

Análisis exploratorio de datos



THE BRIDGE DIGITAL
TALENT
ACCELERATOR

Bootcamp Data Science Online, Mayo 2024

Luis Tamayo Ortiz

Indice

1. Introducción pag. 1

1.1 Contexto pag. 1

1.2 Hipotesis pag. 2

2. Tratamiento de datos pag. 4

2.1 Exploración inicial y limpieza pag. 4

2.2 Unión de dataframes pag. 5

3. Análisis de los datos pag. 6

3.1 Repetitividad de victoria de un mismo piloto en una temporada

pag.6

3.2 Adelantamientos por temporada pag. 10

3.3 Influencia de la lluvia en los adelantamientos pag. 12

3.4 Influencia de los abandonos (DNF – Do Nor Finsish) en los

adelantamientos pag. 15

3.5 Influencia del circuito en los abandonos pag. 18

3.6 Estudio Circuito – Propuesta de Calendario F1 pag.23

1 | Introducción

Se va a realizar un análisis exploratorio de datos sobre como la Formula 1 ha ido cambiando durante los últimos 40 años.

Se analizará diferentes datos para obtener una visión más clara sobre el estado actual de la Formula 1 en cuanto a la capacidad de generar atracción para el nuevo aficionado y de retener al ya fidelizado.

El análisis se basará en la comparación estadística entre diferentes periodos, así como la repercusión de los cambios normativos.

1.1 | Contexto

Para el análisis se han usado datos provenientes de diferentes fuentes.

Todas las fuentes no disponían de la misma resolución en cuanto al rango de años que abarcaban.

Por este motivo, aunque inicialmente se iban a analizar los 75 años de historia de la F1, se optó por centrarse únicamente en los últimos 40 años, por disponer de una mejor y abundante cantidad de datos para poder ser trabajados.

Los datos manejados durante el EDA:

Fuente de datos 1

- **driver.csv:** datos sobre los pilotos de F1.
- **races.csv:** datos sobre los grandes premios de F1.
- **circuits.csv:** datos sobre los circuitos de F1.
- **status.csv:** listado de códigos sobre como la F1 interpreta el estado en el que un piloto finalizo un gran premio.
- **driver_standings.csv:** datos sobre los resultados de los pilotos.

Fuente de datos 2

- **overtakes.csv:** datos sobre los adelantamientos de cada Gran Premio en la F1.
- **overtaking_season_average.csv:** datos generales sobre adelantamientos en F1.

Fuentes de datos 3

- **weather_overtake_bs_completed.csv:** tabla de datos creada mediante código durante el EDA. Se extrae el dato de si la carrera fue declarada en mojado o en seco.

1.2 | Hipótesis

La Formula 1 es el máximo exponente del deporte de motor y durante los últimos 40 años ha sufrido innumerables cambios de equipos, pilotos, circuitos y reglamentos técnicos.

Sobre todo, el último punto es el que la FIA (Federación Internacional del Automóvil) más atención le pone, pues siempre se intenta que las normas favorezcan la igualdad entre escuderías, proporcionando temporadas emocionantes para el aficionado.

El reglamento técnico tiene un peso en la ecuación que garantiza Grandes Premios emocionantes y llenos de incertidumbre, pero no es único factor.

Otros factores influyen en el desarrollo de un Gran Premio y estos van a ser analizados para poder encontrar las claves que hacen que un Gran Premio sea o no un éxito para el aficionado.

Por todo ello, la hipótesis o pregunta a responder es:

¿Es la Formula 1 actual más atractiva que la F1 anterior?

Para poder dar la respuesta más clarificadora posible, se van a contestar a las siguientes subpreguntas:

¿Cuántas veces un mismo piloto ha ganado Grandes Premios en una misma temporada?

- Es entendible que una temporada en la cual más pilotos se disputen la victoria tendrá un impacto directo en la atracción del público objetivo.
- Se analizará la relación de los pilotos con más victorias en la misma temporada para entender si es posible tener

¿Cuántas adelantamientos ha habido por temporada?

- Los adelantamientos son la esencia de la F1. Cuantos más de ellos haya, más impredecible será el resultado final y más momentos de interés generar para el aficionado.
- Se tendrán en cuenta el número de circuitos de cada temporada para obtener un valor medio.
- Se estudiarán los cambios normativos que se ejecutaron en pro de mejorar este aspecto.
- Esta pregunta desemboca ineludiblemente en:

¿Qué circuitos son más propensos a tener un número mayor de adelantamientos?

- Los circuitos, que es donde se juega la "partida", a diferencia de otros deportes, son muy diferentes entre sí, lo que implica que tienen un peso en la ecuación, del resultado y desarrollo del Gran Premio, muy alto.

¿Cuántos piloto han podido finalizar el Gran Premio?

