la teoría de los modelos lineales generalizados y los resultados de las estimaciones. En el Capítulo 5, se describe la implementación de los árboles de decisión. Finalmente, el Capítulo 6 presenta los resultados del análisis, comparando los hallazgos del modelo con carreras reales de la temporada 2024.

Capítulo 1

El mundo de la Fórmula 1

1.1. ¿Qué es la Fórmula 1?

La Fórmula 1 es la categoría reina del automovilismo y la competición de carreras de coches más prestigiosa del mundo. Actualmente, 20 pilotos y 10 equipos conforman la parrilla de la Fórmula 1, con cada equipo compuesto por dos pilotos y sus respectivos monoplazas. Los pilotos compiten por el Campeonato de Pilotos, mientras que como compañeros de equipo suman puntos para su equipo en el Campeonato de Constructores.

La primera temporada de Fórmula 1 se celebró en 1950, y la carrera inaugural tuvo lugar en el circuito de Silverstone, Gran Bretaña, un trazado que sigue siendo parte del calendario actual. Cada temporada se organiza en una serie de Grandes Premios, que se llevan a cabo en distintos circuitos alrededor del mundo.

Cada Gran Premio se desarrolla a lo largo de un fin de semana, de viernes a domingo, y sigue generalmente la siguiente estructura:

Prácticas Libres: Tres sesiones de una hora cada una en las que

los equipos tienen la oportunidad de ajustar la configuración de los coches y recopilar datos clave. Durante estas sesiones, los pilotos realizan vueltas rápidas para preparar la clasificación y tandas largas para evaluar el ritmo de carrera. Las Prácticas Libres 1 y 2 se celebran el viernes, mientras que la Práctica Libre 3 tiene lugar el sábado antes de la clasificación.

Clasificación: El sábado, después de la última sesión de prácticas, se lleva a cabo la clasificación para determinar el orden de la parrilla de salida para la carrera del domingo. La clasificación se divide en tres sesiones: Q1, Q2 y Q3. Todos los coches salen a pista para intentar registrar su vuelta más rápida; al final de cada sesión, los cinco coches más lentos son eliminados. El piloto más rápido en Q3 gana la *pole position* y parte desde la primera posición en la carrera del domingo.

Carrera: El evento principal de cada Gran Premio se celebra el domingo. La carrera suele durar entre una hora y media y dos horas, dependiendo de las interrupciones, como la salida del coche de seguridad o las banderas rojas. El número de vueltas varía según el circuito, ya que la distancia total de cada carrera debe exceder los 305 kilómetros. Al finalizar, los pilotos que terminan en las primeras diez posiciones suman puntos para el campeonato.

1.1.1. ¿Por qué nadie quiere parar en *pits*?

Este proyecto se centrará en la carrera principal de cada Gran Premio. Una vez completada la clasificación y con los coches alineados en la parrilla de salida, todos los equipos y pilotos tienen un único objetivo: cruzar la línea de meta lo más rápido posible, es decir, completar la distancia de carrera en el menor tiempo posible.

En la Fórmula 1, la distancia de carrera es de al menos 305 kilómetros. El número de vueltas varía dependiendo del circuito, ya

que se ajusta para alcanzar dicha distancia mínima, excepto en Mónaco, donde la carrera cubre poco más de 260 kilómetros debido a las características de este trazado corto y lento, lo que permite que la duración de la carrera oscile entre 80 y 100 minutos.

El desafío surge porque los neumáticos pierden rendimiento tras varias vueltas, lo que provoca un aumento en los tiempos de vuelta. Para contrarrestar esta degradación, los equipos deben optar por realizar una parada en *pits* para cambiar a un nuevo juego de neumáticos. Pero surge una pregunta clave: ¿por qué los equipos no desean realizar paradas en *pits*? La respuesta es simple: cada segundo perdido en *boxes* se traduce en tiempo perdido en pista, lo que puede resultar en la pérdida de posiciones.

Por tanto, los equipos enfrentan un *trade-off*: hacer una parada en *pits* y perder tiempo, pero salir con neumáticos nuevos y mejor rendimiento, o continuar con neumáticos desgastados que siguen perdiendo efectividad.

El costo de una parada en *pits* se desglosa en tres componentes principales:

Inlap: Es el tiempo de la vuelta en la que el coche ingresa a los *pits*. Esta vuelta suele ser más lenta que una vuelta normal, ya que el piloto debe reducir la velocidad antes de entrar al carril de *pits*.

Tiempo en *pits*: Se refiere al tiempo total desde que el coche entra al carril de *pits* (*pitlane*) hasta que sale de él. Esto incluye el desplazamiento dentro del *pitlane*, donde el límite de velocidad suele ser de 80 km/h (aunque en algunos circuitos puede ser más bajo), y el tiempo que el coche pasa detenido en el *box* para el cambio de neumáticos.

Outlap: Es el tiempo de la primera vuelta tras salir de los *pits*. Esta vuelta suele ser más lenta que una vuelta normal, ya que los

neumáticos nuevos necesitan calentarse y el coche debe recuperar su ritmo de carrera.

1.1.2. La estrategia de carrera en la Fórmula 1

Aquí entra en juego la estrategia de carrera. Por reglamento, cada piloto debe usar al menos dos compuestos de neumáticos diferentes durante la carrera, lo que obliga a realizar al menos una parada en *pits*, aunque esto no sea ideal para muchos equipos. Ha habido casos en los que los pilotos han hecho su única parada en la penúltima vuelta para evitar la descalificación, como el caso de Alex Albon en el Gran Premio de Australia de 2022, donde corrió 57 vueltas con el compuesto duro (*Hard*) y paró en la penúltima vuelta, logrando terminar décimo y obteniendo un punto para Williams.

La estrategia de carrera consiste en decidir cuántas paradas realizar y qué compuestos usar. En la Fórmula 1 actual, hay cinco compuestos de neumáticos: tres para condiciones secas (*Soft*, *Medium* y *Hard*) y dos para condiciones de lluvia (*Intermediate* y *Full Wet*).

La elección del compuesto implica un balance entre durabilidad y rendimiento. Los *Soft* son los más rápidos pero menos duraderos, los *Medium* ofrecen un equilibrio entre ambos factores, y los *Hard* son los más duraderos, pero también los más lentos. No hay un compuesto perfecto para todas las carreras, ya que algunos circuitos presentan mayor degradación de neumáticos, lo que hace inviable un *stint* largo con *Softs*, mientras que otros permiten mayor flexibilidad en la elección del compuesto.

El punto de partida de cualquier estrategia es cómo completar la distancia de carrera en el menor tiempo posible. Luego, se evalúan diferentes escenarios para ganar posiciones durante la carrera, donde la posición en pista es crucial, sobre todo en circuitos con pocas

oportunidades de adelantamiento. Aquí es donde las estrategias de *pits* pueden ser decisivas, ya que si un piloto no puede adelantar en pista, puede hacerlo a través de las paradas en *pits*. A estas estrategias se les conoce como *Undercut* y *Overcut*.

El *Undercut* consiste en parar en *pits* antes que el rival y ser más rápido en la *outlap* con neumáticos nuevos. El *Overcut*, en cambio, implica parar después que el rival y tratar de ser más rápido en la *inlap* con neumáticos aún a buena temperatura, mientras el rival ya tiene neumáticos fríos. El *Undercut* se usa principalmente cuando hay alta degradación de neumáticos, permitiendo al piloto ser más rápido con neumáticos nuevos en comparación con el coche de adelante, que lleva neumáticos desgastados. También puede ser útil cuando el piloto está en tráfico, ya que al estar a menos de un segundo de otro coche entra en lo que se conoce como *aire sucio*, lo que afecta negativamente el rendimiento aerodinámico del coche y la *downforce* crucial en las curvas (Collins, 2024, 129). Esto también afecta las temperaturas de componentes clave, como los frenos y el motor, lo que dificulta el manejo de la degradación de los neumáticos. En cambio, si el piloto ha gestionado bien sus neumáticos, el *Overcut* puede ser una opción, ya que puede mantener un ritmo competitivo con neumáticos desgastados y luego ser más rápido en su *inlap* cuando finalmente decida parar en *pits*.

Cabe destacar que tanto el *Undercut* como el *Overcut* no solo funcionan con diferencias de una vuelta antes o después, sino que también pueden extenderse a varias vueltas.

1.1.3. La suerte también juega en la pista

Por más que los estrategas planifiquen la estrategia óptima desde la vuelta 1 hasta el final de la carrera, hay ciertos eventos que obligan a

cambiar la estrategia. No necesariamente tiene que ser un incidente en pista; puede ser una pérdida de posiciones en la arrancada o una parada lenta en *pits*.

Para los incidentes en pista, los oficiales tienen diferentes banderas y señales para informar a los pilotos y a los equipos sobre la situación actual en la pista:

Green Flag (Bandera verde): significa que la pista está limpia y que no hay ningún incidente que impida a los pilotos circular a velocidades normales.

Yellow Flag (Bandera amarilla): significa que hubo un incidente en pista, como un trompo, un choque o restos en la pista. Esta se usa para advertir a los pilotos de que hay un peligro en la pista y deben reducir la velocidad. Estas pueden aplicarse en toda la pista o solo en sectores específicos.

Safety Car o **Virtual Safety Car**: Coche de seguridad o coche de seguridad virtual. Cuando el incidente que provocó la bandera amarilla no se puede resolver rápidamente, esta puede convertirse en un periodo de **Safety Car** o **Virtual Safety Car (VSC)**. El **VSC** se usa cuando el incidente no es tan grave; en este caso, los pilotos deben respetar un *delta time* y mantener los intervalos de tiempo entre los coches. Una vez resuelto el incidente, los pilotos pueden volver a las velocidades normales cuando se dé la bandera verde. Un **VSC** puede convertirse en un **Safety Car** si lo determinan los oficiales de pista. El **Safety Car** es un coche físico que obliga a los pilotos a conducir a una velocidad determinada, y los intervalos de tiempo entre los coches desaparecen porque los vehículos se agrupan nuevamente. Los oficiales de pista pueden decidir si un incidente requiere ir directamente a **Safety Car** sin pasar por **VSC**.

Red Flag (Bandera roja): Si hay un incidente muy grave en la

pista, como un choque a alta velocidad o de varios coches, o condiciones meteorológicas muy adversas, es común que los oficiales declaren bandera roja. La carrera se detiene y los pilotos regresan a *pits* en el orden en que estaban antes de la bandera roja, esperando a que la carrera se reanude.

Es importante notar que durante un *Safety Car* se pierde menos tiempo en *pits*, y durante una bandera roja se pueden cambiar llantas. Por lo tanto, si sale un *Safety Car* cuando todos los rivales ya han parado, esto podría significar una ventaja para quienes no lo han hecho. De manera similar, podría anular la ventaja de un piloto que haya parado antes.

Todo esto debe ser considerado, y los equipos deben ser capaces de reaccionar rápidamente a estos cambios. También, es necesario tener un poco de suerte.

1.2. Trabajo relacionado

1.2.1. Literatura

El problema de la estrategia de carrera en la Fórmula 1 ha ganado interés recientemente, impulsado por el avance en las capacidades de extracción y análisis de datos, así como la disponibilidad de datos más detallados en este deporte. Diversos estudios han abordado este problema desde distintos enfoques y técnicas.

Heilmeier et al. (2020) proponen un *Virtual Strategy Engineer* (VSE) que, mediante redes neuronales, aborda el problema de determinar la estrategia de *pits* en una carrera considerando factores como la presencia de oponentes y las estrategias de estos. El modelo predice si el piloto debe entrar a *pits* en una vuelta determinada y qué tipo de compuesto

de llanta se debería montar, usando datos de las temporadas 2014-2019. Entre las variables clave que optimiza el modelo se encuentran el número de vueltas restantes, la degradación de los neumáticos y el número de paradas en *pits* restantes. Sin embargo, una limitación importante de este enfoque es que fija de antemano el número de paradas en *pits*, lo cual reduce su flexibilidad en escenarios reales donde dicho número podría variar en respuesta a la evolución de la carrera. Los autores sugieren que, en futuras investigaciones, un modelo basado en aprendizaje por refuerzo podría mejorar el rendimiento al permitir la adaptación dinámica de la estrategia e incorporar elementos de teoría de juegos.

El enfoque de redes neuronales también ha sido aplicado a la predicción de resultados de carrera. Por ejemplo, Stoppels (2017) utiliza redes neuronales para predecir los resultados de varias carreras de las temporadas 2016 y 2017, considerando datos históricos de carreras y factores climáticos. Este tipo de modelado busca anticipar el rendimiento de los equipos bajo condiciones similares en el pasado.

Otro enfoque al problema de la estrategia de carrera ha sido el de la programación dinámica. Thraves et al. (2022) desarrollan un modelo en el que el desgaste de llantas y el ritmo de carrera son tratados como funciones en un marco de programación dinámica, optimizando las paradas en *pits* para minimizar el tiempo total de carrera. Este modelo se amplía mediante programación dinámica estocástica para incorporar eventos aleatorios, como la aparición de *Safety Cars* o cambios en el clima, permitiendo una estrategia que se adapta a estos eventos inciertos. En sus conclusiones, los autores mencionan que el modelo tiende a retrasar las paradas en *pits* en espera de un posible *Safety Car*. No obstante, una limitación señalada es la falta de consideración de los otros pilotos en pista; sugieren que futuras investigaciones podrían beneficiarse de la inclusión de teoría de juegos.

Además, los modelos lineales generalizados han sido aplicados en el contexto de la Fórmula 1. Por ejemplo, McCarthy and Rotthoff (2013) implementan un modelo *Probit* para analizar la probabilidad de que un piloto no finalice la carrera (*DNF*, por sus siglas en inglés) en función de su posición de salida. Este tipo de modelado estadístico proporciona *insights* sobre el impacto de la posición inicial en el riesgo de abandono durante la carrera.

Los avances en la optimización de estrategias para la Fórmula 1 han sido sustanciales gracias a la incorporación de técnicas como redes neuronales, programación dinámica y modelos estadísticos. Los enfoques basados en redes neuronales ofrecen un gran potencial para la automatización de decisiones estratégicas en tiempo real; sin embargo, su flexibilidad se ve limitada por la dependencia de parámetros fijos, como el número de paradas en *pits*. Por su parte, la programación dinámica estocástica proporciona la ventaja de ajustarse a eventos aleatorios, aunque no incluye interacciones con las estrategias de otros pilotos. Los modelos estadísticos, en cambio, permiten una mejor comprensión de riesgos y patrones específicos, como la probabilidad de abandono según la posición de partida. En conjunto, estos estudios destacan la importancia de adoptar un enfoque multidimensional para desarrollar una estrategia de carrera ideal en Fórmula 1. Las investigaciones futuras podrían beneficiarse de integrar técnicas de aprendizaje por refuerzo y teoría de juegos, logrando así un modelo con mayor adaptabilidad y realismo competitivo.

1.2.2. En la Fórmula 1

La mayoría de los equipos de Fórmula 1 desarrollan su software de estrategia *in-house* y mantienen un alto nivel de secretismo, ya que esto puede representar una ventaja competitiva. Como resultado, la

información sobre estos programas no es pública. Los equipos de F1 también aprovechan sus alianzas estratégicas con empresas de tecnología, como Darktrace (McLaren) y Oracle (Red Bull Racing), para desarrollar y mejorar estos sistemas.

No obstante, el equipo Mercedes ha utilizado un software desarrollado por Catapult llamado *RaceWatch*, que les permite, a través de vídeos y análisis de datos, estudiar a sus rivales en aspectos como la degradación de neumáticos y la *racing line*. *RaceWatch* combina datos de diversas fuentes, incluyendo cronometraje, posicionamiento GPS, telemetría, mensajes de control de carrera y condiciones meteorológicas. Además, integra comunicaciones de radio y transmisiones en vivo desde cámaras a bordo, de pista, en *pits* y de difusión, proporcionando así una visión integral para la toma de decisiones estratégicas en competencia.

Este sistema también ofrece análisis de rendimiento, permitiendo identificar en qué partes del circuito el piloto pierde tiempo frente a sus rivales. Además, facilita la planificación de *stints* y la estrategia de carrera mediante modelos de aprendizaje automático que optimizan parámetros clave para generar predicciones precisas sobre la estrategia ideal. Este sistema personalizado asegura que se tomen las mejores decisiones posibles durante la carrera, mejorando tanto la planificación táctica como estratégica de la competencia.

Capítulo 2

El motor oculto de la F1: Datos

Aunque los datos a menudo pasan desapercibidos durante las carreras, la Fórmula 1 no podría funcionar sin ellos. Desde sus inicios, cuando los tiempos de vuelta se medían con cronómetro en mano, hasta el actual avance tecnológico que permite extraer y analizar grandes volúmenes de datos, la importancia de estos ha crecido significativamente.

De igual manera, esta tesis no podría avanzar sin datos; por lo tanto, la extracción y creación de nuevas variables que faciliten el análisis posterior resultaron cruciales para el desarrollo de este trabajo.

