



Informe 1

“F1 Data Analytics: evaluación de la competitividad entre equipos mediante técnicas de ciencia de datos”

Diana Luz Hernandez Torres



Objetivos

01

Analizar las brechas de rendimiento entre escuderías a lo largo de las últimas temporadas mediante datasets, con el proposito de identificar patrones, tendencias y factores que influyen en la competitividad entre equipos.

02

Evaluar el rendimiento de las escuderías en clasificación y carrera a lo largo de diferentes temporadas, mediante el análisis de tiempos por vuelta, posiciones finales y consistencia en distintos circuitos.

03

Detectar patrones y técnicos que influyen en las diferencias de rendimiento entre escuderías, considerando variables como los circuitos, condiciones meteorológicas, degradación de neumáticos.

04

Proponer acciones y estrategias que probablemente podrían contribuir a reducir las brechas de rendimiento entre escuderías.

Descripción del problema

El problema radica que en la formula uno es muy habitual que existan brechas entre las escuderías lo que genera dominio por parte de algunos equipos como lo fue mercedes desde 2014-2020, red bull 2021-2024 y actualmente mclaren, esto como espectador llega a ser aburrido.

A pesar de las reglas técnicas y financieras que buscan equilibrar las competencias como el límite en el presupuesto y cambios en el diseño aerodinámico aun así hay diferencias grandes en el rendimiento entre las escuderías.

Estas diferencias se ven reflejadas en los tiempos por vuelta, en los resultados de clasificación y carreras, y en la capacidad de los equipos para mejorar durante la temporada.

También factores como cómo gestionan los neumáticos, las condiciones del circuito, las estrategias para las paradas y las adaptaciones técnicas afectan estas desigualdades.

Hacer un análisis de datos que ayudaría a entender y mostrar las causas y el crecimiento de estas diferencias en rendimiento, y así obtener información útil para entender mejor como se podría mejorar la competencia en la Fórmula uno hoy en día.





Recursos disponibles:

Herramientas:

- Python como lenguaje principal para la manipulación y análisis de la información.
- Pandas para la limpieza, transformación y estructuración de los datasets.
- Matplotlib y Seaborn para la visualización inicial y la creación de gráficos que permitan identificar patrones de rendimiento entre escuderías.
- Numpy para realizar cálculos numéricos y operaciones matemáticas relacionadas con tiempos por vuelta, diferencias de rendimiento y promedios.

Datos:

Se utilizarán 3 datasets de Fórmula 1 obtenidos de Kaggle. Estos contienen información detallada sobre el tipos de climas, clasificaciones, carreras, equipos y pilotos desde temporadas anteriores.

Estos datos permitirán realizar un análisis comparativo entre equipos, identificar patrones de rendimiento por temporada, circuito y condiciones externas, así como visualizar la evolución de las brechas a lo largo del tiempo.

Hipótesis iniciales



01

Los equipos ubicados en la parte alta de la parrilla (como Red Bull, McLaren) mantienen una brecha de rendimiento significativamente constante con respecto a los equipos de mitad de tabla, medida en diferencia promedio de tiempo por vuelta, a lo largo de cada temporada.



02

Las diferencias de rendimiento entre escuderías se amplían en circuitos con condiciones climáticas extremas (altas temperaturas o lluvia), afectando especialmente a los equipos con menor capacidad de adaptación técnica y estratégica.



03

Los equipos que introducen actualizaciones aerodinámicas de forma más frecuente logran reducir progresivamente la brecha de rendimiento con respecto al líder de la temporada en las rondas finales.

Definición de Stakeholders clave

01 Escuderías

Obtendrían información útil para identificar áreas de mejora estratégicas y técnicas, así como patrones de rendimiento en comparación con rivales a lo largo de la temporada.

02 Departamentos de Estrategia y Análisis de Datos

Utilizarían los datos para optimizar la toma de decisiones en carrera como estrategias de parada, ritmo de carrera, uso de neumáticos y para planificar el desarrollo técnico del monoplaza.

03 FIA y Organizadores de la F1

Podrían emplear el análisis para evaluar si las regulaciones técnicas y financieras están cumpliendo su objetivo de reducir brechas de rendimiento entre equipos.

04 Analistas deportivos

Usarían los resultados para explicar de forma más objetiva las dinámicas competitivas, ofrecer análisis más profundos y contextualizar rendimientos en cada Gran Premio.

05 Aficionados de F1

Se beneficiarían al acceder a visualizaciones claras y comprensibles que expliquen por qué ciertos equipos dominan o sufren en determinados circuitos o condiciones, fomentando un mayor entendimiento del deporte.



Preguntas clave

1-¿Cuál es la diferencia promedio de tiempo por vuelta entre los equipos de tabla alta y los de mitad de tabla en cada temporada analizada?

2-¿Cómo varía esta brecha entre clasificación y carrera?

3-¿En qué circuitos se amplían o reducen más estas diferencias?