- Como en otros deportes, no siempre acaban los mismos que empiezan.
- El mundo del automovilismo tiene unos hándicaps, además del propio piloto, como lo es la parte mecánica, el Formula 1.
- Los Formulas 1 sufren mecánica y electrónicamente durante cada Gran Premio, lo que ocasiona fallos que impiden que el piloto pueda finalizar el Gran Premio.
- Hay otra pregunta relacionada con esta que es bastante determinante, y es la de saber si:

¿Qué influencia tiene el tiempo atmosférico en la cantidad de adelantamientos que hay en un gran premio?

- El clima es determinante a la hora de cambiar radicalmente las condiciones de agarre en la pista.
- Si bien el proveedor de neumáticos proporciona variantes específicas para este fin, la introducción de un nuevo elemento a la ecuación, el agua, debe de ser estudiado para poder entender que efecto tiene.

Todas estas hipótesis se han basado en un entendimiento extenso por parte del autor del EDA sobre el “negocio”, en este caso la Formula 1.

Como es sabido, un EDA debe de fundamentarse en el conocimiento del negocio donde se va a aplicar la exploración, pues de lo contrario, podrían pasarse por alto conclusiones importantes o la no relación de observaciones con otros factores propios del negocio, los cuales podrían no a ver sido estudiados en este EDA.

Durante el desarrollo del EDA, podrán aparecer otras hipótesis resultado de las conclusiones y por tanto no estar listado en los párrafos iniciales de esta sección.

2 | Tratamiento de datos

Los datos obtenidos de las diferentes fuentes van a ser analizados para poder acondicionarlos para el posterior análisis.

Debido al alto número de archivos diferentes con datos, para la limpieza de estos se ha creado un Jupyter Notebook, específico.

De esta manera no se condensa toda la información en un solo Jupyter Notebook, lo que dificultaría la lectura del EDA.

Jupyter Notebook - Analisis_inicial_&limpieza de datos

```
df_drivers.shape - (859, 4)
df_races.shape - (1125, 7)
df_circuits.shape - (77, 8)
df_status.shape - (139, 2)
df_drivers_standings.shape - (34595, 7)
df_overtakes.shape - (780, 5)
df_overtakes_avegare.shape - (40, 8)
df_results.shape - (26519, 18)
```

2.1 | Exploración inicial y limpieza

- **drivers**
 - No presenta NaN
 - No presenta duplicados
 - Eliminación de columnas
 - Creacion de columnas
 - Concatenacion de columnas
 - Conversion string a fecha
- **races**
 - No presenta NaN
 - No presenta duplicados
 - Eliminacion de columnas
 - Conversion string a fecha
- **circuits**
 - No presenta NaN
 - No presenta duplicados
 - Eliminacion de columnas
- **status**
 - No presenta NaN
 - No presenta duplicados
- **drivers_standigs**
 - No presenta NaN
 - No presenta duplicados

- **overtakes**
 - No presenta NaN
 - No presenta duplicados
 - Eliminación de filas con datos que no van a ser usados en el EDA
 - Creación de columnas y asignación de valores para el filtrado
- **overtaking_average**
 - No presenta NaN
 - No presenta duplicados
- **results**
 - No presenta NaN
 - No presenta duplicados
- **weather_overtake_bs_completed**
 - Dataframe creado a propósito del EDA.
 - este data frame ha sido creado mediante la búsqueda de una palabra clave dentro del texto de una columna
 - No presenta NaN
 - No presenta duplicados

*Si durante el transcurso del proyecto se evidenciara algún problema con los datos del data set, este sería revisado y modificado.

2.2 | Unión de dataframes

Para poder realizar las consultas adecuadas, varios data frames se han unido entre sí:

- Creación data frame “**df_F1**”

```
df_F1 = pd.merge(df_results, df_races, on = 'raceId', how = 'left')
df_F1 = pd.merge(df_F1, df_circuits, on = 'circuitId', how = 'left')
df_F1 = pd.merge(df_F1, df_drivers, on = 'driverId', how = 'left')
df_F1 = pd.merge(df_F1, df_status, on = 'statusId', how = 'left')
df_F1 = pd.merge(df_F1, df_constructors, on = 'constructorId', how = 'left')
```

- Debido a la gran cantidad de uniones de data frames que ha habido que hacer, no se van a listar pues no hay un aporte real a la memoria y si una compleja interacción. En los Jupyter Notebooks que acompañan al análisis se ha detallado cada paso para un correcto seguimiento de las interacciones entre data frames.

3 | Análisis de los datos

3.1 | Repetitividad de victoria de un mismo piloto en una temporada.

Los datos muestran que la media de victorias de un mismo piloto en una temporada es del **47.4%**.

Eso quiere decir, que casi la mitad de las carreras es siempre ganada por un mismo piloto.