2.1. Extracción de datos

La Fórmula 1 es un deporte en el que la tecnología y los datos juegan un rol crucial en el desarrollo de una carrera. Durante un fin de semana de competencia, cada coche puede tener más de 250 sensores que generan

información sobre la temperatura, presión, desplazamiento, inercia, etc. Todos estos datos, así como los tiempos de vuelta y resultados, están disponibles a través de la API del Live Timing de la Fórmula 1¹.

En este trabajo de investigación, se utilizó la biblioteca de Python FastF1² para usar la API y descargar los datos. Esta biblioteca proporciona los datos en forma de DataFrames de Pandas, lo que permite su manipulación utilizando todas las herramientas que Pandas ofrece.

Usualmente, el flujo de trabajo con FastF1 comienza con la creación de un objeto de Sesión (`session`) utilizando la función `get_session()`. Esto permite acceder a todos los datos dentro de este objeto, que corresponde usualmente a una sesión específica de un Gran Premio. Por ejemplo, para acceder a los datos de la carrera del Gran Premio de México de 2023, se puede crear el objeto de esta sesión de la siguiente manera:

```
session = fastf1.get_session(2023, "Mexico", "Race")  
  
session.load()
```

A partir de este punto, la sesión ya está creada, y se pueden descargar o analizar todos los datos de la misma. También existen diferentes indicadores para la clasificación y las tres prácticas libres. Los datos disponibles dentro de una sesión incluyen:

Información del evento: Lugar, hora de inicio, fechas, etc.

Resultados: Posiciones, nombres de los equipos y pilotos, estado al finalizar (finished status).

¹Formula 1 (2024a): API oficial de la Fórmula 1

²Schaefer (2024): Documentación de la biblioteca de Python FastF1

Timing Data: Tiempos de vuelta, tiempos por sectores, información sobre los neumáticos y paradas en *pits*. Para este trabajo de investigación, estos son los datos de mayor importancia, ya que aportan la información necesaria para analizar las estrategias de carrera.

Estatus de la pista: Usualmente los status posibles son: bandera verde, bandera amarilla, coche de seguridad, coche de seguridad virtual o bandera roja.

Telemetría: Velocidades, revoluciones por minuto, posición en la pista, cambios de marcha, etc.

Los datos de telemetría y timing data están disponibles desde 2018 en adelante, pero es posible acceder a los datos de resultados desde la temporada de 1950 a través de la API Ergast, otra funcionalidad que ofrece la biblioteca FastF1.

2.2. Creación de variables y limpieza de datos

Se descargaron los datos de todas las carreras de las temporadas de 2019 a 2023. Las variables se dividieron en tres partes importantes para una estrategia de carrera exitosa:

Tiempos de vuelta: Incluyen tiempos de vuelta, tipo de neumático, vida del neumático, posición, número de *stint*, número de vuelta, piloto y equipo.

Información del circuito: Incluye las características del circuito.

Costo de la parada en *pits*: Incluye el tiempo de parada en *pits*, el tiempo de vuelta de la *inlap* y el tiempo de vuelta de la *outlap*.

2.2.1. Tiempos de vuelta

Vueltas rápidas

Se filtraron las vueltas de todos los pilotos mediante el método:

```
pick_quicklaps()
```

Esto garantiza que todas las vueltas sean más rápidas que el 107 % del tiempo de la vuelta más rápida.³ Por ejemplo, si la vuelta más rápida de la carrera es de 1 minuto y 40 segundos (100 segundos), la regla del 107 % establece que cualquier vuelta más lenta que 1 minuto y 47 segundos (107 segundos) no se considera como vuelta rápida. Filtrar las vueltas de esta manera asegura que todas las vueltas son consideradas rápidas, es decir, que no se hayan realizado bajo un coche de seguridad o una bandera amarilla. El objetivo es analizar los tiempos de vuelta de cada piloto en condiciones normales.

Tiempo de vuelta

Se transformó la variable Tiempo de vuelta (*LapTime*) de su estado original que era un pandas.Timedelta⁴ a segundos para facilitar su manipulación.

Vueltas en seco

Un componente muy interesante que puede llegar a ser esencial en la estrategia de carrera es el clima, en especial la lluvia, por lo que la

³Regla del 107% en carrera de la F1: Para que una vuelta rápida durante la carrera sea considerada competitiva, debe estar dentro del 107 % del tiempo de la vuelta más rápida registrada durante la carrera. Si una vuelta es más lenta que este umbral, no se considera competitiva para estos fines.

⁴Tipo de objeto en Pandas que representa una diferencia de tiempo

Fórmula 1 cuenta con neumáticos para estas condiciones.

Para este trabajo de investigación, solo se consideran estrategias de carrera en eventos no afectados por la lluvia, por lo que se filtraron los datos para incluir únicamente vueltas con compuestos de neumáticos para condiciones secas. En el Cuadro 2.1 se muestran los diferentes tipos de neumáticos para dichas condiciones.

| Compuesto | Descripción |
|-----------|------------------|
| HARD | Compuesto duro |
| MEDIUM | Compuesto medio |
| SOFT | Compuesto blando |

Cuadro 2.1. Tipos de compuestos de neumáticos para vueltas en seco en Fórmula 1

Equipos

En la Fórmula 1, es común que los equipos sufran *rebrandings* o cambios de nombre a lo largo de los años debido a diferentes patrocinadores, adquisiciones de equipos, entre otros factores. Esto puede convertirse en un problema, ya que los datos de un mismo equipo pueden estar registrados con nombres diferentes simplemente porque el equipo cambió de patrocinador en un año determinado. Para resolver este problema, se unificaron los nombres de los equipos utilizando las denominaciones oficiales de 2023. Esto permitió mantener la consistencia en los datos y evitar la pérdida de información valiosa sobre los equipos. En el cuadro 2.2 se presentan los nombres de los equipos en temporadas anteriores y los nombres en la temporada 2023 para los equipos que cambiaron de nombre.

| Nombre antiguo | Nombre en 2023 |
|-------------------------------|----------------|
| Toro Rosso (2006-2019) | AlphaTauri |
| Renault (2002-2020) | Alpine |
| Alfa Romeo Racing (2019-2022) | Alfa Romeo |
| Racing Point (2018-2020) | Aston Martin |

Cuadro 2.2. Cambio de nombres de equipos en la Fórmula 1

Longitud del *stint* (*StintLength*)

Un *stint* es el número de vueltas que un piloto realiza con los mismos neumáticos antes de hacer una parada en *pits* o también entre paradas. Para crear la variable *StintLength*, se agrupó la información por número de *stint*, piloto, tipo de neumático, Gran Premio (GP) y año.

Tiempo de vuelta por kilómetro (*LapTimePerKM*)

Dado que no todos los circuitos de la Fórmula 1 tienen la misma longitud en kilómetros, se creó la variable *LapTimePerKM*. Esta variable se calculó dividiendo el tiempo de vuelta por el número de kilómetros del circuito.

Porcentaje de carrera (*RacePercentage*)

Conocer cuántas vueltas faltan en una carrera es crucial para decidir la estrategia de carrera, por lo que se creó la variable *RacePercentage*. Esta variable se calculó dividiendo el número de la vuelta actual entre el número total de vueltas de la carrera.

2.2.2. Costo de la parada en *pits*

El tiempo combinado de la *inlap*, el tiempo transcurrido en *pits* y la *outlap* es lo que los equipos deben considerar como la pérdida total durante una parada en *pits*, y es fundamental para planificar la estrategia de carrera.

Inlaps

Al descargar los datos de las vueltas, no existía un indicador específico para identificar las *inlaps*. No obstante, los datos incluían la variable *PitInTime*, que registra el tiempo exacto en que un coche entra a *pits*. Cualquier vuelta con un valor de *PitInTime* mayor a cero fue clasificada como una *inlap*. Posteriormente, se aplicó un filtro adicional para conservar únicamente las *inlaps* realizadas con neumáticos de compuesto seco. Finalmente, se calculó también el *LapTimePerKM* de cada *inlap*.

Tiempo transcurrido en *pits*

Los datos sobre el tiempo transcurrido en *pits* se obtuvieron a través de la API de Ergast, accesible también mediante la biblioteca *FastF1*. Un desafío común es la inconsistencia en la forma en que se registran los datos entre las API. En Ergast, los tiempos transcurridos en *pits* están asociados al nombre del circuito, en lugar de la ciudad o el país donde se corre la carrera, como ocurre en la API oficial de la Fórmula 1. Para evitar problemas en el análisis y la modelización, se reemplazaron los nombres de los circuitos por los nombres de los países o ciudades (en caso de que haya más de un Gran Premio en el mismo país) y se añadió el equipo del piloto.

Outlaps

Similar a las *inlaps*, no existía un indicador específico para las *outlaps*, pero se pudo utilizar la variable *PitOutTime*, que registra el tiempo en que un coche sale de *pits*, para identificarlas. Se filtraron todas las vueltas con un valor de *PitOutTime* mayor a cero. Después, se aplicó un filtro adicional para mantener únicamente las *outlaps* con neumáticos de compuesto seco, y finalmente se calculó también el *LapTimePerKM* de cada *outlap*.

2.2.3. Coches de seguridad (*Safety Cars*)

Una fuente de incertidumbre que puede afectar el diseño de una estrategia de carrera son los Coches de Seguridad (*Safety Cars*).

El *Safety Car* suele salir a la pista después de una bandera amarilla. Las banderas amarillas se despliegan cuando ocurre un incidente en la pista que representa un peligro, alertando a los pilotos para que reduzcan la velocidad. Estos incidentes pueden variar, desde un coche que se sale de la pista, un choque, hasta la presencia de escombros en la pista.

Una bandera amarilla puede escalar a un *Safety Car* cuando el incidente no puede resolverse rápidamente o representa un peligro significativo. En ese caso, el *Safety Car* entra en la pista y lidera el grupo a una velocidad reducida mientras el incidente se resuelve.

Lo interesante de los *Safety Cars* es que obligan a todos los coches a mantener una velocidad controlada, lo que generalmente incrementa los tiempos de vuelta en un 60 %. Esta reducción de velocidad abre la posibilidad para que los equipos decidan realizar una parada en *pits* durante la presencia del *Safety Car*, sacrificando la posición en pista, pero ganando tiempo en la parada. Si un coche tiene una ventaja considerable respecto al coche que le sigue, puede beneficiarse de una

parada en *pits* “gratis”, lo que significa que puede entrar a *pits* y regresar a la pista sin perder posiciones, ya que el tiempo perdido en *pits* disminuye debido a la velocidad reducida de los demás coches.

| Etiqueta | Descripción |
|---------------------------|----------------------------------|
| <i>Green Flag</i> | <i>TrackStatus</i> contiene al 1 |
| <i>Yellow Flag</i> | <i>TrackStatus</i> contiene al 2 |
| <i>Safety Car</i> | <i>TrackStatus</i> contiene al 4 |
| <i>Red Flag</i> | <i>TrackStatus</i> contiene al 5 |
| <i>Virtual Safety Car</i> | <i>TrackStatus</i> contiene al 6 |

Cuadro 2.3. Descripción de las etiquetas de estatus de la pista

En los datos que se tienen, cada registro de vuelta incluye la variable *TrackStatus*, que indica el estado de la pista en esa vuelta. En el Cuadro 2.3 se muestran las etiquetas construidas a partir de *TrackStatus* y el proceso seguido para crearlas.

Por último se creó una variable indicadora de *Safety Car*, para identificar las vueltas en donde salió el *Safety Car* más fácil.

2.3. Base de datos final

Después de crear nuevas variables y realizar la limpieza de los datos, se generaron cinco conjuntos de datos finales. El principal contiene todas las vueltas rápidas en condiciones de pista seca. Luego las *inlaps* y las *outlaps* se guardaron en archivos diferentes para diferenciarlas de las vueltas normales pero estos conjuntos de datos contienen las mismas variables que el conjunto de las vueltas rápidas en seco. En el Cuadro 2.4 se describen las variables incluidas en estos conjuntos de datos.

| Variable | Descripción |
|-----------------------|--|
| <i>Driver</i> | Piloto que realizó la vuelta |
| <i>Team</i> | Equipo al que pertenece el piloto |
| <i>LapNumber</i> | Número de la vuelta en la carrera |
| <i>LapTime</i> | Tiempo registrado en la vuelta |
| <i>Stint</i> | Número de <i>stint</i> |
| <i>Compound</i> | Tipo de neumático utilizado |
| <i>TyreLife</i> | Edad de los neumáticos, medida en el número de vueltas que ha completado |
| <i>Position</i> | Posición del piloto al completar esa vuelta |
| <i>Year</i> | Año |
| <i>Circuit</i> | Circuito |
| <i>Length</i> | Longitud del circuito |
| <i>LapTimePerKM</i> | Tiempo de vuelta por kilómetro |
| <i>Laps</i> | Número de vueltas de la carrera |
| <i>RacePercentage</i> | Porcentaje completado de la carrera |

Cuadro 2.4. Descripción de variables en el conjunto de datos de vueltas rápidas

Se tiene otro conjunto de datos de *pitstops*. En el Cuadro 2.5 se detallan las variables de este conjunto de datos.

| Variable | Descripción |
|-----------------|--|
| <i>Circuit</i> | Circuito donde se realizó la parada en <i>pits</i> . |
| <i>PitstopT</i> | Tiempo transcurrido durante la parada en <i>pits</i> . |
| <i>Driver</i> | Piloto que realizó la parada en <i>pits</i> . |
| <i>Year</i> | Año en el que se registró la parada. |
| <i>Team</i> | Equipo al que pertenece el piloto que realizó la parada. |

Cuadro 2.5. Descripción de variables del conjunto de datos de *pitstops*.

Por último, se tiene el conjunto de datos de los *Safety Cars*. En el Cuadro 2.6 se describen las variables que contiene este conjunto de datos.

| Variable | Descripción |
|--------------------|--|
| <i>Circuit</i> | Circuito |
| <i>LapNumber</i> | Número de la vuelta en la carrera |
| <i>Year</i> | Año |
| <i>TrackStatus</i> | Indicador del estatus de la pista en esa vuelta |
| <i>Label</i> | Etiqueta asociada al número de <i>TrackStatus</i> |
| <i>SafetyCar</i> | Indicador de si hubo un <i>Safety Car</i> en esa vuelta (1 = sí, 0 = no) |
| <i>Risk</i> | Categoría de riesgo de que un <i>Safety Car</i> salga en ese circuito. |

Cuadro 2.6. Descripción de variables del conjunto de datos de *Safety Cars*.

Capítulo 3

Vueltas de reconocimiento: Exploración de Datos

Antes de estacionar los coches en la parrilla de salida para la carrera, los pilotos realizan las llamadas vueltas de reconocimiento mientras se dirigen a sus posiciones de salida. Estas vueltas permiten evaluar las condiciones de la pista y hacer los últimos ajustes antes de iniciar la carrera.

De la misma manera, la exploración de datos permite familiarizarse con el conjunto de datos, encontrar patrones y comprender su estructura antes de proceder a la estimación de modelos. Así como las vueltas de reconocimiento preparan a los pilotos para la carrera, la exploración de datos sienta las bases para un análisis sólido y efectivo.

3.1. Vueltas Rápidas en Seco y Neumáticos

Como se mencionó anteriormente, solo se incluyeron en los datos para el modelo las vueltas rápidas realizadas con compuestos de neumáticos

para condiciones secas. Esto resulta en un total de 64,516 registros de vueltas.

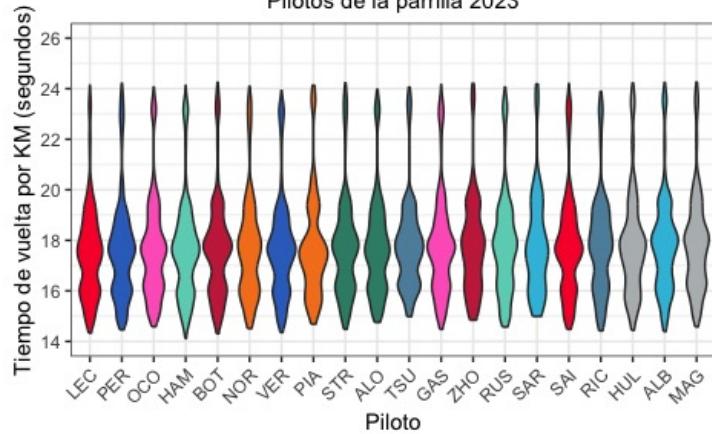
El objetivo de esta tesis es estimar el tiempo esperado de una estrategia determinada para seleccionar la estrategia óptima. Para ello, es necesario calcular el tiempo de vuelta por kilómetro, lo que permite estimar el tiempo esperado de cada *stint*. Por lo tanto, se busca identificar qué variables pueden ser buenos predictores para esta estimación.