4-¿Qué impacto tienen las estrategias de pit stop en las brechas de rendimiento finales?

5-¿Cómo afecta la gestión de neumáticos a las diferencias entre escuderías?

6-¿Cómo influyen las condiciones meteorológicas (temperatura, lluvia) en el rendimiento relativo entre equipos?

7-¿Qué equipos muestran una mayor mejora de rendimiento carrera a carrera a lo largo de una temporada?

8-¿Existe correlación entre la frecuencia de actualizaciones técnicas y la reducción de la brecha con el líder?

9-¿Qué variables tienen mayor peso en la explicación de las diferencias de rendimiento (ritmo, estrategia, clima, circuito)?



Preguntas clave

- 10-¿Se pueden identificar tendencias generales en la evolución de las brechas entre temporadas (aumento, disminución o estabilidad)?
- 11-¿Qué equipos muestran mayor consistencia en sus tiempos por vuelta a lo largo de las carreras?
- 12-¿Qué relación existe entre la posición en parrilla y el resultado final según el equipo?
- 13-¿Cómo varía la brecha de rendimiento en circuitos urbanos vs permanentes?
- 14-¿Los equipos de mitad de tabla realizan más o menos paradas que los equipos de tabla alta?
- 15-¿Qué equipos logran mejores tiempos promedio en boxes y cómo influye eso en su posición final?
- 16-¿Se observan diferencias en la eficacia de estrategias undercut/overcut entre equipos grandes y pequeños?
- 17-¿Qué equipos logran acortar brechas en la segunda mitad del campeonato y cuáles las amplían?
- 18-¿Qué impacto tendría una reducción en el presupuesto técnico en la competitividad general?
- 19-¿En qué momento de la temporada suelen aparecer las mayores mejoras de rendimiento por parte de los equipos medios?



Fuentes de datos identificadas

Los datos de los datasets que me van a ayudar a comprobar o negar mis hipótesis son los siguientes:

Generales:

- Datos sobre puntos de la escudería.
- Datos de secciones.
- Datos sobre circuitos.

Rendimiento:

- Datos del tiempo de clasificación.
- Datos de los podios.
- Datos sobre los puntos de pilotos.
- Datos de los puntos de constructores.
- Datos sobre los puntos totales por temporada.

Clima:

- Datos del clima nublado.
- Datos del clima lluvioso.
- Datos del clima calido.
- Datos del clima frio.
- Datos del clima seco.
- Datos sobre la altitud y latitud de la ciudad donde se corre.



Justificación

Analizar las diferencias de rendimiento es importante para entender la Fórmula 1. Al medir cuánto varía el rendimiento entre equipos en varias temporadas, el proyecto ofrece una visión general sobre si las escuderías líderes siempre dominan o si hay momentos y factores que ayudan a las escuderías más rezagadas a igualar el campo de juego.

Al observar cómo varía la diferencia, se puede ver si la competencia se vuelve más igualada o se divide en extremos. Esto afecta a reguladores como la FIA, que ajustan reglas como el límite de presupuesto o cambios en aerodinámica para mantener el interés y la sorpresa en la carrera. Los datos ayudarán a ver si estas reglas funcionan bien.

El impacto del clima y la altitud es importante para entender qué tan fuerte es la ingeniería de un equipo. Si la diferencia aumenta en condiciones extremas (lluvia, alta altitud), esto indica una debilidad en el diseño del chasis, en el sistema de refrigeración o en la estrategia de configuración. Este factor hace que la competitividad pase de ser solo una cuestión de velocidad a una cuestión de adaptación y manejo de riesgos. Esto es muy importante para el desarrollo técnico de cualquier equipo.

Al medir cómo mejoran los equipos a lo largo de una temporada (el progreso en su desarrollo técnico), el análisis ayuda a entender qué tan eficiente es ese proceso. Los resultados permiten a un equipo en mitad de la parrilla saber si sus actualizaciones están funcionando o si necesitan cambiar su enfoque.

En resumen, este proyecto ayuda a obtener una idea clara de los patrones que determinan cuál equipo rinde más. Esto proporciona información útil a equipos, analistas y a los responsables del campeonato lo que ayuda a generar una fórmula 1 más competitiva.

¿Cuántos datos son y que tipo de datos son?

En total por los tres datasets son 14,701 que se clasifican de la siguiente manera:

- Identificadores Categóricos: season, round, circuit_id, driver, constructor, year, Driver, Car.
- Identificadores Geográficos: circuitId, name, location, country.
- Métricas de Resultado: podium, driver_points, driver_wins, constructor_points, constructor_wins.
- Métricas de Rendimiento: qualifying_time.
- Métricas de Clasificación: Pos (Posición final), PTS (Puntos totales de la temporada).
- Variables Categóricas: weather_warm, weather_cold, weather_dry, weather_wet, weather_cloudy
- Variables Físicas: alt (Altitud).
- Variables Geográficas: lat (Latitud).