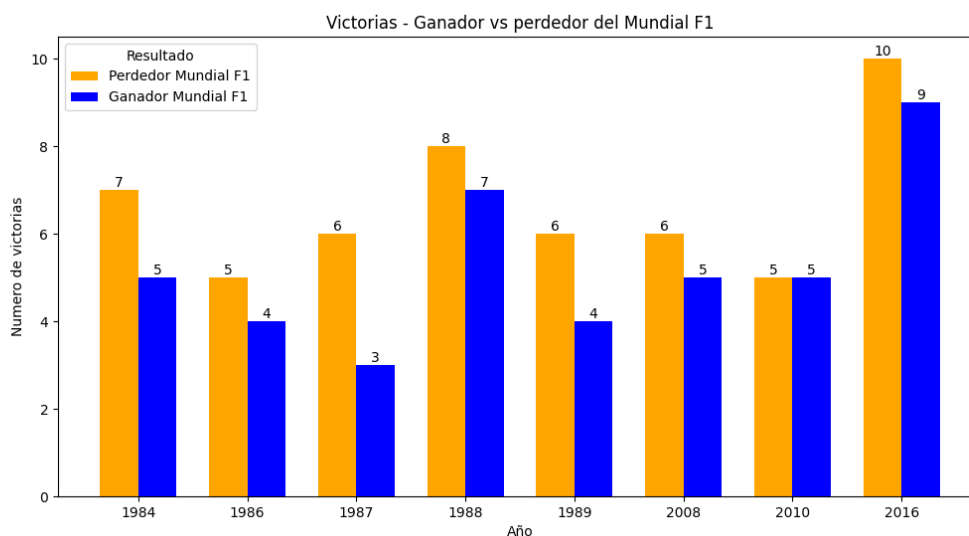
Se analiza si el piloto que más veces ha ganado es también el que fue campeón en esa temporada, para entender si hay una relación real entre el número de carreras ganadas y la consecución del campeonato.

La lógica diría que sí, pero para eso están los datos.

En los últimos 40 años, solo en 8 ocasiones, un piloto que no ha sido el que más carreras ha ganado dentro de una temporada, se ha coronado vencedor.

Por lo que se puede decir que el **80% de las veces**, el piloto que más grandes premios gana suele ganar el campeonato.

Las temporadas donde ocurrió fuero: 1984, 1986, 1987, 1988, 1989, 2008, 2010, 2016.

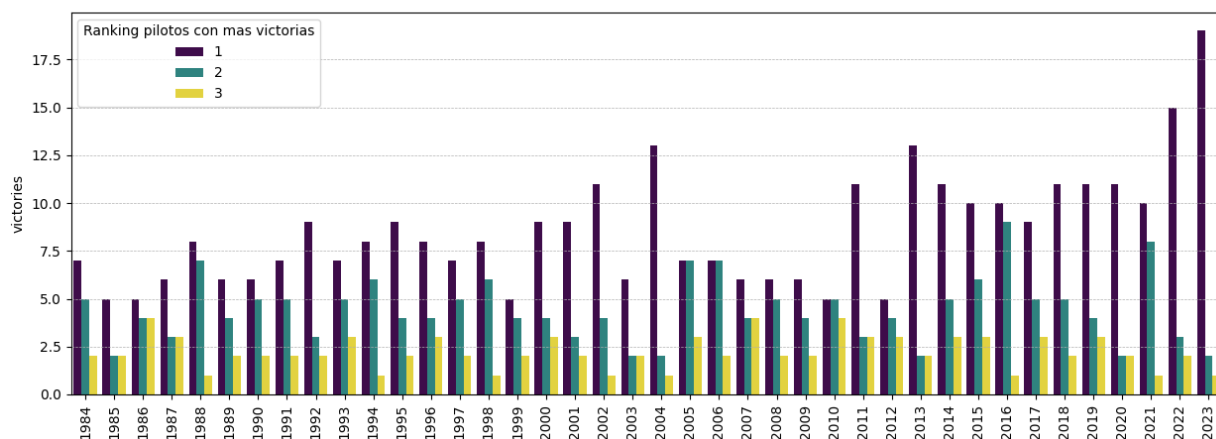


Teniendo en cuenta ambos porcentajes, si durante el transcurso de un campeonato, alguno de los pilotos ha ganado ya el 50% de las carreras aun no habiendo finalizo este, la probabilidad de que ese piloto gane, aun no ganando más carreras desde ese momento es del 80%.

Esto hace que los campeonatos parecieran más o menos atractivos cuando uno de los pilotos alcance esa cifra, el 50 % de carreras ganadas.

Cuanto más se retrase este hecho, más incertidumbre habrá sobre el piloto ganador y por ende mayor atracción para el público.