3.1.1. Tiempos de vuelta por kilómetro por Piloto y Equipo

En la Figura 3.1, los diagramas de violín están ordenados según el tiempo de vuelta por kilómetro promedio. La primera gráfica muestra la distribución de los tiempos de vuelta por kilómetro por piloto, etiquetados con la abreviación oficial del apellido de cada uno. En la segunda gráfica se presenta la distribución de los tiempos de vuelta por kilómetro por equipo. Se pueden observar diferencias en los tiempos de vuelta por kilómetro tanto entre equipos como entre pilotos. Nico Rosberg, campeón de la Fórmula 1 en 2016, ha mencionado en varias ocasiones su regla del 80-20. Según Rosberg, el 80 % del éxito en Fórmula 1 se debe al coche y al equipo, mientras que el 20 % se debe a las habilidades del piloto. La validez de estos porcentajes no es el foco de este proyecto de investigación; lo relevante es que tanto el equipo como el piloto pueden influir de manera significativa en los tiempos de vuelta por kilómetro.

Distribución de tiempos de vuelta por piloto (2019-2023)

Pilotos de la parrilla 2023



Distribución de tiempos de vuelta por equipo

(2019-2023)

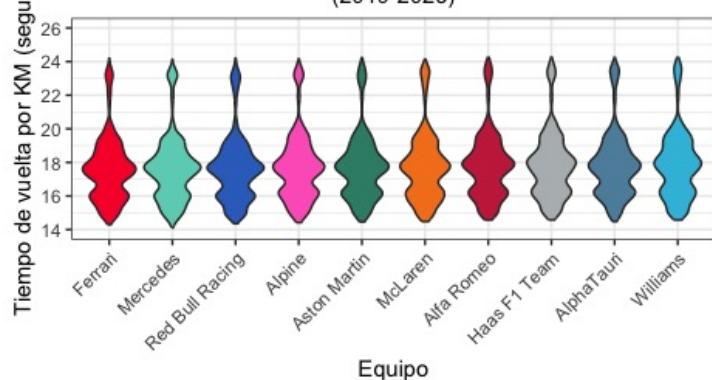


Figura 3.1. Tiempos de vuelta por KM por piloto y por equipo

3.1.2. Neumáticos

En la Figura 3.2 se puede observar el comportamiento de cada tipo de neumático. Se puede observar una clara diferencia en el tiempo de vuelta por kilómetro entre los distintos compuestos. El compuesto *Medium* es, en promedio, el más rápido durante la carrera, a pesar de que, teóricamente, el compuesto *Soft* debería ser el más veloz a una vuelta. Sin embargo, como se muestra en tercera gráfica, el *Soft* tiene la vida útil promedio más corta debido a su rápida degradación, lo cual explica por qué es el segundo compuesto más rápido en carrera. El compuesto más utilizado es el *Hard*, lo cual es comprensible dado que es el más duradero. Por todo esto, el tipo de llanta podría tener un efecto significativo sobre los tiempos de vuelta por kilómetro.

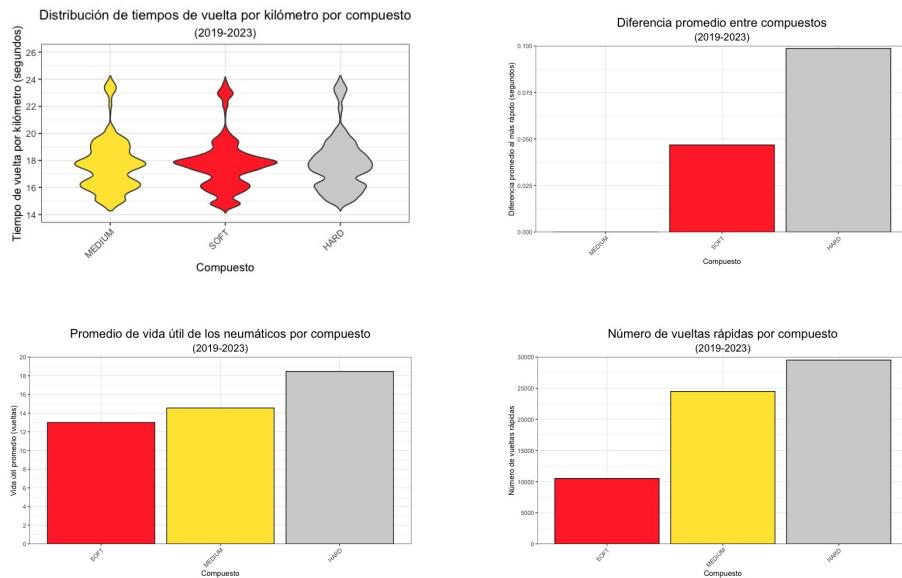


Figura 3.2. Comportamiento del tipo de neumáticos

Las correlaciones entre pares de predictores son, en su mayoría, muy bajas. Las únicas correlaciones que resultan ligeramente más altas son entre *TyreLife* y *RacePercentage*, así como entre *Stint* y *RacePercentage*. Estas relaciones son evidentes, ya que, conforme aumenta el porcentaje completado de la carrera, también incrementa la edad de los neumáticos y el número de *stint*. Esto podría indicar que cada uno de los predictores aporta información distinta sobre el tiempo de vuelta por kilómetro y, por ende, podrían ser útiles para la estimación del modelo.

3.2. Circuitos y pitstops

Otra variable que debe estimarse para cumplir con el objetivo es el tiempo transcurrido en *pits*. Por ello, también se busca identificar qué variables pueden ser buenos predictores para esta estimación.

| Circuito | Longitud (km) | Circuito | Longitud (km) |
|------------|---------------|---------------|---------------|
| Monaco | 3.337 | Netherlands | 4.259 |
| Mexico | 4.304 | Brazil | 4.309 |
| Austria | 4.318 | Canada | 4.361 |
| Hungary | 4.381 | Imola | 4.909 |
| Singapore | 4.940 | Australia | 5.278 |
| Miami | 5.412 | Bahrain | 5.412 |
| Qatar | 5.419 | China | 5.451 |
| Austin | 5.513 | Monza | 5.793 |
| Japan | 5.807 | Great Britain | 5.891 |
| Azerbaijan | 6.003 | Saudi Arabia | 6.174 |
| Belgium | 7.004 | | |

Cuadro 3.1. Longitud de los circuitos

Cada circuito en el calendario es diferente; algunos son más largos, otros más rápidos, y todo esto influye en el tiempo de vuelta y la vida útil de los neumáticos. Además, las diferentes características de los pitlanes pueden afectar el tiempo perdido en pits.

En el Cuadro 3.1 están las longitudes de los circuitos actuales de la Fórmula 1. El más corto es Mónaco y el más largo es Bélgica. Mientras más corto sea el circuito, más vueltas se dan en la carrera.

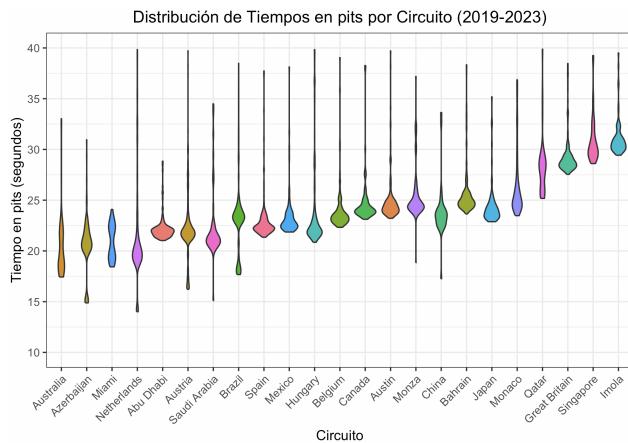


Figura 3.3. Tiempos en *pits* por circuito

Como se puede observar en la Figura 3.3, gracias a las diferentes características de los *pitlanes* de los circuitos, se notan grandes diferencias en el tiempo perdido en *pits*. Australia es el circuito con menor tiempo promedio perdido, mientras que Imola tiene el mayor.

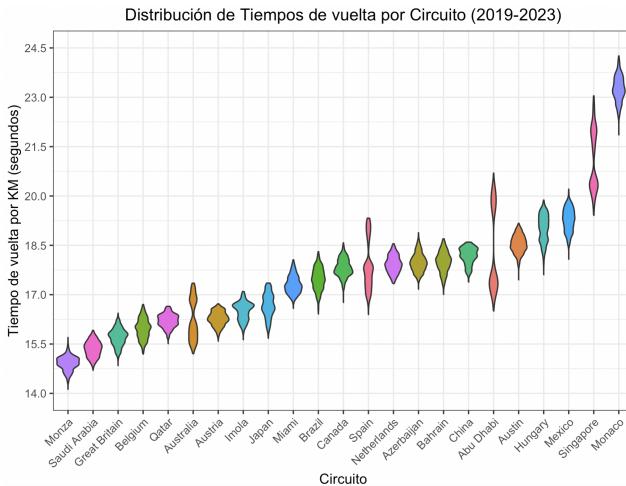


Figura 3.4. Tiempos de vuelta por KM por circuito

En la Figura 3.4, también se puede observar una diferencia notable en el tiempo de vuelta por KM por circuito. Monza, apodado *el templo de la velocidad*, hace honor a su apodo, teniendo el tiempo por vuelta por KM más rápido en promedio, mientras que Mónaco, aunque es legendario en la Fórmula 1, es el circuito más lento en promedio.

Por todo esto, el circuito podría tener un efecto significativo sobre los tiempos de vuelta por kilómetro y los tiempos en pits.

3.3. Safety Cars

La estrategia de carrera también requiere adaptabilidad ante eventos inciertos, como la salida de un *Safety Car*. Por ello, resulta interesante estimar la probabilidad de que un *Safety Car* salga en una vuelta durante la carrera, lo que implica identificar qué variables podrían ser buenos predictores para esta estimación.

En la Figura 3.5 se puede observar la distribución de los *Safety Cars*

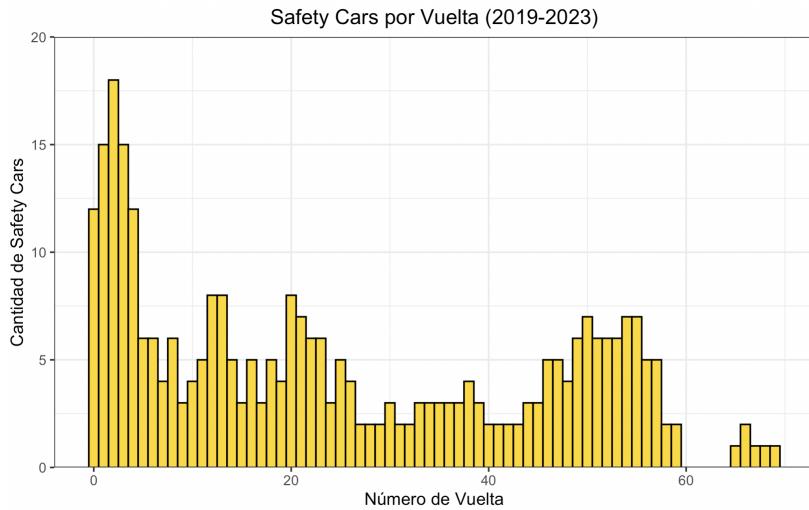


Figura 3.5. Safety Cars por vuelta

por número de vuelta en los últimos 5 años. El mayor número de *Safety Cars* se da en las primeras vueltas; esto se debe generalmente a que los coches están muy juntos y luchan por posición, lo que conlleva a incidentes. Hacia el final de la carrera también se producen varios *Safety Cars* por la misma razón. Sin embargo, en la mitad de la carrera, la situación se estabiliza y no hay muchas peleas cercanas, por lo que no se registran muchos *Safety Cars*.

Cuadro 3.2. Riesgo según el porcentaje de ocurrencia

| Rango de porcentaje (%) | Categoría de Riesgo |
|-------------------------|---------------------|
| 0 - 30 | Riesgo Bajo |
| 30 - 70 | Riesgo Medio |
| 70 - 100 | Riesgo Alto |

En el Cuadro 3.3 se muestra el porcentaje de ocurrencia de *safety*

Cuadro 3.3. Porcentaje de Ocurrencia y Categoría de Riesgo por Circuito (2019-2023)

| Circuito | Ocurrencia (%) | Riesgo |
|-----------------|-----------------------|---------------|
| Brazil | 100.00 | Riesgo Alto |
| Great Britain | 100.00 | Riesgo Alto |
| Imola | 100.00 | Riesgo Alto |
| Saudi Arabia | 100.00 | Riesgo Alto |
| Singapore | 100.00 | Riesgo Alto |
| Bahrain | 80.00 | Riesgo Alto |
| Canada | 75.00 | Riesgo Alto |
| Australia | 66.67 | Riesgo Medio |
| Japan | 66.67 | Riesgo Medio |
| Netherlands | 66.67 | Riesgo Medio |
| Belgium | 60.00 | Riesgo Medio |
| Monaco | 60.00 | Riesgo Medio |
| Monza | 60.00 | Riesgo Medio |
| Azerbaijan | 50.00 | Riesgo Medio |
| Mexico | 50.00 | Riesgo Medio |
| Miami | 50.00 | Riesgo Medio |
| Qatar | 50.00 | Riesgo Medio |
| Abu Dhabi | 40.00 | Riesgo Medio |
| Austria | 40.00 | Riesgo Medio |
| Spain | 40.00 | Riesgo Medio |
| Austin | 25.00 | Riesgo Bajo |
| Hungary | 20.00 | Riesgo Bajo |
| China | 0.00 | Riesgo Bajo |

cars por circuito. Además los circuitos se categorizaron de acuerdo a las categorías en el Cuadro 3.2. Las diferencias en los porcentajes de ocurrencia entre los circuitos podrían influir en la estimación de la probabilidad de que ocurra un *Safety Car*.

Capítulo 4

El plan perfecto para la victoria: Modelos

Los ingenieros de estrategia dedican largas horas la tarde del sábado y la mañana del domingo a diseñar la estrategia óptima que maximice los puntos para los pilotos al final de la carrera. Si bien los imprevistos pueden alterar el plan inicial, es fundamental contar con una estrategia bien definida que conduzca a la victoria.

Para alcanzar este objetivo, es esencial utilizar herramientas de estimación que permitan evaluar y comparar las diferentes estrategias. En el contexto de esta tesis, estas herramientas son los modelos lineales generalizados.

4.1. Modelos lineales generalizados

Los modelos lineales generalizados, al igual que los modelos lineales clásicos, tienen como objetivo expresar la relación entre una variable dependiente Y y variables explicativas $\mathbf{X} = (X_1, X_2, \dots, X_n)$.

La principal ventaja de los modelos lineales generalizados sobre los modelos lineales clásicos es que no requieren que la variable dependiente siga una distribución normal. En su lugar, la variable dependiente puede seguir cualquier distribución de la familia exponencial. Además, estos modelos incorporan una función liga, la cual permite modelar relaciones no lineales dentro de un marco lineal, lo que hace que el análisis sea más flexible en comparación con el enfoque clásico.

Los modelos de regresión lineal clásicos están restringidos a datos que cumplen los supuestos de linealidad, normalidad, homocedasticidad e independencia.

Linealidad: En los modelos clásicos, la relación entre Y y las variables explicativas \mathbf{X} es lineal. En los modelos lineales generalizados, esta relación se modela a través de una función de enlace, lo que permite que la relación no sea necesariamente lineal, ya que la función de enlace transforma la media de Y .

Normalidad: En los modelos clásicos, los residuos (la diferencia entre el valor observado y el estimado) deben seguir una distribución normal. En los modelos lineales generalizados, no es necesario que los residuos se distribuyan normalmente, ya que la distribución de Y puede pertenecer a la familia exponencial, sin ser necesariamente normal.

Homocedasticidad: En los modelos clásicos, la varianza de los residuos debe ser constante. En los modelos lineales generalizados, la varianza de los residuos puede no ser constante, y puede depender de la media de la variable dependiente.

Independencia: Los residuos deben ser independientes entre sí. La independencia entre observaciones también es necesaria para los modelos lineales generalizados.

De acuerdo con (Dunn et al., 2018, 211-229) los modelos lineales generalizados tienen 2 componentes principales:

Distribución de los errores (Familia): Se refiere a la distribución de la variable dependiente Y . La elección de esta distribución está guiada por la naturaleza de la variable dependiente, que puede ser continua, binaria, de conteo, etc. Esta distribución debe pertenecer a la familia exponencial.

Función de enlace: La función de enlace ($g()$) conecta el predictor lineal (η), que es una combinación lineal de las variables independientes, con la media de la variable dependiente, $\mu = \mathbb{E}(Y)$, de manera que $g(\mu) = \eta = X\beta$, donde X es la matriz de las variables independientes y β es el vector de parámetros. El predictor lineal se define como:

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n, \quad (4.1)$$

donde:

- η es el predictor lineal,
- β_0 es el intercepto del modelo,
- β_i son los coeficientes asociados a cada variable independiente X_i ,
- X_i son las variables independientes, donde $i = 1, 2, \dots, n$.