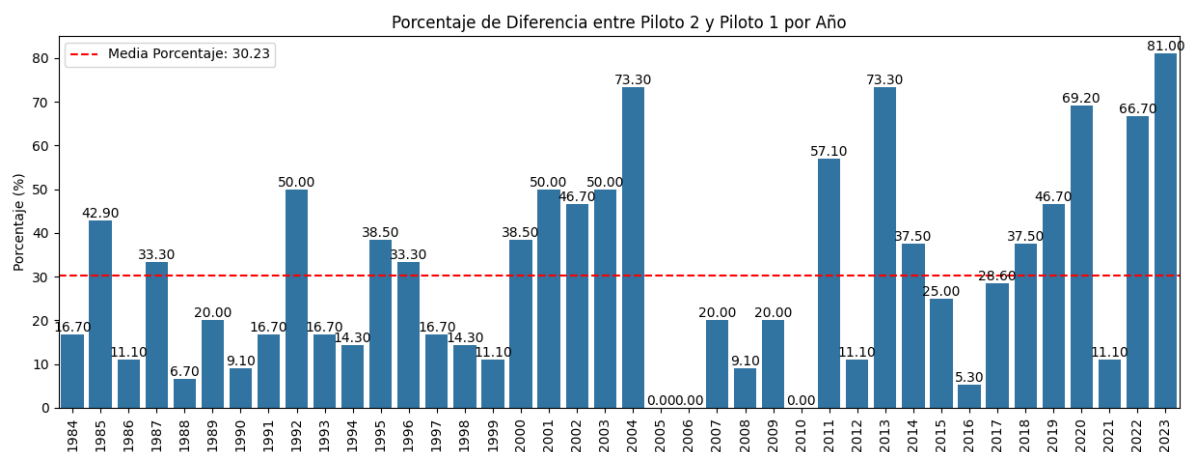
Para entender mejor este último hecho, se ha graficado el número de victorias de los 3 pilotos que más han tenido en cada temporada.



Cuanto mayor sea el parecido entre en la altura de las barras verdes y moradas, de una misma temporada, más disputado fue el campeonato.

Se aprecia bastante irregularidad en los datos, áreas temporales donde sí que hay paridad entre las columnas y zonas donde el piloto que gano más carreras está muy por delante del segundo.

Para poder entender mejor este comportamiento se creó el siguiente gráfico:



Aquí se tiene el dato de la diferencia en carreras ganadas en %, entre el piloto con más victorias (Piloto 1) y el segundo (Piloto 2). Se ha marcado la media de los últimos 40 años.

Desde 1984 hasta 1999 los números nos hacen ver que había bastante igualdad a la hora de repartirse los triunfos entre los dos mejores pilotos.

A partir del año 2000 se aprecia una hegemonía creciente hasta el año 2005, momento en el cual, parece que comienza el mejor momento de la F1 en cuanto a la rivalidad entre pilotos, acabando este ciclo en 2011.

A partir de ese momento, comienza tendencias positivas y negativas en la evolución de las diferencias entre los dos pilotos con más victorias.

Si bien es cierto, el pico máximo lo encontramos en el último año del gráfico, 2023, donde se consigue el récord de victorias de un mismo piloto con Max Verstappen, ganador de 19 Grandes Premios de 22 posibles.

Conclusión:

Aun no habiendo un patrón claro en cuanto al reparto de carreras ganadas con el transcurso de los años, sí que la Formula 1 actual tiene periodos más prolongados de hegemonía de un piloto.

El ideal sería que siempre hubiese muchos pilotos diferentes ganadores de Grandes Premios, o al menos dos de ellos que se repartieran estas victorias.

Por lo tanto, la Formula 1 actual sí que es menos atractiva en cuanto a cómo se reparten las victorias y, por lo tanto, el resultado final de los campeonatos.

Es una Formula 1 más predecible.

Análisis histórico de los acontecimientos durante el periodo estudiado

1987-1999 → Época con cambios en las regulaciones de los motores (prohibición de motores turbo en 1988), se permitió el control de tracción, aparecieron los primeros cambios semiautomáticos y hubo cambios en la distribución del peso de los vehículos y en las especificaciones de los neumáticos. Estos son grandes cambios que generaron la gran competitividad de esa época.

2000-2004 → Combinación de un coche superior (F2002/F2004), un equipo sólido y estabilidad normativa. La estabilidad normativa permitió a Ferrari desarrollar su vehículo con gran éxito. Ferrari es un equipo que dispone de un circuito al lado de la factoría donde se fabrican los F1. Esto, en una época donde los túneles del viento no eran tan normales, le dio mucha ventaja sobre el resto de los equipos.

2005-2010 → Introducción de neumáticos únicos, cambio en las regulaciones aerodinámicas (Ferrari no encontró el compromiso entre velocidad punta y agarre en las curvas, el coche era lento en recta), nuevas regulaciones sobre combustible y limitaciones técnicas que afectaron negativamente a Ferrari.

Renault (Fernando Alonso) tenía un nuevo motor V10 muy fiable y potente, además de que introdujo un sistema revolucionario para mejorar la estabilidad del coche.

Fue prohibido en 2007, momento en el que Renault dejó de ganar.

En 2009 Brawn GP (antigua Honda) presentó el doble difusor. Su rendimiento era tal que todos los equipos lo copiaron y la FIA en 2010, limitó su diseño mediante el reglamento.

En 2009 también se introdujo el KERS (Kinetic Energy Recovery System), un sistema que recuperaba energía de las frenadas y que el piloto podía usar en el momento que él considerara. Quizá este punto ayudó a incrementar los adelantamientos en 2010 (los equipos necesitan tiempos para comprender el uso de sistemas tan novedosos).

En 2010 el KERS se siguió usando, lo que mejoró los adelantamientos.

En 2010, Red Bull junto con su distribuidor de motores Renault, desarrollaron el difusor soplado, mediante el cual, cuando el coche estaba en la fase de entrada a curva (piloto frenando), el motor seguía expulsando gases de escape sobre el difusor, aumentando la carga aerodinámica.

Este hecho hizo que Red Bull ganara 4 campeonatos seguidos.

En **2010** y **2012**, Fernando Alonso estuvo muy cerca de ganar el título, pero fallos de estrategia le privaron de ello.

2011 → Red Bull sigue mandando, su coche es muy superior al resto.

Se introduce el sistema DRS (Drag Reduction System). Este sistema permite cambiar la carga aerodinámica del alerón trasero cambiando su ángulo de ataque. Gracias a él, los fórmulas 1, incrementan su velocidad punta, pudiendo rebasar a otros pilotos con más facilidad.

Se siguió usando el sistema KERS.

También en este año se introdujeron los neumáticos Pirelli, los cuales estaban diseñados para sufrir degradación, no como los antiguos Bridgestone, los cuales duraban prácticamente toda la carrera.

Estos tres factores fueron determinantes para que se consiguiera el récord de adelantamientos medios por temporada en Formula 1.

2012 → En este año hubo cambios aerodinámicos que empeoraron la capacidad de poder seguir un coche cerca del otro sin sufrir las turbulencias generadas por el coche de delante.

Este hecho sumado a que los equipos, viendo la degradación de los neumáticos, optaron por gestionar el desgaste en vez de intentar ir los más rápido posible, genero una bajada en los adelantamientos.

La introducción en el campeonato de nuevos circuitos con curvas complicadas y con zonas no optimizadas para el uso del DRS, consiguió que los adelantamientos fuera cayendo

2014-2015 → Los motores de formula pasan a ver híbridos y V6. Este hecho pone en cabeza a Mercedes con Lewis Hamilton.

Mercedes fue la que mejor interpreto el reglamento para fabricar estos motores. Había incluso diferencias de 80 cv entre el motor Mercedes y otros motoristas.

2016 – Un año donde Mercedes siguió siendo el equipo que batir, pero esta vez fue Nico Rosberg el que se llevó el título. Fue una temporada muy emocionante donde Nico gano sin ser el piloto que más victorias obtuvo.

2017-2020 – Mercedes seguía mandando de la mano de Lewis Hamilton.

2021- Honda junto con Red Bull pudieron finalmente alcanzar a Mercedes, con un campeonato

2022 – 2023 – La FIA presenta el cambio de normativa más grande en las últimas décadas.

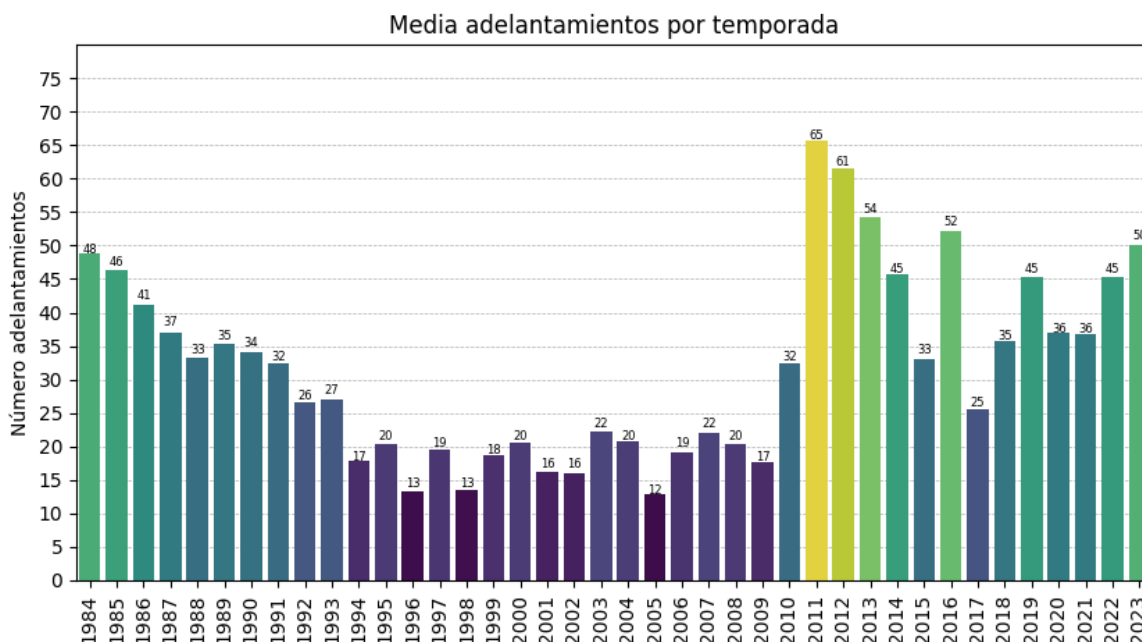
Se vuelve a usar el efecto suelo como recurso aerodinamico para mejorar las turbulencias y así mejorar los adelantamientos.