La función de enlace es monótona y diferenciable, lo que asegura que cada valor de η se relacione de manera única con un valor de μ . Su propósito es transformar el valor esperado de la variable dependiente a la escala del predictor lineal.

4.1.1. Estimación y ajuste del modelo

Los modelos lineales generalizados emplean el método de máxima verosimilitud para estimar sus parámetros (β). Este método busca los valores de los parámetros que maximizan la función de verosimilitud,

es decir, la probabilidad de observar los datos bajo el modelo propuesto. La forma de la función de verosimilitud depende de la distribución seleccionada para la variable dependiente. Maximizar la log-verosimilitud, que es el logaritmo de la función de verosimilitud, resulta más práctico debido a la naturaleza aditiva del logaritmo.

Dado que se puede despejar μ de la relación mencionada anteriormente ($g(\mu) = \eta = X\beta$), el modelo general se expresa como:

$$Y = g^{-1}(X\beta) + \varepsilon, \quad (4.2)$$

donde:

- Y es la variable dependiente,
- g es la función de enlace,
- X es la matriz de variables independientes,
- β es el vector de parámetros,
- ε es el vector de errores.

4.1.2. Evaluación del ajuste y selección de modelos

Determinar si el modelo propuesto representa adecuadamente los datos es crucial para garantizar un buen desempeño predictivo. Asimismo, es fundamental comparar diferentes modelos para seleccionar aquel que mejor se ajuste a los datos y proporcione una mejor capacidad de predicción.

Bondad de Ajuste

Para evaluar la bondad de ajuste en los modelos lineales generalizados se utiliza la **devianza**, que es la diferencia entre la

log-verosimilitud del modelo ajustado y la de un modelo saturado con un ajuste perfecto.

Un modelo saturado tiene la misma distribución de la variable dependiente que el modelo propuesto y la misma función de enlace. En el modelo saturado se estima un parámetro para cada observación, por lo que tiene un ajuste perfecto.

La log-verosimilitud del modelo saturado es siempre estrictamente mayor que la del modelo propuesto. Cuanto mejor sea el ajuste del modelo propuesto, más se parecerán entre ellas. Por lo tanto, el estadístico conocido como devianza que está dado por:

$$D = 2 [l(\boldsymbol{\beta}_{\max}; y) - l(\boldsymbol{\beta}; y)], \quad (4.3)$$

donde:

- D es la devianza,
- $l(\boldsymbol{\beta}_{\max}; y)$ es la log verosimilitud del modelo completo, es decir el modelo que tiene el mismo número de parámetros como de variables,
- $l(\boldsymbol{\beta}; y)$ es la log verosimilitud del modelo estimado.

Debe ser lo más cercano a cero y, por lo tanto, si el modelo tiene un ajuste adecuado, D se distribuirá como $D \sim \chi^2_{n-p}$, donde n es el número de parámetros en el modelo saturado y p es el número de parámetros en el modelo propuesto. Una devianza baja indica un buen ajuste.

Comparación entre modelos

El **Criterio de Información de Akaike** (AIC) es útil para comparar modelos estadísticos, ya que busca balancear la bondad de

ajuste con la complejidad del modelo penalizando aquellos con un alto número de parámetros. En general, se prefiere el modelo con el menor valor de AIC. Sin embargo, el valor de AIC en sí mismo no tiene un significado; solo es útil para la comparación entre modelos. El AIC se define como:

$$\text{AIC} = -2 \ln(L) + 2k, \quad (4.4)$$

donde L es la verosimilitud del modelo y k es el número de parámetros en el modelo propuesto.

El **Criterio de Información Bayesiano** (BIC), similar al AIC, también busca balancear la bondad de ajuste y la complejidad del modelo, pero penaliza el número de parámetros de manera más estricta. Esto lo lleva a seleccionar modelos más parsimoniosos, es decir, más simples que aquellos que tiende a elegir el AIC. Al igual que con el AIC, se prefieren los modelos con menores valores de BIC. El BIC se define como:

$$\text{BIC} = -2 \ln(L) + k \ln(n), \quad (4.5)$$

donde:

- L es la verosimilitud del modelo propuesto,
- k es el número de parámetros en el modelo propuesto,
- n es el número de observaciones o datos.

Según (Dunn et al., 2018, 71), el enfoque del AIC es crear un modelo orientado a hacer buenas predicciones, permitiendo la inclusión de variables explicativas adicionales incluso si la evidencia de su importancia no es concluyente. En cambio, el BIC exige una evidencia

más fuerte para añadir variables adicionales, por lo que selecciona modelos más simples con interpretaciones más sencillas. De esta forma, el AIC se enfoca principalmente en la predicción, mientras que el BIC busca un equilibrio entre interpretación y predicción.

Por esta razón, en esta tesis se utiliza el AIC como criterio principal de comparación entre modelos, ya que, conforme al objetivo de esta investigación, se requieren predicciones de alta precisión. Sin embargo, en una etapa del proceso de selección de modelos se emplea la función `bestglm`, la que utiliza el criterio BIC.

También se puede utilizar la **devianza** para comparar modelos, pero los modelos comparados deben ser anidados, es decir, deben tener la misma distribución para los errores y utilizar la misma función de enlace, solo difieren en el número de variables independientes.

4.2. *LapTimePerKM*

El primer modelo lineal generalizado empleado para determinar la estrategia óptima es el de *LapTimePerKM*. La variable dependiente es el tiempo de vuelta por kilómetro, lo que resulta útil, ya que los circuitos de Fórmula 1 tienen longitudes distintas. Para estimar el tiempo de vuelta en un circuito determinado, basta con multiplicar el número de kilómetros del circuito por el tiempo de vuelta por kilómetro estimado.

4.2.1. Selección de variables y modelo

Se seleccionaron las variables con posible impacto en el *LapTimePerKM* según el análisis exploratorio de datos (EDA). Posteriormente, se utilizó la función `bestglm` de la biblioteca de R `bestglm`, que compara todas las combinaciones posibles de esas variables y selecciona el mejor modelo utilizando el criterio BIC.

La función `bestglm` se aplicó para tres familias de distribuciones de errores con sus respectivas funciones de enlace: la distribución Gamma con función de enlace inversa, la Normal con enlace logarítmico, y la Gaussiana Inversa¹. Estas distribuciones se eligieron porque se ajustan al soporte de los tiempos de vuelta, que son positivos. Los resultados de la comparación fueron los siguientes:

| | Normal | Gamma | Gaussiana Inversa |
|-------------------|-----------|---------|-------------------|
| Función de enlace | Logaritmo | Inversa | Inversa cuadrada |
| AIC | 85009 | 79833 | 78114 |

Cuadro 4.1. Comparación de diferentes familias de distribución para el modelo *LapTimePerKM*

Se utilizó el criterio de Akaike (AIC) para seleccionar el mejor modelo. Como se muestra en el Cuadro 4.1, el modelo con el menor AIC fue el de la Gaussiana Inversa, por lo que se eligió este modelo, que se especifica a continuación:

¹De acuerdo con (Dunn et al., 2018, 431-432), la distribución Gaussiana inversa, similar a la distribución Gamma, es útil para modelar datos continuos positivos, especialmente cuando los datos presentan un sesgo aún mayor que el que típicamente ajustaría una distribución Gamma. Además, esta distribución está relacionada con el movimiento Browniano (movimiento aleatorio de partículas a través del tiempo): describe el tiempo necesario para que una partícula alcance un punto a una distancia positiva fija δ . En contraste, la distribución normal (o Gaussiana) describe la distancia de la partícula al origen en un tiempo fijo. La distribución Gaussiana inversa recibe su nombre debido a esta relación con la distribución normal.

$$\begin{aligned}
LapTimePerKM = g^{-1}(&\beta_0 + \beta_1 Circuit + \beta_2 RacePercentage + \beta_3 Driver \\
&+ \beta_4 Team + \beta_5 TyreLife + \beta_6 Compound \\
&+ \beta_7 Position + \beta_8 Stint) + \varepsilon
\end{aligned} \tag{4.6}$$

4.2.2. Resultados

Cuadro 4.2. *LapTimePerKM - Intercepto y otras variables*

| Variable | Coeficiente |
|-------------------|---------------------------|
| TyreLife | -1.421e-06*** (1.046e-07) |
| MEDIUM | -0.00003*** (0.00000) |
| SOFT | -0.00005*** (0.00000) |
| Position | -3.895e-06*** (2.327e-07) |
| Stint | -0.00001*** (0.00000) |
| Constant | 0.003*** (0.00001) |
| Observaciones | 67,265 |
| Akaike Inf. Crit. | 108,231.600 |

Nota: *p<0.1; **p<0.05; ***p<0.01

En el Cuadro 4.2 se presentan los resultados de la estimación del modelo, excluyendo las variables de equipos, circuitos y pilotos. Se observa que todos los coeficientes son significativos, lo que tiene sentido según el análisis exploratorio de datos (EDA), ya que existen diferencias en los compuestos de neumáticos, la vida útil de los mismos y la posición, que pueden influir debido a factores como aire sucio, aire limpio o tráfico.

Cuadro 4.3. *LapTimePerKM* - Equipos

| Equipo | Coeficiente |
|-----------------|----------------------|
| AlphaTauri | 0.00001 (0.00001) |
| Alpine | 0.00003** (0.00001) |
| Aston Martin | 0.0001*** (0.00001) |
| Ferrari | 0.0001*** (0.00001) |
| Haas F1 Team | 0.0001*** (0.00002) |
| McLaren | 0.00002 (0.00001) |
| Mercedes | 0.0001*** (0.00001) |
| Red Bull Racing | 0.0001*** (0.00001) |
| Williams | 0.00003*** (0.00001) |

En el Cuadro 4.3 se presentan las estimaciones de los coeficientes de los equipos, tomando a Alfa Romeo como el equipo de referencia. La mayoría de los coeficientes son significativos, lo que indica que estos pueden influir en los tiempos de vuelta por km.

El Cuadro 4.4 muestra las estimaciones de los coeficientes de los circuitos, con Abu Dhabi como circuito de referencia. La mayoría de los circuitos son significativos, excepto el de Netherlands (Países Bajos).

En el Cuadro 4.5 se encuentran las estimaciones de los coeficientes por piloto, con Albon (ALB) como el piloto de referencia. En general, se puede observar como los pilotos sí pueden influir en el tiempo de vuelta por kilómetro.

Cuadro 4.4. *LapTimePerKM* - Circuitos

| Circuito | Coeficiente | Circuito | Coeficiente |
|-----------|----------------------|---------------|----------------------|
| Austin | -0.0001*** (0.00000) | Austria | 0.001*** (0.00000) |
| Australia | 0.001*** (0.00001) | Azerbaijan | 0.0001*** (0.00000) |
| Bahrain | 0.0001*** (0.00000) | Belgium | 0.001*** (0.00001) |
| Brazil | 0.0003*** (0.00001) | Canada | 0.0002*** (0.00000) |
| China | 0.00004*** (0.00001) | Great Britain | 0.001*** (0.00001) |
| Hungary | -0.0002*** (0.00000) | Imola | 0.001*** (0.00001) |
| Japan | 0.001*** (0.00001) | Mexico | -0.0003*** (0.00000) |
| Miami | 0.0003*** (0.00001) | Monaco | -0.001*** (0.00000) |
| Monza | 0.002*** (0.00001) | Netherlands | -0.00001 (0.00000) |
| Qatar | 0.001*** (0.00001) | Saudi Arabia | 0.001*** (0.00001) |
| Singapore | -0.001*** (0.00000) | Spain | 0.0001*** (0.00000) |

4.3. *Pitstops*

Estimar el tiempo perdido en el *pitlane*, es decir, desde que un coche cruza la línea de entrada a los *pits*, se detiene en su *box* para cambiar llantas y sale del *pitlane*, es crucial para determinar la estrategia óptima. Este tiempo representa un periodo en el que los pilotos no están en la pista, lo que podría resultar en la pérdida de posiciones.

4.3.1. Selección de variables y modelo

De manera similar al modelo anterior, se utilizó la función `bestglm` para comparar modelos con diferentes familias de distribuciones. Los resultados obtenidos fueron los siguientes:

El modelo con el menor AIC, como se observa en el Cuadro 4.6, es

Cuadro 4.5. *LapTimePerKM* - Pilotos

| Piloto | Coeficiente | Piloto | Coeficiente |
|--------|-----------------------|--------|-----------------------|
| ALO | 0.00002* (0.00001) | LAW | 0.0001*** (0.00002) |
| BOT | 0.00003*** (0.00001) | LEC | 0.00000 (0.00001) |
| DEV | 0.0001*** (0.00001) | MAG | -0.0001*** (0.00001) |
| FIT | -0.001*** (0.00003) | MAZ | -0.0001*** (0.00002) |
| GAS | 0.00003*** (0.00001) | MSC | -0.0001*** (0.00001) |
| GIO | 0.00004*** (0.00001) | NOR | 0.00003*** (0.00001) |
| GRO | -0.0001*** (0.00001) | OCO | 0.00001 (0.00001) |
| HAM | 0.00001 (0.00001) | PER | 0.00002** (0.00001) |
| HUL | -0.00002 (0.00001) | PIA | 0.0001*** (0.00001) |
| KUB | -0.00005*** (0.00001) | RAI | 0.00004*** (0.00001) |
| KVY | 0.00000 (0.00001) | RIC | 0.00001 (0.00001) |
| LAT | -0.00003*** (0.00001) | RUS | -0.00002*** (0.00001) |
| VER | 0.00003*** (0.00001) | SAI | 0.00001 (0.00001) |
| STR | -0.00004*** (0.00001) | SCH | -0.00002* (0.00001) |
| TSU | -0.00002 (0.00001) | ZHO | 0.0001*** (0.00001) |

| | Normal | Gamma | Gaussiana Inversa |
|-------------------|-----------|---------|-------------------|
| Función de enlace | Logaritmo | Inversa | Inversa cuadrada |
| AIC | 18468 | 17675 | 17469 |

Cuadro 4.6. Comparación de diferentes familias de distribución para el modelo de *Pitstops*

nuevamente el de la Gaussiana Inversa, por lo que se procedió con dicho modelo, cuya especificación es la siguiente:

$$Pitstop T = g^{-1}(\beta_0 + \beta_1 Circuit) + \varepsilon \quad (4.7)$$

4.3.2. Resultados

Los Cuadros 4.7 y 4.8 muestran los resultados de la estimación del modelo.

Cuadro 4.7. *Pitstop Time - Intercepto*

| Variable | Coeficiente |
|-------------------|-------------------|
| Constante | 0.002*** (0.0001) |
| Observaciones | 3,330 |
| Akaike Inf. Crit. | 17,469.140 |

Nota: *p<0.1; **p<0.05; ***p<0.01

Como se puede observar en el Cuadro 4.8, los circuitos influyen en la variabilidad del tiempo transcurrido en los *pits*, ya que tanto la longitud como las características del pitlane varían en cada uno de ellos. Estos factores pueden hacer que el tiempo de parada sea mayor o menor. Por ejemplo, el circuito de Interlagos en Brasil tiene uno de los pitlanes

Cuadro 4.8. *Pitstop Time* - Circuitos

| Circuito | Coeficiente | Circuito | Coeficiente |
|---------------|---------------------|------------|---------------------|
| Austin | -0.0004*** (0.0001) | Australia | 0.0003*** (0.0001) |
| Austria | -0.00001 (0.0001) | Azerbaijan | 0.0003*** (0.0001) |
| Bahrain | -0.001*** (0.0001) | Belgium | -0.0003*** (0.0001) |
| Brazil | -0.0001** (0.0001) | Canada | -0.0004*** (0.0001) |
| China | -0.001*** (0.0001) | Eifel | -0.0002* (0.0001) |
| France | -0.001*** (0.0001) | Germany | 0.0001* (0.0001) |
| Great Britain | -0.001*** (0.0001) | Hungary | -0.0003*** (0.0001) |
| Imola | -0.001*** (0.0001) | Japan | -0.001*** (0.0001) |
| Las Vegas | -0.0002** (0.0001) | Mexico | -0.0002*** (0.0001) |
| Miami | 0.0001 (0.0001) | Monaco | -0.001*** (0.0001) |
| Monza | -0.0005*** (0.0001) | Mugello | -0.00001 (0.0001) |
| Netherlands | 0.0001 (0.0001) | Portugal | -0.0004*** (0.0001) |
| Qatar | -0.001*** (0.0001) | Russia | -0.001*** (0.0001) |
| Saudi Arabia | -0.00002 (0.0001) | Singapore | -0.001*** (0.0001) |
| Spain | -0.0001** (0.0001) | Turkey | -0.0004*** (0.0001) |

más largos del calendario, con una salida que incluye una curva muy prolongada. En contraste, el circuito de Zandvoort en los Países Bajos se caracteriza por un pitlane estrecho, mientras que en Yas Marina, Abu Dhabi, los autos atraviesan un túnel al salir de los *pits*. Todos estos elementos contribuyen a las diferencias en el tiempo perdido en las paradas.