Red Bull, de la mano de Max Verstappen y con el mejor ingeniero en aerodinámica del, consiguieron hacer el coche de los récords, consiguiendo ganar en 19 de las 22 carreras disputadas.

Posiblemente el campeonato donde menos emoción hubo por ver quien sería el ganador,

3.2 | Adelantamientos por temporada

Tras el análisis de los adelantamientos, se ven 3 bloques diferenciados.



1984-1993 → Tendencia negativa. → 35,9 de media por temporada.

1994-2009 → Valores estabilizados, pero al nivel más bajo de los últimos 40 años. → 17,75 de media.

2010-2023 → Zona con la mayor media de adelantamientos, aunque bastante irregular. → 43,85 de media.

Solo con este análisis se podría decir que los adelantamientos no tienen relación con el avance de los años en la F1, aun así, se van a analizar lo sucedido en los años en los que hubo un cambio de tendencia para poder entender si fue condicionado por algún acontecimiento.

1994 – 2010 – se podría decir que ambos periodos son el mismo, lo único que hubo una estabilización en los adelantamientos durante 15 años. Se podría decir que los valores de ese periodo son seguramente el límite que tiene la F1 y que más bajos es complicado alcanzarlos.

En toda esta etapa hay 3 factores que destacan sobre el resto que:

- Aumento de la carga aerodinámica, impidiendo que el coche que persigue pueda acercarse sin entrar en aire turbulento.
- Aumento del tamaño de los Formula 1. Se hace más difícil batallas en paralelo.
- Neumáticos con muy poca degradación que no favorecían los adelantamientos ni las estrategias desde boxes.

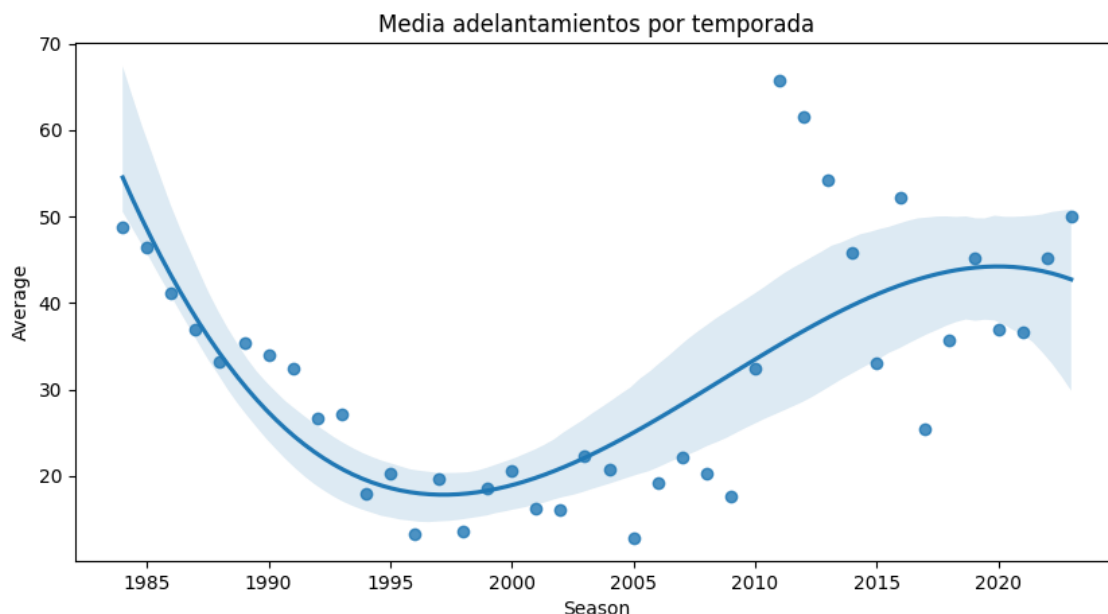
2011 – En este punto se introduce el sistema DRS, que permite aumentar la velocidad punta cuando se activa. Gracias a él, el aumento de adelantamientos tuvo un crecimiento impresionante, aunque muchos de los adelantamientos son en rectas debido a este gran diferencial de velocidad.

Desde entonces hasta el año **2017**, los monoplazas fueron aumentando dotados de mayor tamaño y peso, así como de neumáticos más anchos, que permitían ir más rápido en las curvas y frenar más tarde, lo que perjudicó notablemente los adelantamientos.

Dejando de lado el repaso histórico de los hechos y volviendo a los datos, se aprecia que la incertidumbre de los datos que crean la media de adelantamientos de cada temporada es creciente conforme avanza el eje de ordenadas (temporadas).

Este hecho nos podría estar diciendo, que, aunque el dato absoluto es mayor, quizá haya más diferencia entre número total de adelantamientos entre carreras, lo que podría provocar carreras con muchos adelantamientos y carreras con pocos.

Para el público medio, ver tanta diferencia entre adelantamientos según qué Gran Premio se corra, no genera atracción.



Conclusión:

Los adelantamientos han mejorado en números absolutos tomando como referencia el último periodo de la F1 (**2010-2023**)

1984-1993 → Una mejoría del **18.12 %** respecto de este periodo

1994-2009 → Una mejoría del **59.57 %** respecto de este periodo.

Se puede decir que la F1 actual mantiene un nivel optimo de adelantamientos.

Análisis estadístico - Adelantamientos vs. temporada

Para comprobar si estadísticamente hablando hay relación entre los adelantamientos y el avance temporal de los años se ha usado el Test de Pearson para tal efecto:

Variables comparadas:

- Media adelantamientos
- Temporadas F1

Coefficiente de correlación: 0.26904411947288936, Valor p: 0.09320268134315272

El valor es cercano a 0,05 pero no lo suficiente, no presenta relación estadística el avance temporal y los adelantamientos

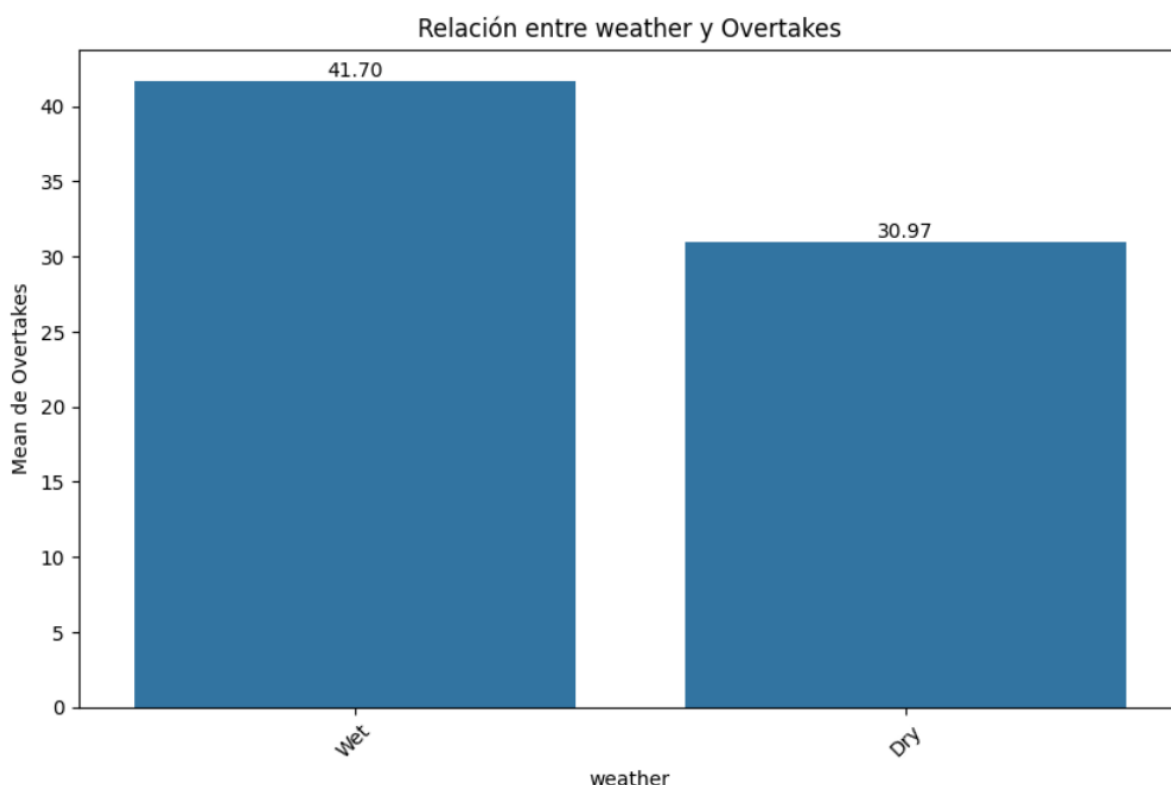
3.3 | Influencia de la lluvia en los adelantamientos

El aspecto que más pueda cambiar el desarrollo de una carrera de Formula 1 es saber si esta se corre en seco o en húmedo.

Los pilotos y equipos se ven expuestos a mayor presión, pues hay un elemento más en juego, el agua, y este cambia la manera de pilotar, así como la toma de decisiones por parte de los equipos.

Volviendo al análisis, de media, una carrera con lluvia genera casi **42** adelantamientos, al contrario que una carrera en seco, que solo aporta **31** adelantamientos.

Esto es más de un **26%** de adelantamientos.



Estos datos son bastante reveladores de que las carreras en húmedo son carreras, al menos, desde el punto de vista de los adelantamientos, más entretenidas.