4.4. *Inlaps*

Aunque el objetivo sigue siendo estimar el tiempo de vuelta por kilómetro, en este caso se utilizarán datos de las *inlaps*. Las *inlaps*, al ser las vueltas en las que los pilotos ingresan a *pits*, son cruciales para la estrategia de carrera. Al igual que las *outlaps*, pueden determinar si un *undercut* fue exitoso o no.

4.4.1. Selección de variables y modelo

Se empleó el mismo procedimiento que en los modelos anteriores, utilizando la función `bestglm` para comparar distintas familias de distribuciones. Los resultados obtenidos fueron los siguientes:

| | Normal | Gamma | Gaussiana Inversa |
|-------------------|-----------|---------|-------------------|
| Función de enlace | Logaritmo | Inversa | Inversa cuadrada |
| AIC | 12954 | 12218 | 12053 |

Cuadro 4.9. Comparación de diferentes familias de distribución para el modelo de *Inlaps*

De acuerdo con el AIC, como se puede observar en el Cuadro 4.9, el mejor modelo es el de la Gaussiana Inversa, cuya especificación es la siguiente:

$$LapTimePerKM = g^{-1}(\beta_0 + \beta_1 Circuit + \beta_2 Compound + \beta_3 TyreLife + \beta_4 Stint) + \varepsilon \quad (4.8)$$

4.4.2. Resultados

Cuadro 4.10. *LapTimePerKM-Inlaps* Intercepto y otras variables

| Variable | Coeficiente |
|-------------------|----------------------|
| MEDIUM | 0.0002*** (0.00003) |
| SOFT | 0.0001*** (0.00003) |
| TyreLife | 0.00002*** (0.00000) |
| Stint | 0.00003** (0.00001) |
| Constant | 0.002*** (0.0001) |
| Observaciones | 2,698 |
| Akaike Inf. Crit. | 12,052.580 |

Note: *p<0.1; **p<0.05; ***p<0.01

Como se puede observar en el Cuadro 4.10, la vida útil de los neumáticos, el tipo de compuesto y el número de *stint* también influyen en el tiempo de vuelta por km de las *inlaps*.

Como se puede observar en el Cuadro 4.11 los circuitos también influyen en el tiempo de vuelta por km de las *inlaps* gracias a las diferentes características de estos, ya que la mayoría de los coeficientes son significativos.

Cuadro 4.11. *LapTimePerKM-Inlaps* (Circuitos)

| Circuito | Coeficiente | Circuito | Coeficiente |
|-----------|---------------------|---------------|---------------------|
| Austin | 0.0002*** (0.0001) | Austria | -0.0001 (0.0001) |
| Australia | -0.0003*** (0.0001) | Azerbaijan | -0.0001 (0.0001) |
| Bahrain | 0.0002*** (0.0001) | Belgium | 0.001*** (0.0001) |
| Brazil | -0.001*** (0.0001) | Canada | -0.001*** (0.0001) |
| China | 0.0001 (0.0001) | Great Britain | 0.001*** (0.0001) |
| Hungary | -0.0002*** (0.0001) | Imola | -0.001*** (0.0001) |
| Japan | 0.001*** (0.0001) | Mexico | -0.001*** (0.0001) |
| Miami | 0.0001 (0.0001) | Monaco | -0.002*** (0.0001) |
| Monza | 0.001*** (0.0001) | Netherlands | -0.0004*** (0.0001) |
| Qatar | 0.0002*** (0.0001) | Saudi Arabia | 0.001*** (0.0001) |
| Singapore | -0.001*** (0.0001) | Spain | -0.00003 (0.0001) |

Note:

*p<0.1; **p<0.05; ***p<0.01

4.5. *Outlaps*

Las *outlaps* son fundamentales, ya que constituyen la primera vuelta que un piloto realiza al salir de *pits*. Esta vuelta es crucial para llevar las llantas nuevas a la temperatura óptima y puede marcar la diferencia entre un buen o mal comienzo de *stint*.

4.5.1. Selección de variables y modelo

Siguiendo el mismo procedimiento empleado en los modelos anteriores, se utilizó la función `bestglm` para comparar diferentes familias de distribuciones. Los resultados fueron los siguientes:

| | Normal | Gamma | Gaussiana Inversa |
|-------------------|-----------|---------|-------------------|
| Función de enlace | Logaritmo | Inversa | Inversa cuadrada |
| AIC | 12124 | 11593 | 11402 |

Cuadro 4.12. Comparación de diferentes familias de distribución para el modelo de *Outlaps*

Como se observa en el Cuadro 4.12, el modelo con menor AIC es nuevamente el de la Gaussiana Inversa:

$$LapTimePerKM = g^{-1}(\beta_0 + \beta_1 Circuit + \beta_2 Compound) + \varepsilon \quad (4.9)$$

4.5.2. Resultados

Como se puede observar en el Cuadro 4.13, los diferentes tipos de compuestos influyen en el tiempo de vuelta por km de las *outlaps* debido a la diferencia de rendimiento de estos.

Como se puede observar en el Cuadro 4.14, los circuitos también influyen en el tiempo de vuelta por km de las *outlaps* gracias a las

Cuadro 4.13. *Outlaps* - Intercepto y Compuestos

| Variable | Coeficiente |
|-------------------|---------------------------|
| MEDIUM | -1.097e-04 (1.943e-05)*** |
| SOFT | -1.381e-04 (2.331e-05)*** |
| Constante | 0.002*** (0.0000) |
| Observaciones | 2,538 |
| Akaike Inf. Crit. | 11,402.25 |

Cuadro 4.14. *Outlaps* - Circuitos

| Circuito | Coeficiente | Circuito | Coeficiente |
|----------|---------------------|----------------|---------------------|
| Austin | -0.0000 (0.0001) | Australia | 0.001*** (0.0001) |
| Austria | -0.0001 (0.0000) | Azerbaiyán | 0.0001** (0.0001) |
| Baréin | 0.0000 (0.0000) | Bélgica | 0.001*** (0.0001) |
| Brasil | -0.0003*** (0.0000) | Canadá | 0.0002*** (0.0001) |
| China | 0.0000 (0.0001) | Gran Bretaña | 0.0002*** (0.0001) |
| Hungría | -0.0003*** (0.0000) | Imola | -0.0002*** (0.0001) |
| Japón | 0.0004*** (0.0001) | México | -0.0003*** (0.0001) |
| Miami | 0.0002*** (0.0001) | Mónaco | -0.001*** (0.0000) |
| Monza | 0.001*** (0.0001) | Países Bajos | -0.0003*** (0.0001) |
| Catar | 0.0005*** (0.0001) | Arabia Saudita | 0.0003*** (0.0001) |
| Singapur | -0.001*** (0.0001) | España | -0.0000 (0.0000) |

diferentes características de estos, ya que la mayoría de los coeficientes son significativos.

4.6. *Safety Car*

El *Safety Car* es uno de los eventos inciertos que pueden influir significativamente en la estrategia de carrera, ya que puede ofrecer la oportunidad de realizar un pit stop “barato” o incluso “gratis” para los equipos. Durante su intervención, los coches deben reducir su velocidad, lo que permite a los equipos realizar paradas en boxes sin perder posición en pista o perder menos tiempo en comparación con una parada en condiciones normales. Por esta razón, es crucial para los equipos estimar la probabilidad de un *Safety Car* en una vuelta determinada para planificar adecuadamente su estrategia de carrera.

4.6.1. Selección de variables y modelo

Se compararon dos modelos con una regresión logística (familia binomial y función de enlace logit). El primer modelo incluye únicamente el número de vuelta (*LapNumber*), mientras que el segundo incorpora también el circuito (*Circuit*).

| | Solo LapNumber | Circuito y LapNumber |
|-----|----------------|----------------------|
| AIC | 2309.6 | 2261.2 |

Cuadro 4.15. Comparación de modelos para la probabilidad de un *Safety Car*

El modelo con el menor AIC, como se muestra en el Cuadro 4.15, es el que incluye las dos variables. Su especificación sería:

$$SafetyCar = g^{-1}(\beta_0 + \beta_1 LapNumber + \beta_2 Circuit) + \varepsilon \quad (4.10)$$

4.6.2. Resultados

Los Cuadros 4.16 y 4.17 muestran los resultados de la estimación del modelo:

Cuadro 4.16. Probabilidad de *Safety Car* - Intercepto y LapNumber

| Variable | Coeficiente |
|-------------------|------------------|
| Constant | -2.92*** (0.35) |
| LapNumber | -0.02*** (0.003) |
| Observaciones | 5,210 |
| Akaike Inf. Crit. | 2,261.24 |

Nota: *p<0.1; **p<0.05; ***p<0.01

Tiene sentido que el número de vuelta sea significativo en la estimación de la probabilidad de un *Safety Car*, ya que estas intervenciones suelen ocurrir cuando los coches están más cerca unos de otros, como al inicio de la carrera o durante las rondas de paradas en boxes.

Cuadro 4.17. Probabilidad de *Safety Car* - Circuitos

| Circuito | Coeficiente | Circuito | Coeficiente |
|---------------|----------------|------------|-----------------|
| Australia | 1.13*** (0.43) | Azerbaijan | 0.41 (0.47) |
| Austria | 0.61 (0.42) | Bahrain | 0.88** (0.41) |
| Belgium | 0.73 (0.45) | Brazil | 1.43*** (0.39) |
| Canada | 1.13*** (0.40) | China | -13.14 (318.59) |
| Great Britain | 1.29*** (0.39) | Hungary | -1.24* (0.67) |
| Imola | 1.37*** (0.44) | Japan | 0.43 (0.52) |
| Mexico | -0.02 (0.49) | Miami | 0.53 (0.54) |
| Monaco | 0.42 (0.43) | Monza | 0.64 (0.43) |
| Netherlands | 0.79* (0.44) | Qatar | 0.33 (0.57) |
| Saudi Arabia | 0.93** (0.45) | Singapore | 1.41*** (0.41) |
| Spain | 0.02 (0.47) | Austin | 0.12 (0.50) |

Es lógico que el circuito también sea relevante para estimar la probabilidad de un *Safety Car*, ya que cada circuito tiene un porcentaje diferente de ocurrencia de estos eventos. Los circuitos no significativos tienen tasas de ocurrencia similares al circuito base, Abu Dhabi (Yas Marina). En cambio, circuitos con alta probabilidad de *Safety Car*, como Gran Bretaña (Silverstone) y Singapur (Marina Bay), son altamente significativos.

Capítulo 5

Arranca la carrera: Implementación

Desde el arranque hasta la línea de meta, los pilotos deben tomar decisiones cruciales: ser agresivos y intentar adelantar en la primera curva, a pesar del riesgo, o ser pacientes y conservar las llantas, entre otras consideraciones.

De igual manera, en esta tesis es necesario tomar decisiones sobre la estrategia. Aquí es donde entran en juego los árboles de decisión, que permiten comparar estrategias potenciales y seleccionar la mejor opción.

5.1. Árboles de Decisión

Los datos siempre han jugado un papel crucial en la Fórmula 1, incluso antes de la disponibilidad de herramientas sofisticadas. Unas décimas de segundo pueden marcar la diferencia entre comenzar desde la pole position o desde la décima posición.

Sin embargo, los datos por sí solos no garantizan decisiones más

eficientes. Es necesario contar con herramientas que permitan extraer y utilizar el potencial de estos datos para tomar decisiones informadas. Los árboles de decisión son una de estas herramientas.

Según Magee (1964), un árbol de decisión es una representación gráfica de un problema de decisión, que comienza con un nodo raíz representando una decisión inicial. A partir de este nodo, el árbol se ramifica en opciones o alternativas, representadas por ramas. Cada nodo de decisión representa un punto en el que se debe tomar una decisión adicional, y sus ramas representan las opciones disponibles en ese punto. Al final de cada secuencia de decisiones, se llega a un nodo terminal que muestra el resultado de la secuencia.

Para encontrar la mejor decisión, se resuelve el árbol hacia atrás, comenzando desde el nodo terminal con el resultado más deseado y rastreando la ruta de decisiones que lleva a ese resultado. De esta manera, el árbol de decisión ayuda a identificar la serie de decisiones más efectiva para alcanzar el objetivo deseado.

En el contexto de esta tesis, cada nodo terminal del árbol representa el tiempo total esperado de la estrategia, que se calcula utilizando las estimaciones de los modelos presentados en el capítulo 4. Estos modelos incluyen el tiempo de vuelta por kilómetro, el tiempo de las *inlaps*, las *outlaps*, y el tiempo perdido en *pits* por cada parada.

El tiempo total esperado de cada estrategia se calcula combinando los tiempos de los *stints* con el costo de la parada en *pits*.

La Figura 5.1 muestra el árbol de decisión utilizado por los modelos implementados. El nodo raíz representa la decisión sobre el compuesto de neumático a utilizar al inicio de la carrera, es decir, para el primer *stint*. Las opciones disponibles son los tres compuestos de neumático en seco (*Slicks*): *HARD* (H), *MEDIUM* (M) o *SOFT* (S).

El siguiente nodo de decisión corresponde a la primera parada en

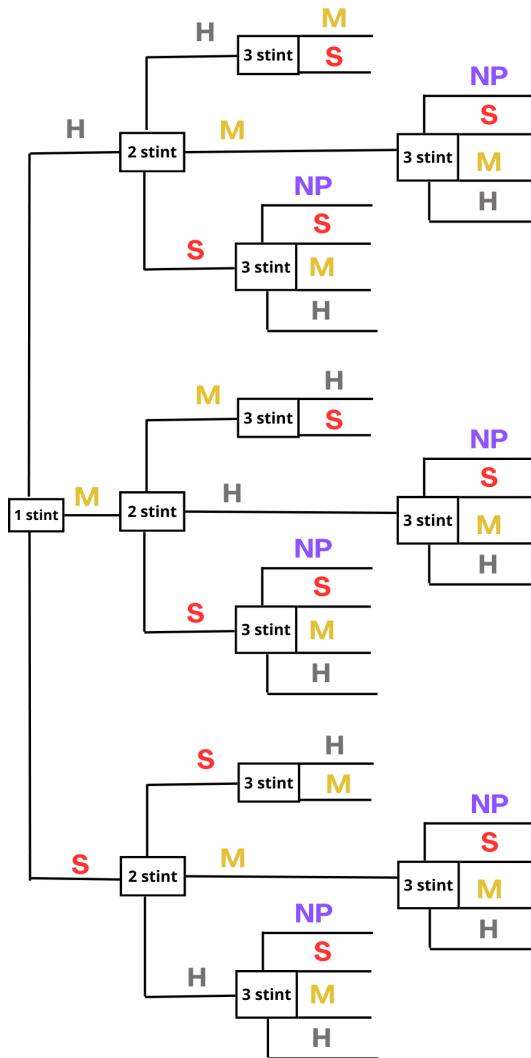


Figura 5.1. Árbol de decisión que siguen los modelos

pits para cambiar llantas. Nuevamente, las opciones son los tres compuestos de neumático. Esta parada es obligatoria por reglamento, ya que, si la pista se declara seca, todos los pilotos deben usar al menos dos compuestos diferentes durante la carrera.

El último nodo de decisión se refiere al tercer *stint*. Si se eligió el mismo compuesto en el segundo *stint* que en el primero, la segunda parada en *pits* es obligatoria. En este caso, las opciones disponibles son los compuestos que no se han usado para evitar una parada adicional. Si en la primera parada se cambió de compuesto, se tienen cuatro opciones: los tres compuestos restantes o la opción de no parar (NP), es decir, no entrar a *pits* para cambiar llantas y continuar hasta el final de la carrera.

Al final de cada ruta, los nodos terminales muestran el tiempo total estimado de la carrera, que incluye los tiempos de los *stints* y el tiempo perdido en *pits*. Con este tiempo estimado, se puede seleccionar la mejor estrategia basada en el menor tiempo total.

5.2. Funciones comunes de los modelos

En esta sección se presentan las funciones comunes que fueron creadas para esta tesis y que son utilizadas en todos los modelos, permitiendo realizar cálculos recurrentes de manera eficiente.

5.2.1. vidapromedio

Esta función tiene como objetivo calcular el máximo, promedio y la desviación estándar de la duración de los *stints* por compuesto de neumáticos.

Recibe como argumentos un conjunto de datos con la información de todos los *stints* (2019-2023) y el circuito. La función comienza filtrando

los datos de acuerdo con el circuito proporcionado, seleccionando un subconjunto de *stints* correspondientes a dicho circuito.

Posteriormente, calcula el máximo, promedio y desviación estándar de la duración de los *stints* por compuesto de neumático. Al finalizar, devuelve un *dataframe* con las estadísticas de cada tipo de neumático.

5.2.2. `tiempoStint`

El objetivo de esta función es calcular el tiempo estimado de un *stint*. La función recibe como argumentos todos los parámetros necesarios para el modelo de *LapTimePerKM*, excepto la vida del neumático y el porcentaje de carrera transcurrido, junto con un intervalo de vueltas sobre el que se realizará el cálculo.

En primer lugar, se calculan las estadísticas para cada compuesto de neumático utilizando la función `vidapromedio`. Posteriormente, se determinan el número total de vueltas de la carrera y los kilómetros por vuelta. Se inicializa una variable que almacenará el tiempo total del *stint*, inicialmente establecida en 0.

A continuación, se ejecuta un bucle que forma un *dataframe* con todos los datos necesarios para predecir el tiempo por vuelta utilizando el modelo. En cada iteración del bucle, se actualizan la vida del neumático y el porcentaje de carrera transcurrido para cada vuelta dentro del intervalo. Para cada vuelta, se predice el tiempo utilizando el modelo *LapTimePerKM*, se multiplica por los kilómetros por vuelta y se añade al tiempo acumulado del *stint*.

De esta forma, al final del bucle, se obtiene el tiempo total estimado para el *stint* en cuestión, considerando la degradación progresiva del neumático y el avance de la carrera.

5.2.3. pitstopcost

Esta función estima el tiempo total perdido durante una parada en *pits*, combinando los tiempos de la *inlap*, la parada en *pits*, y la *outlap*. El proceso consiste en predecir los tiempos de cada segmento utilizando los modelos correspondientes y sumarlos.

Para la *inlap* y la *outlap*, el tiempo predicho se multiplica por los kilómetros por vuelta, ajustando así el cálculo a la longitud de la pista. Al final, la función devuelve el tiempo total perdido en *pits*, que incluye tanto el tiempo de la parada como el tiempo extra debido a las vueltas de entrada y salida de *pits*.

5.3. Modelo Determinista

En Fórmula 1, los equipos a veces se preocupan por las acciones de sus rivales, pero según Bernie Collins, su enfoque principal está en sus propios pilotos. “[...] calcularíamos la estrategia más rápida para un coche en particular. En otras palabras, si fueras el único coche en la pista, ¿cuál sería el tiempo más rápido posible para completar la distancia de carrera y cómo lo lograrías? ”(Collins, 2024, 126).

Este enfoque es conocido como una *single car race* y representa el punto de partida para cualquier estrategia de carrera. Según Collins, la estrategia óptima en carreras con muchas oportunidades de adelantamiento tiende a parecerse mucho a la de una *single car race*, ya que, en teoría, el piloto debería poder adelantar con neumáticos más frescos.

El primer modelo que se presenta está basado en una *single car race*, con la particularidad de que las paradas en *pits* sólo pueden realizarse en la vuelta correspondiente a la vida útil esperada del compuesto de neumáticos. Por esta razón, este modelo se denomina determinista, ya

que calcula todas las posibles alternativas estratégicas siguiendo el árbol de decisiones descrito previamente.

Para simplificar el flujo del modelo, se abordan los grupos de ramas en orden: primero el grupo que comienza con el primer *stint* de neumáticos duros (*Hards*), luego el primer *stint* de neumáticos medios (*Mediums*), y finalmente el de neumáticos blandos (*Softs*). En la Figura 5.2 se muestra el ejemplo del grupo que comienza con el primer *stint* de neumáticos duros (*Hards*).

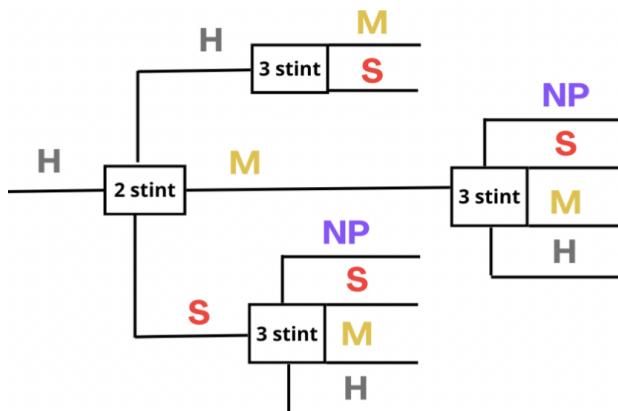


Figura 5.2. Rama del Primer *stint* con *Hards*

En primer lugar, se calculan los tiempos del primer *stint* para los tres compuestos utilizando la función `tiempoStint` hasta la vuelta correspondiente a la vida útil esperada de cada compuesto.

5.3.1. Segundo *stint* con el mismo compuesto

Una vez calculado el tiempo del primer *stint* con neumáticos duros, el modelo avanza hacia el grupo de ramas del segundo *stint* con el mismo compuesto. Al tiempo ya calculado del primer *stint*, se le añade el tiempo

perdido en *pits* por cambiar a un nuevo juego de neumáticos duros. Posteriormente, se evalúa si la vida útil de este nuevo set sería suficiente para finalizar la carrera.

Si la vida útil permite completar la distancia restante, el piloto podría terminar con este set de neumáticos. Sin embargo, debido a la obligación reglamentaria de utilizar dos compuestos diferentes, sería necesaria otra parada. El segundo *stint* se calcula hasta la antepenúltima vuelta, dejando la penúltima para la entrada a *pits* (*inlap*) y la última para la salida (*outlap*), ambos tiempos considerados dentro de la función *pitstopcost* para calcular el tiempo total perdido en *pits*.

Si la vida útil del compuesto no es suficiente para completar la carrera, se calcula el tiempo del segundo *stint* hasta la vuelta anterior a su desgaste y se determina el tiempo necesario para las vueltas restantes con los otros dos compuestos.

5.3.2. Segundo *stint* con un compuesto diferente

Si se decide hacer el segundo *stint* con un compuesto diferente, se pasa al grupo de ramas donde hay cuatro opciones para el tercer *stint*. Primero se calcula el tiempo perdido en *pits* para el nuevo compuesto elegido.

Para la opción de no parar nuevamente, se evalúa si una sola parada es suficiente. Si no lo es, el modelo genera una advertencia con el número de vueltas que exceden la vida útil del compuesto, además de señalar la vida máxima del mismo. No obstante, el modelo no descarta esta estrategia y procede a calcular el tiempo del tercer *stint* para las vueltas restantes de la carrera.

En el caso de decidir hacer una segunda parada, se calcula el tiempo del segundo *stint* hasta la vuelta correspondiente a la vida útil esperada

del compuesto elegido. Posteriormente, se calcula el tiempo perdido en *pits* y el tiempo del tercer *stint* para cada uno de los compuestos.

5.3.3. Resultados

Después de aplicar el mismo proceso a todas las ramas del modelo, este devuelve un vector con las diferentes estrategias posibles, junto con el tiempo estimado para cada una en segundos y las advertencias correspondientes. La estrategia óptima se puede determinar seleccionando la opción con el menor tiempo esperado.

En algunas carreras, debido al elevado tiempo perdido en *pits* o a la baja degradación de los neumáticos, los equipos optan por una estrategia de una sola parada. Para estos casos, el modelo determinista también ha sido configurado para considerar dos *stints*. Su funcionamiento es similar al del modelo descrito previamente, con la diferencia de que, en lugar de un árbol de decisiones con tres *stints*, se utiliza uno con solo dos. Esto simplifica el proceso de decisión sin perder precisión en la estimación del tiempo total de carrera.

5.4. Modelo Ventana

El punto de partida de cualquier estrategia en Fórmula 1 es la *single car race*, ya modelada en el enfoque determinista. Sin embargo, los estrategas también se interesan en evaluar el impacto de modificar el momento de la parada en *pits*. Por ejemplo, si se realiza la parada dos vueltas antes de lo planeado, “[...] en este caso estarías evitando parte de la degradación al detenerte antes en tu primer *stint*, pero también tendrías que sumar dos vueltas más lentas al final del segundo *stint*” (Collins, 2024, 127).

A este concepto se le conoce como *pit window* o ventana de parada, que se refiere al intervalo de vueltas en el que se espera que un piloto entre a *pits*. Usualmente, abarca algunas vueltas antes y después de la parada prevista en la estrategia óptima.

Basado en este principio, se desarrolló el modelo de ventana, que sigue el mismo árbol de decisiones que el modelo determinista, pero con una diferencia clave: la restricción de detenerse exactamente en la vuelta correspondiente a la vida esperada de los neumáticos se relaja para la primera parada. En lugar de ello, se suma un valor de ventana a la vuelta esperada para determinar el nuevo punto de parada.

La función `calculaVentana` implementa este modelo utilizando diferentes valores de ventana {-3, -2, -1, 0, 1, 2, 3}. Un valor de -3 significa detenerse tres vueltas antes de la vida esperada de los neumáticos, 0 representa la parada en la vuelta exacta de la vida esperada, y 3 implica detenerse tres vueltas después de la vida esperada.

Los resultados del modelo consisten en una serie de estrategias posibles para cada valor de ventana, permitiendo elegir la opción con el menor tiempo esperado. Además, estas estrategias pueden compararse entre diferentes escenarios.

Este modelo proporciona mayor flexibilidad a los equipos, permitiéndoles ajustar los *stints* en función de factores externos durante la carrera, como incidentes o cambios en las condiciones climáticas.

5.5. Modelo Rival

Hasta ahora, los modelos presentados han sido estáticos, lo que significa que calculan las estrategias óptimas basándose en la posición

del piloto después de la primera vuelta sin considerar que esta posición puede variar. Tampoco recalculan la estrategia con base en estos cambios.

El Modelo Rival introduce un primer enfoque hacia un modelo dinámico, ya que tiene en cuenta el intervalo de tiempo entre los coches. La idea es recalcular la estrategia en cada vuelta, considerando la proximidad del rival más cercano y la posibilidad de realizar un undercut o un overcut.

El modelo genera un conjunto de estrategias posibles junto con la diferencia de tiempo que se ganaría o perdería en comparación con el rival elegido. En esta versión del modelo, no se consideran diferentes compuestos de neumáticos, sino que se centra en el momento de la parada en *pits* en relación con la vida útil esperada de los neumáticos.

Las tres estrategias posibles son:

- Parar tres vueltas antes de la vida esperada del neumático.
- Parar en la vuelta correspondiente a la vida esperada del neumático.
- Parar tres vueltas después de la vida esperada del neumático.

Este modelo es un primer paso hacia una dinámica de carrera que, sin embargo, aún no es completamente probable, ya que para una precisión total se necesitarían los intervalos exactos de tiempo durante la carrera. Estos pueden variar debido a factores como *safety cars*, paradas en *pits* y otros incidentes en pista. La diferencia entre los tiempos de vuelta solamente no es suficiente para calcular con exactitud los tiempos vuelta por vuelta.

En el Cuadro 5.1 se presenta un ejemplo del modelo aplicado en el Gran Premio de Mónaco. En este caso, el equipo McLaren, con Lando

| Equipo | Rival | DifTiempo | Estrategia |
|---------------|--------------|------------------|-------------------|
| 1 | 1 | 42.48 | Neutral |
| 1 | 2 | -416.42 | Undercut |
| 1 | 3 | -875.20 | Undercut |
| 2 | 1 | 504.74 | Overcut |
| 2 | 2 | 45.84 | Neutral |
| 2 | 3 | -412.94 | Undercut |
| 3 | 1 | 966.88 | Overcut |
| 3 | 2 | 507.98 | Overcut |
| 3 | 3 | 49.20 | Neutral |

Cuadro 5.1. Ejemplo del modelo rival

Norris en la tercera posición, enfrenta a su rival Charles Leclerc de Ferrari, ubicado en la segunda posición con un intervalo de 3 segundos.

En este caso, el modelo sugiere un undercut de tres vueltas antes, ya que esta estrategia ofrece la mayor ganancia en el tiempo total esperado en comparación con el rival.

Capítulo 6

La Línea de Meta: Resultados

Al cruzar la línea de meta, los equipos suelen reflexionar sobre sus errores y aciertos. Se pueden apreciar los frutos de una estrategia acertada, así como las lecciones aprendidas de una decisión errada.

En este capítulo, se evalúan los aciertos y desaciertos de los modelos implementados.

Para analizar los resultados de la implementación de este modelo, se seleccionaron carreras de la temporada 2024 que no estuvieran sesgadas por factores externos, como la aparición de un *Safety Car* o condiciones de lluvia, ya que el modelo no considera estos elementos en la estimación de la estrategia óptima.

La elección del compuesto con el que iniciar la carrera puede depender de múltiples factores, así como de la asignación de neumáticos realizada por cada equipo. Por esta razón, se seleccionó la estrategia más rápida para cada tipo de compuesto inicial: una para el compuesto *Soft*, otra para *Medium* y otra para *Hard*.

La precisión del modelo se evaluó a partir de tres carreras en las que no ocurrieron eventos externos que pudieran distorsionar los tiempos de vuelta o la vida útil de los neumáticos.

A través de los ejemplos que se presentarán a continuación, es importante destacar que varios factores influyen en la decisión sobre la estrategia de carrera. En ocasiones, no se puede optar por la estrategia óptima; aunque se mida la precisión comparando las estrategias reales utilizadas por los equipos, es fundamental considerar que estas pueden no ser las mejores opciones. Factores como la asignación de neumáticos, errores estratégicos y la búsqueda de alternativas para obtener más puntos o ganar la carrera pueden llevar a los equipos a elegir estrategias subóptimas. Por lo tanto, la estrategia real podría diferir de la óptima.

6.1. Gran Premio de Bahréin 2024

En la Figura 6.1 se pueden observar las estrategias de neumáticos utilizadas por los pilotos durante la carrera. El color rojo indica un stint con neumáticos *Soft*, el gris representa un stint con neumáticos *Hard* y el amarillo señala un stint con neumáticos *Medium*. Sin embargo, en esta carrera particular solo se usaron los compuestos *Soft* y *Hard*. El Gran Premio de Bahréin tiene 57 vueltas, por lo que los pilotos cuyas barras llegan hasta la vuelta 57 completaron la carrera. Por otro lado, Zhou (ZHO), Magnussen (MAG), Ricciardo (RIC), Tsunoda (TSU), Albon (ALB), Hulkenberg (HUL), Ocon (OCO), Gasly (GAS) y Bottas (BOT) fueron alcanzados y superados por el líder, terminando la carrera con una vuelta menos. Sargeant (SAR) finalizó con dos vueltas de desventaja.

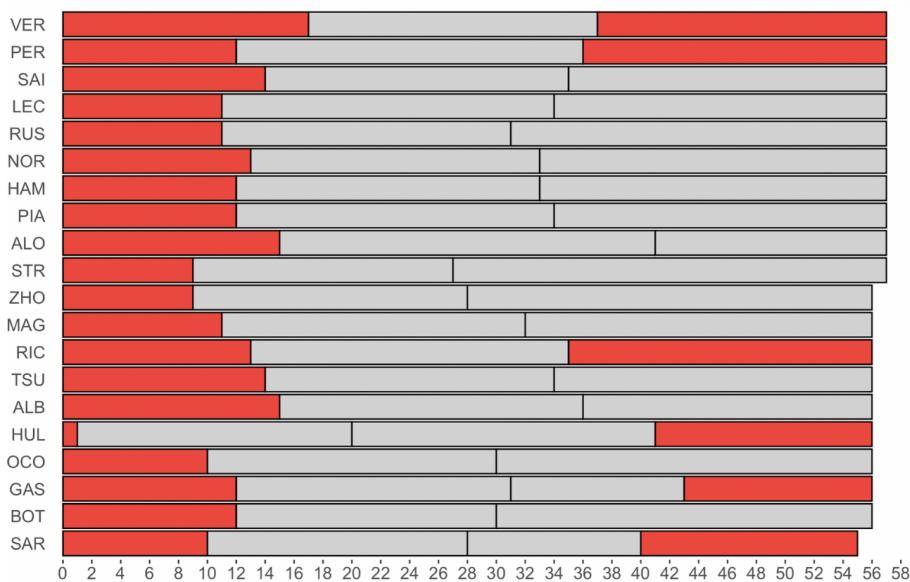


Figura 6.1. Estrategias de neumáticos para el Gran Premio de Bahréin 2024

6.1.1. Modelo Determinista

El Cuadro 6.1 presenta las estrategias óptimas según el modelo determinista. Aunque los tiempos esperados de cada piloto y equipo son diferentes, la estrategia óptima es la misma para todos. En esta carrera, todos los pilotos comenzaron con el compuesto *Soft*, por lo que la mejor estrategia según el modelo fue *Soft-Hard-Hard*. Al compararla con las estrategias reales, se observa que tiene una precisión del 70 %. Esta precisión se calcula como el porcentaje de veces en que la estrategia óptima propuesta por el modelo coincide con la estrategia real utilizada por el piloto en la carrera, la cual se puede observar en la Figura 6.1. Dado que ninguno de los pilotos utilizó *Hards* o *Mediums* para su primer *stint*, no se calculan las precisiones de las estrategias

con estos compuestos iniciales. De las 20 estrategias reales, 14 coincidieron con la estrategia óptima propuesta por el modelo, lo que da una precisión del 70 %.

| Compuesto inicial | Precisión | Estrategia propuesta |
|-------------------|-----------|----------------------|
| HARD | - | HARD-MEDIUM |
| MEDIUM | - | MEDIUM-HARD |
| SOFT | 0.7 | SOFT-HARD-HARD |
| General | 0.7 | |

Cuadro 6.1. Estrategias óptimas propuestas por el modelo para Bahréin 2024

De hecho, la estrategia *Soft-Hard-Hard*, fue la estrategia más popular entre los pilotos. Sin embargo, aquellos que no utilizaron la estrategia óptima según el modelo, como Max Verstappen y Checo Pérez del equipo Red Bull Racing, pueden haber tomado decisiones alternativas debido a la asignación de neumáticos para la carrera. Estos dos pilotos contaban con 1 set de neumáticos *Soft* nuevos y 3 sets de *Soft* usados, pero solo disponían de 1 set de neumáticos *Hard* nuevos Formula 1 (2024d). Esto limitó su capacidad para optar por la estrategia *Soft-Hard-Hard*, a diferencia de otros pilotos del Top 10, que, aunque no tenían sets de *Soft* nuevos, sí contaban con 2 sets de *Hard* nuevos. Este ejemplo ilustra cómo la asignación de neumáticos puede influir en la elección de una estrategia alternativa.

6.1.2. Modelo Ventana

En el Cuadro 6.2 se indica la mejor estrategia para todos los pilotos determinada por el Modelo Ventana que sugiere parar 3 vueltas

antes que la vuelta de la vida esperada de los neumáticos con la mejor estrategia *Medium-Medium-Hard*.

Cuadro 6.2. Mejor Estrategia y Ventana de Parada

| Circuito | Mejor Estrategia | Ventana de Parada |
|----------|--------------------|-------------------|
| Bahréin | MEDIUM-MEDIUM-HARD | -3 |

6.2. Gran Premio de Singapur 2024

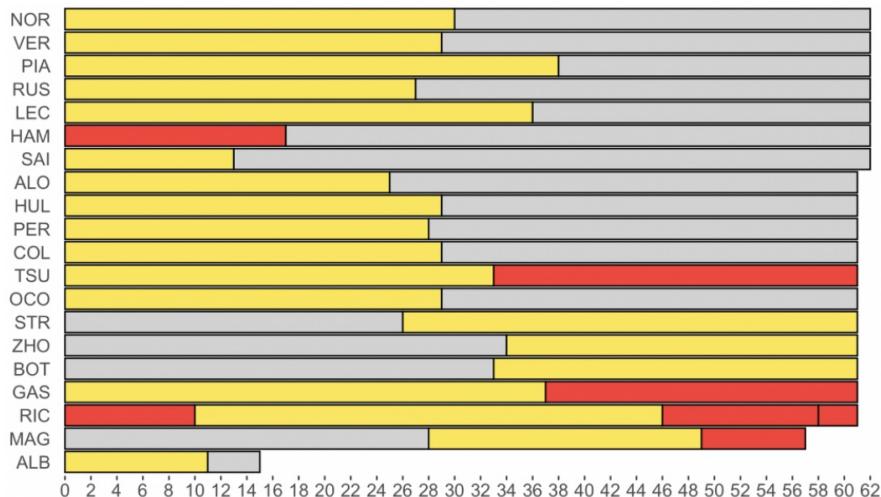


Figura 6.2. Estrategias de neumáticos para el Gran Premio de Singapur 2024

En la Figura 6.2 se pueden observar las estrategias de neumáticos utilizadas por los pilotos durante la carrera. El Gran Premio de Singapur tiene 62 vueltas, por lo que solo los 7 primeros pilotos terminaron todas las vueltas de la carrera. Los pilotos de las posiciones 8 a 18 terminaron

una vuelta atrás, mientras que Albon (ALB) se retiró en la vuelta 15 y Magnussen (MAG) se retiró en la vuelta 57.

6.2.1. Modelo Determinista

| Compuesto inicial | Precisión | Estrategia propuesta |
|-------------------|-----------|----------------------|
| HARD | 0.15 | HARD-MEDIUM |
| MEDIUM | 0.60 | MEDIUM-HARD |
| SOFT | 0.00 | SOFT-HARD-HARD |
| General | 0.75 | |

Cuadro 6.3. Estrategias óptimas y su precisión para Singapur 2024

El Cuadro 6.3 presenta las estrategias óptimas según el modelo determinista. En esta carrera, de las 20 estrategias reales, 15 coincidieron con las estrategias óptimas según el modelo: 12 con *Medium* como compuesto inicial y 3 con *Hard* como compuesto inicial. Así, el modelo mostró una precisión general del 75 %, siendo la estrategia más popular *Medium-Hard*, aunque también se utilizó el compuesto *Hard* como neumático de salida.

El circuito de Singapur es un trazado callejero muy estrecho, con paredes a ambos lados, lo que deja poco margen de error para los pilotos. Además, la carrera es físicamente exigente debido a la humedad y el calor, lo que limita las oportunidades de adelantamiento.

El neumático *Soft*, al ser el más blando, proporciona mayor agarre en la salida, lo que permite un arranque más rápido. Por ello, no es ilógico utilizarlo en la salida para intentar ganar posiciones en las primeras vueltas. Sin embargo, según Formula 1 (2024b), James Alisson, director técnico del equipo Mercedes, mencionó que habían

evaluado incorrectamente la carrera al optar por neumáticos *Soft* para Lewis Hamilton, quien comenzaba en tercera posición. Su intención era aprovechar una posible salida del *Safety Car* para cambiar neumáticos. Esta estrategia parecía razonable, dado que, antes de este año, el *Safety Car* había tenido un 100 % de ocurrencia en este circuito. Sin embargo, el *Safety Car* nunca llegó, y esta decisión costó a Mercedes tres posiciones al final de la carrera.

6.2.2. Modelo Ventana

En el Cuadro 6.4 se indica la mejor estrategia para todos los pilotos determinada por el Modelo Ventana que sugiere parar 3 vueltas antes que la vuelta de la vida esperada de los neumáticos con la mejor estrategia *Soft-Hard-Hard*.

Cuadro 6.4. Mejor Estrategia y Ventana de Parada

| Circuito | Mejor Estrategia | Ventana de Parada |
|----------|------------------|-------------------|
| Singapur | SOFT-HARD-HARD | -3 |

6.3. Gran Premio de Monza 2024

En la Figura 6.3 se pueden observar las estrategias de neumáticos utilizadas por los pilotos durante la carrera. El Gran Premio de Monza tiene 53 vueltas, por lo que todos los pilotos desde la posición 1 hasta la posición 13 terminaron todas las vueltas de la carrera. Los demás pilotos terminaron 1 vuelta atrás, con el único abandono de la carrera, Tsunoda (TSU), quien se retiró en la vuelta 7.

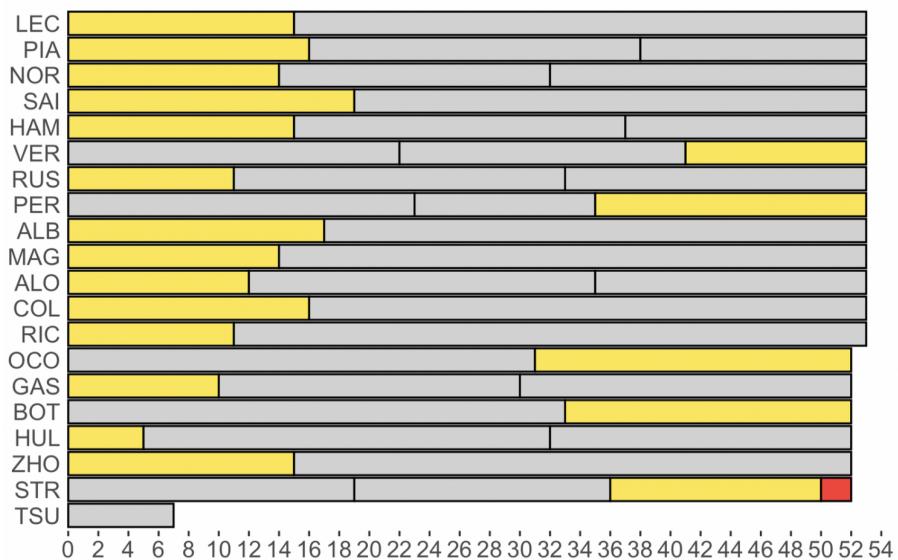


Figura 6.3. Estrategias de neumáticos para el Gran Premio de Monza 2024

6.3.1. Modelo Determinista

El Cuadro 6.5 presenta las estrategias óptimas según el modelo determinista.

De las 20 estrategias reales, 9 coincidieron con las estrategias óptimas del modelo: 7 con el compuesto inicial *Medium* y 2 con el compuesto inicial *Hard*. En esta carrera, el modelo mostró una precisión general del 45 %. En Monza, las estrategias se dividieron entre una y dos paradas, lo que podría explicar por qué la precisión del modelo no fue tan alta en comparación con otras carreras.

Charles Leclerc, piloto de Ferrari, ganó la carrera partiendo desde la tercera posición y utilizando la estrategia óptima de *Medium-Hard* según el modelo.

| Compuesto inicial | Precisión | Estrategia propuesta |
|-------------------|-----------|----------------------|
| HARD | 0.10 | HARD-MEDIUM |
| MEDIUM | 0.35 | MEDIUM-HARD |
| SOFT | 0.00 | SOFT-HARD-HARD |
| General | 0.45 | |

Cuadro 6.5. Estrategias óptimas y su precisión según el modelo para Monza 2024

Por otro lado, los McLaren de Lando Norris y Oscar Piastri, que partieron desde las posiciones 1 y 2, optaron por una estrategia de dos paradas, lo cual podría haberles costado la victoria. Formula 1 (2024c).

6.3.2. Modelo Ventana

En el Cuadro 6.6 se indica la mejor estrategia para todos los pilotos determinada por el Modelo Ventana que sugiere parar 3 vueltas antes que la vuelta de la vida esperada de los neumáticos con la mejor estrategia *Medium-Medium-Hard*.

Cuadro 6.6. Mejor Estrategia y Ventana de Parada

| Circuito | Mejor Estrategia | Ventana de Parada |
|----------|--------------------|-------------------|
| Monza | MEDIUM-MEDIUM-HARD | -3 |

6.4. Gran Premio de China 2024

En la Figura 6.4 se pueden observar las estrategias de neumáticos utilizadas por los pilotos durante la carrera. El Gran Premio de China tiene 56 vueltas, por lo que todos los pilotos terminaron todas las vueltas

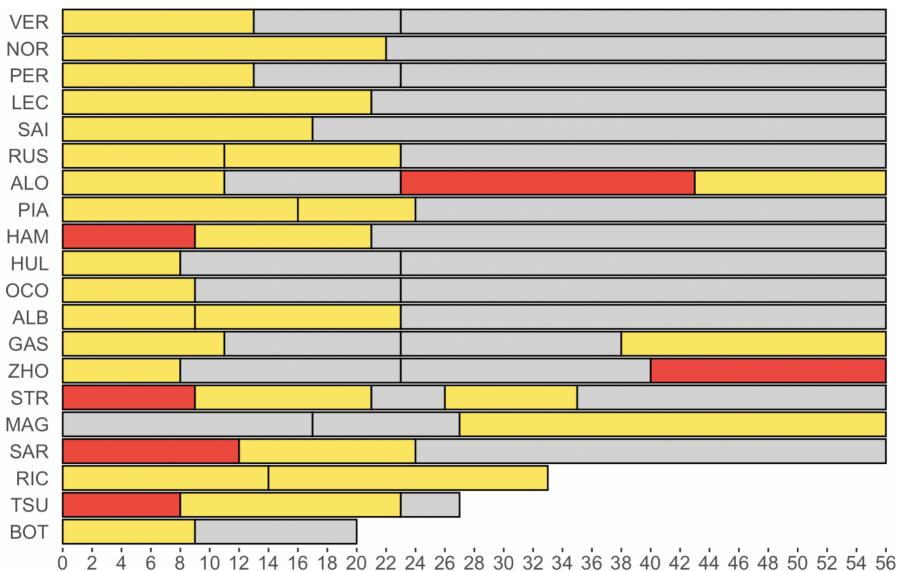


Figura 6.4. Estrategias de neumáticos para el Gran Premio de China 2024

de la carrera, excepto Ricciardo (RIC), Tsunoda (TSU) y Bottas (BOT), quienes se retiraron en las vueltas 33, 27 y 20, respectivamente.

6.4.1. Modelo Determinista

El Cuadro 6.7 presenta las estrategias óptimas según el modelo determinista. De las 20 estrategias reales, 4 coincidieron con las estrategias óptimas propuestas por el modelo, todas ellas con *Medium* como compuesto inicial. En esta carrera, el modelo mostró una precisión del 20 %. Aunque el compuesto inicial más popular fue el *Medium*, también hubo pilotos que optaron por iniciar con el compuesto *Hard* y *Soft*.

Mientras que en los ejemplos anteriores se analizaron carreras en

| Compuesto inicial | Precisión | Estrategia propuesta |
|-------------------|-----------|----------------------|
| HARD | 0.0 | HARD-MEDIUM |
| MEDIUM | 0.20 | MEDIUM-HARD |
| SOFT | 0.00 | SOFT-HARD-HARD |
| General | 0.20 | |

Cuadro 6.7. Estrategias óptimas y su precisión según el modelo determinista para China 2024

las que no ocurrió ninguna intervención que pudiera haber afectado la vida útil de las llantas, se decidió incluir una carrera en la que sí hubo intervenciones, para analizar la precisión del modelo en estos escenarios.

En el Gran Premio de China hubo dos intervenciones del *Safety Car* que impactaron el curso de las estrategias de los equipos. Por ejemplo, Verstappen y Pérez de Red Bull Racing habían parado justo unas vueltas antes del *Safety Car* para cambiar a llantas nuevas; sin embargo, volvieron a entrar a pits durante este periodo. En la rearrancada de la carrera, otro incidente causó la aparición de un segundo *Safety Car*. Estas intervenciones dieron ventaja a algunos pilotos: a quienes no habían parado antes del *Safety Car*, les permitió hacer una carrera de una sola parada, mientras que otros aprovecharon la oportunidad para hacer una estrategia diferente o, si tenían daños en las llantas, hacer múltiples paradas Formula 1 (2024e).

Por lo tanto, la precisión del modelo fue baja en este caso, ya que muchos equipos se desviaron de la estrategia óptima según el modelo debido a las intervenciones del *Safety Car*.

6.4.2. Modelo Ventana

En el Cuadro 6.8 se indica la mejor estrategia para todos los pilotos determinada por el Modelo Ventana que sugiere parar 3 vueltas antes que la vuelta de la vida esperada de los neumáticos con la mejor estrategia *Soft-Hard-Hard*.

Cuadro 6.8. Mejor Estrategia y Ventana de Parada

| Circuito | Mejor Estrategia | Ventana de Parada |
|----------|------------------|-------------------|
| China | SOFT-HARD-HARD | -3 |

6.5. Safety Cars

Para probar el modelo de *Safety Cars*, se categorizaron todos los circuitos del calendario 2024 (hasta Singapur) en tres grupos de acuerdo a las categorías descritas en el Cuadro 3.2 del Capítulo 3.

Se corrió el modelo para cada uno de estos grupos y se calcularon las probabilidades de *Safety Car* por porcentaje de carrera completado.

En la Figura 6.5 se muestra la probabilidad estimada por el modelo, según el porcentaje de carrera completado, para los circuitos de riesgo bajo (China y Hungría). Hungría tiene una probabilidad muy baja de *Safety Car* durante toda la carrera, mientras que China, aunque solo tuvo un Gran Premio en 2019 antes de su regreso en 2024 por la pandemia, presenta una tendencia decreciente hacia el final de la carrera.

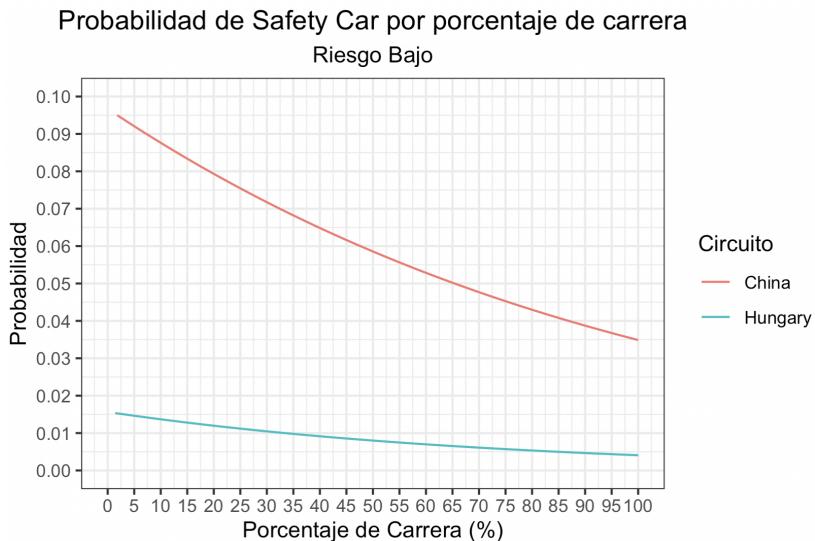


Figura 6.5. Probabilidad de *Safety Car*: Circuitos de riesgo bajo

En la Figura 6.6 se puede observar la probabilidad estimada para los circuitos de riesgo medio. Estos presentan una tendencia similar a la de los circuitos de riesgo bajo, con un punto máximo al inicio de la carrera y una disminución gradual, aunque en general las probabilidades son más altas que en los circuitos de bajo riesgo.

La Figura 6.7 muestra las probabilidades estimadas para los circuitos de riesgo alto. Estos circuitos tienen una mayor probabilidad de *Safety Car* durante toda la carrera.

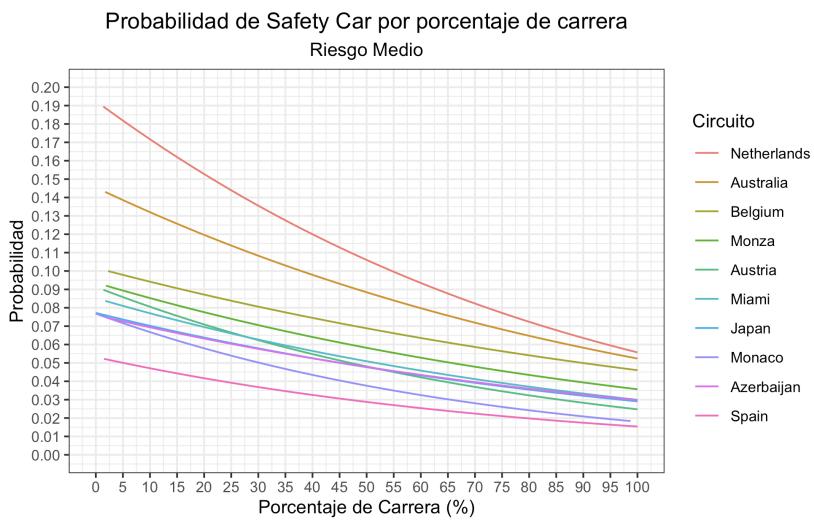


Figura 6.6. Probabilidad de *Safety Car*: Circuitos de riesgo medio

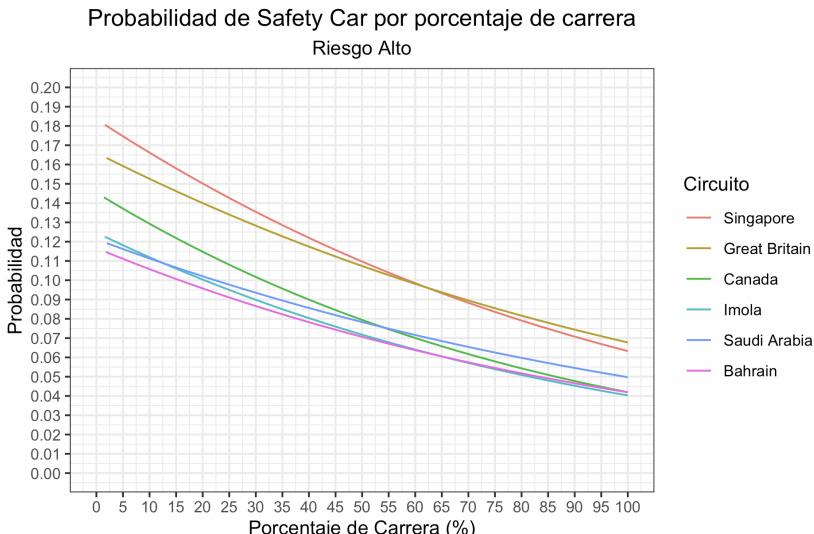


Figura 6.7. Probabilidad de *Safety Car*: Circuitos de riesgo alto

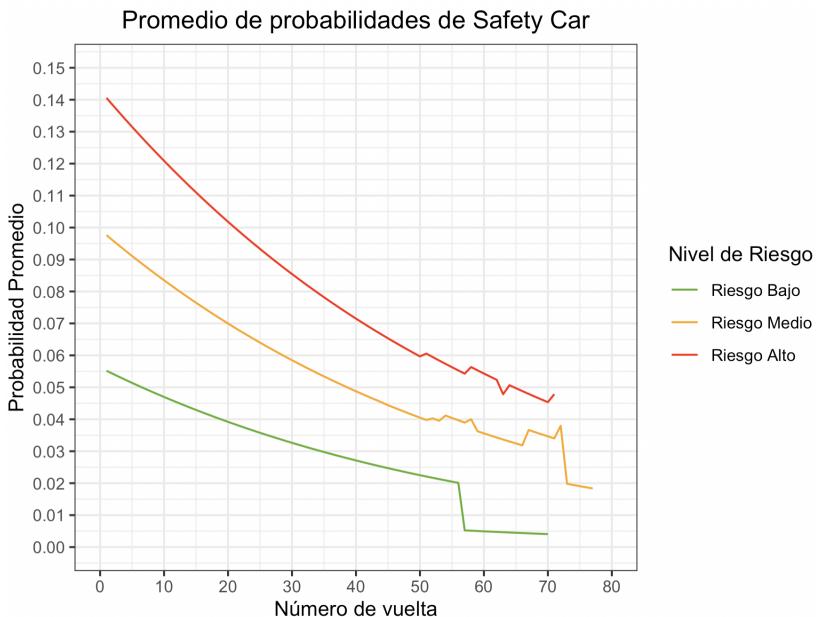


Figura 6.8. Probabilidad promedio de *Safety Car* por nivel de riesgo

En la Figura 6.8 se presenta el promedio de probabilidades de Safety Car por número de vuelta, desglosado por nivel de riesgo. Siguiendo la tendencia observada en las gráficas anteriores, los circuitos de riesgo alto muestran la mayor probabilidad promedio de *Safety Car*, seguidos por los de riesgo medio, y finalmente los de riesgo bajo. La probabilidad máxima ocurre al inicio de la carrera, disminuyendo hacia el final.

Debrief: Discusión

Tras cada carrera, todo el equipo se reúne con sus pilotos en el cuarto de ingenieros para analizar lo sucedido. Este momento de reflexión es crucial, ya que se revisan las decisiones tomadas y se evalúan los aciertos y errores.

Al igual que los equipos de Fórmula 1, esta tesis también requiere un *debrief* para discutir los resultados y proponer futuras investigaciones.

El objetivo de esta tesis fue diseñar un modelo de estrategia para los equipos de Fórmula 1, que les ayudara a tomar decisiones al encontrar la estrategia óptima para una carrera específica. Para lograr este objetivo, se extrajeron datos de la API oficial de la Fórmula 1 y, como base de cualquier análisis, se realizó una exploración inicial de los datos para identificar los factores clave que determinan la estrategia de carrera. Posteriormente, se desarrollaron modelos para estimar la estrategia óptima. Sin embargo, fue necesario complementar esto con un método de decisión entre estrategias, lo que llevó a incorporar árboles de decisión.

En las primeras etapas de esta tesis, se utilizaron modelos de regresión lineal múltiple para estimar los tiempos de vuelta y el tiempo perdido en pits. A pesar de obtener un R^2 alto, las pruebas de normalidad y heterocedasticidad mostraron que estos modelos no eran

adecuados, ya que los datos no se ajustaban a los supuestos. Así, se decidió implementar modelos lineales generalizados, que resultaron más flexibles en estos aspectos.

Se alcanzaron los objetivos planteados, pero me gustaría resaltar algunos puntos sobre los aciertos y limitaciones de los modelos implementados:

Modelo Determinista: Este modelo tuvo una precisión alta en carreras sin intervenciones del clima o *Safety Car*, como en el Gran Premio de Bahréin 2024. Sin embargo, en carreras con múltiples intervenciones del *Safety Car*, como en el Gran Premio de China 2024, la precisión disminuyó considerablemente, algo esperado ya que el modelo no incluye la probabilidad de *Safety Car* como parámetro. La precisión también bajó, aunque en menor medida, en situaciones donde los equipos utilizaron estrategias alternativas en el número de paradas, como se observó en el Gran Premio de Monza 2024.

Evaluación frente a Estrategias Reales: La precisión del modelo se evaluó respecto a las estrategias usadas por los equipos en la carrera, aunque no se puede afirmar con certeza que esas fueron siempre las estrategias óptimas. Ejemplos como el Gran Premio de Singapur 2024 ilustran que los equipos también pueden cometer errores estratégicos.

Modelo Ventana: En todos los casos analizados, el modelo sugiere un *undercut* de tres vueltas antes del límite de vida esperado de los neumáticos, lo cual es consistente con el hecho de que los equipos suelen intentar cambiar los neumáticos cuando están desgastados o si el piloto está atrapado en un tren de *DRS*.

Este trabajo es un primer paso hacia un modelo de estrategia de carrera, y futuras investigaciones podrían mejorar la precisión en carreras con *Safety Car* incorporando la probabilidad de intervención. También sería valioso incluir información sobre el estado de la pista en

condiciones de lluvia, para optimizar las estrategias de neumáticos en dichas circunstancias.

Además, cabe señalar que los modelos no son dinámicos, ya que se desarrollan con la posición de los pilotos tras la primera vuelta y no consideran incidentes que causen cambios significativos en la clasificación. Una posible mejora sería recalcular la estrategia tras un cambio significativo en la posición, aunque esto puede ser complicado por el ritmo acelerado de las decisiones en Fórmula 1. En este contexto, el **Modelo Rival** representa un primer intento de modelo dinámico, al considerar los intervalos de tiempo entre pilotos.

Este trabajo contribuye a la literatura existente proponiendo un enfoque basado en modelos lineales generalizados y árboles de decisión para determinar la estrategia de carrera, en contraste con enfoques previos como redes neuronales o programación dinámica. Estos resultados sientan una base prometedora para desarrollar modelos más complejos y adaptativos que se acerquen a la toma de decisiones en tiempo real de los equipos de Fórmula 1.

En conclusión, este trabajo representa un avance inicial en el desarrollo de modelos de estrategia para equipos de Fórmula 1, proporcionando una herramienta útil para la toma de decisiones estratégicas. A través del uso de modelos lineales generalizados y árboles de decisión, se logró abordar la complejidad de las carreras, ofreciendo una metodología que puede adaptarse y mejorar con la incorporación de factores adicionales, como la probabilidad de *Safety Car* y condiciones de pista. Los resultados aquí obtenidos demuestran el potencial de estos modelos y sugieren el camino hacia investigaciones futuras que puedan llevar estos enfoques hacia aplicaciones más dinámicas y específicas para la competencia de Fórmula 1. Con ello, se abre la puerta a modelos aún más precisos y

flexibles que reflejen las condiciones cambiantes de cada carrera y ayuden a los equipos a optimizar sus decisiones en tiempo real.

Glosario

- **Aire sucio:** Flujo de aire turbulento que se genera cuando un coche está detrás de otro. Este fenómeno puede reducir la carga aerodinámica del coche que va detrás, dificultando su manejo y rendimiento.
- **Box:** Área específica dentro del *pitlane* donde cada equipo tiene su espacio asignado frente a su garaje para atender a sus coches durante las paradas. También se usa en la comunicación con los pilotos para indicarles que es tiempo de su parada, y comúnmente se escucha "box, box".
- **Downforce:** Fuerza aerodinámica que empuja el coche hacia el suelo, mejorando su agarre y estabilidad en las curvas.
- **DRS:** *Drag Reduction System*, es un mecanismo que reduce la resistencia al aire del coche para aumentar su velocidad máxima, facilitando los adelantamientos. Sin embargo, solo se puede utilizar en las zonas del circuito asignadas y únicamente si el piloto está a menos de 1 segundo de distancia del coche que tiene adelante.
- **Green Flag:** Bandera verde que indica que las condiciones de la pista son seguras y que la sesión puede continuar sin restricciones.

- **Hard:** Compuesto de neumáticos duros que tienen menor agarre pero mayor durabilidad, ideal para *stints* largos.
- **Inlap:** Vuelta en la que un piloto ingresa al *pitlane* para realizar una parada.
- **Medium:** Compuesto de neumáticos medios que equilibran agarre y durabilidad.
- **Outlap:** Vuelta inicial después de salir del *pitlane*, donde los neumáticos necesitan alcanzar su temperatura óptima.
- **Overcut:** Estrategia opuesta al *undercut*, en la que el piloto realiza su parada en boxes después que su rival, buscando ser más rápido durante su *inlap* con neumáticos aún a buena temperatura, mientras su rival enfrenta neumáticos fríos debido a que acaba de salir del *pitlane*.
- **Pitlane:** Zona de la pista donde se encuentran los garajes de los equipos y los *boxes*. Es utilizada para realizar paradas técnicas, como cambios de neumáticos o reparaciones.
- **Pit window:** Período de vueltas durante la carrera en el que los equipos planean realizar paradas en *pits* para maximizar la estrategia.
- **Pits:** Sinónimo de *box*.
- **Red Flag:** Bandera roja que detiene la carrera debido a condiciones peligrosas, como un accidente grave o clima extremo.
- **Safety Car:** Coche de seguridad que entra a la pista para agrupar a los coches y reducir la velocidad en caso de peligro, como accidentes o condiciones inseguras.

- **Single car race:** Estrategia que se hace pensando en que un coche es el único coche en la pista.
- **Soft:** Compuesto de neumáticos blandos que ofrecen mayor agarre pero se desgastan rápidamente.
- **Stint:** Número de vueltas que un piloto realiza con los mismos neumáticos antes de hacer una parada en *pits* o también entre paradas.
- **Undercut:** Estrategia en la que un piloto para en *pits* antes que su rival y trata de ser más rápido en la *outlap* con neumáticos nuevos, aprovechando que su rival aún está en pista con neumáticos desgastados.
- **Virtual Safety Car:** Sistema que reduce la velocidad de los coches en toda la pista mediante un límite de velocidad establecido, sin necesidad de un coche de seguridad físico.
- **Yellow Flag:** Bandera amarilla que indica peligro en la pista, como un accidente, obligando a los pilotos a reducir la velocidad.

Referencias

- Collins, B. (2024). *How to Win a Grand Prix: From Pit Lane to Podium - The Inside Track*. Quercus.
- Dunn, P. K., Smyth, G. K., et al. (2018). *Generalized linear models with examples in R*, volume 53. Springer.
- Formula 1 (2024a). F1 live timing. <https://www.formula1.com/en/timing/f1-live>.
- Formula 1 (2024b). It Ruined His Race: Mercedes Admit to Clear Mistake with Hamilton's Singapore GP Strategy. <https://www.formula1.com/en/latest/article/it-ruined-his-race-mercedes-admit-to-clear-mistake-with-hamiltons-singapore.4VteePMCYiSgKE7NAq9TY6>.
- Formula 1 (2024c). Leclerc Thrills the Tifosi to Triumph at Monza Ahead of Piastri and Norris with Bold Ferrari Strategy Paying Off. <https://www.formula1.com/en/latest/article/leclerc-thrills-the-tifosi-to-triumph-at-monza-ahead-of-piastri-and-norris.1aiYZF3rWZp2Q9yQtuvqV>.
- Formula 1 (2024d). Strategy Guide: What Are the Possible Race Strategies for the 2024 Bahrain Grand

- Prix. <https://www.formula1.com/en/latest/article/strategy-guide-what-are-the-possible-race-strategies-for-the-2024-bahrain.1NT25ROUaq4grnLyJmOojV>.
- Formula 1 (2024e). Verstappen charges to victory over Norris and Perez in action-packed Chinese GP. <https://www.formula1.com/en/latest/article/verstappen-charges-to-victory-over-norris-and-perez-in-action-packed-chinese.3Uz5CwNh5tEQt62umIGhob>.
- Heilmeier, A., Thomaser, A., Graf, M., and Betz, J. (2020). Virtual strategy engineer: Using artificial neural networks for making race strategy decisions in circuit motorsport. *Applied Sciences*, 10(21):7805.
- Magee, J. F. (1964). *Decision trees for decision making*. Harvard Business Review Brighton, MA, USA.
- McCarthy, L. and Rotthoff, K. W. (2013). Incentives on the starting grid in formula one racing. *The Journal of Sport*, 2(2).
- Schaefer, P. (2024). FastF1. <https://docs.fastf1.dev/>.
- Stoppels, E. (2017). Predicting race results using artificial neural networks. Master's thesis, University of Twente.
- Thraves, C. et al. (2022). On the optimization of pit stop strategies via dynamic programming. *Central European Journal Of Operations Research*, 31(1).