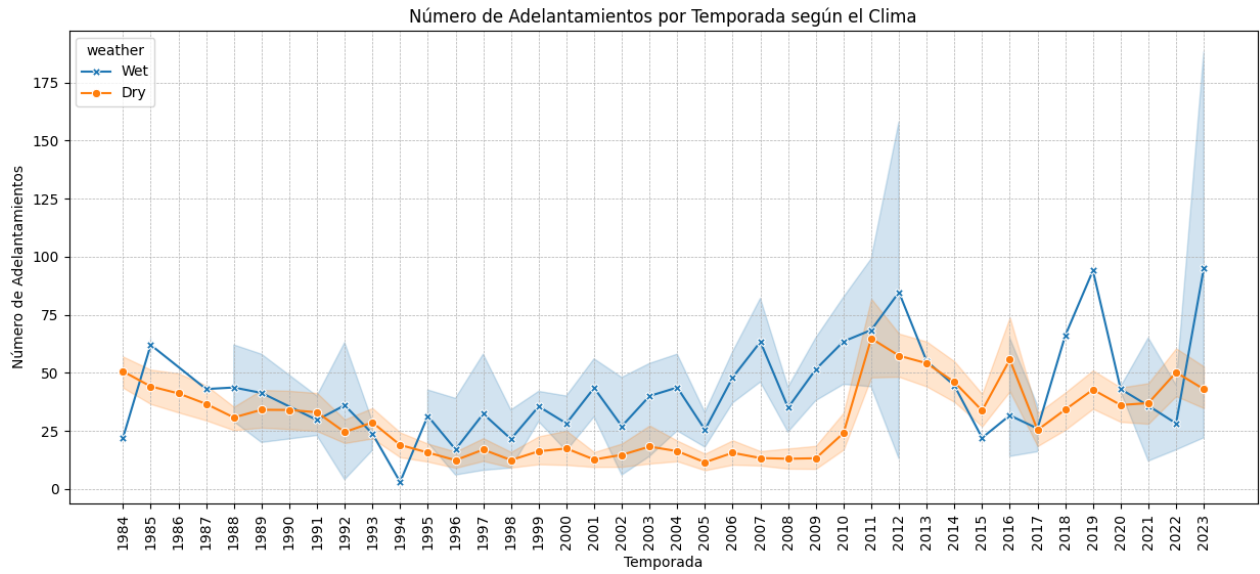
Vamos a ver durante el transcurso de las temporadas como ha ido evolucionando este dato.

El gráfico de abajo muestra el total de adelantamientos en seco y húmedo que ha habido cada temporada.

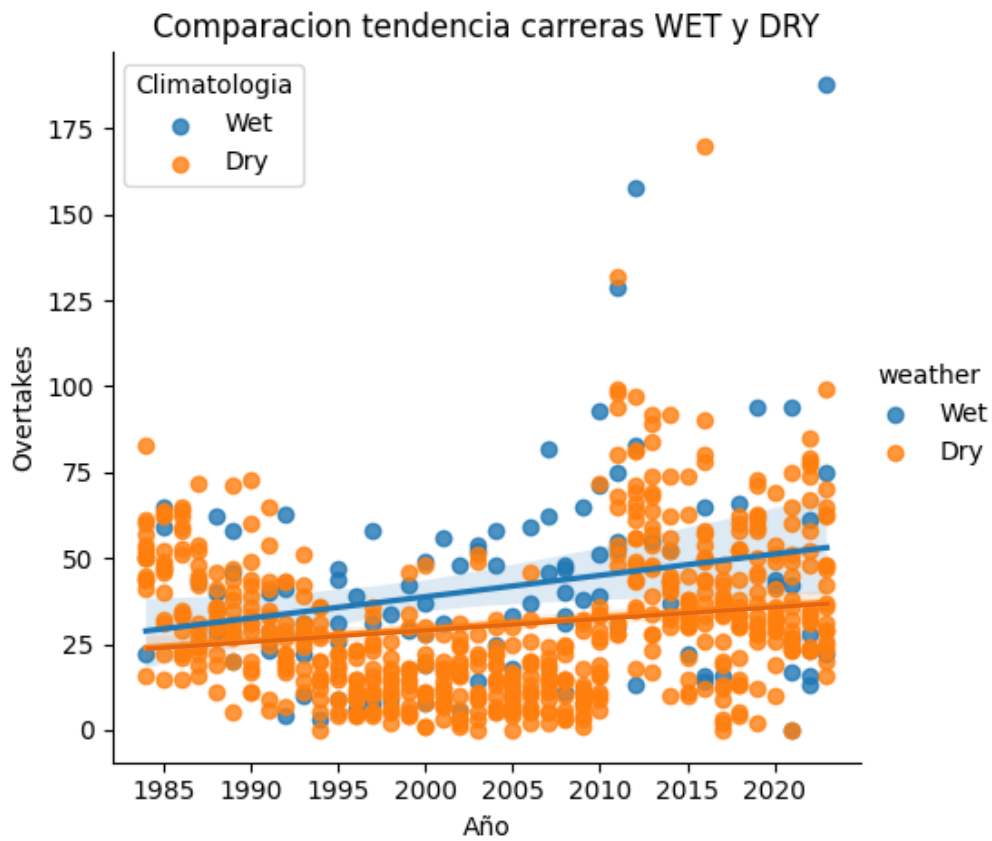
Se ve como en el **82%** de las temporadas, ha habido más adelantamientos en húmedo que en seco.

Si bien, el aumento de adelantamientos en seco a partir del año 2010, aumento el rango, haciendo que la diferencia entre seco y húmedo no fuera tan grande.

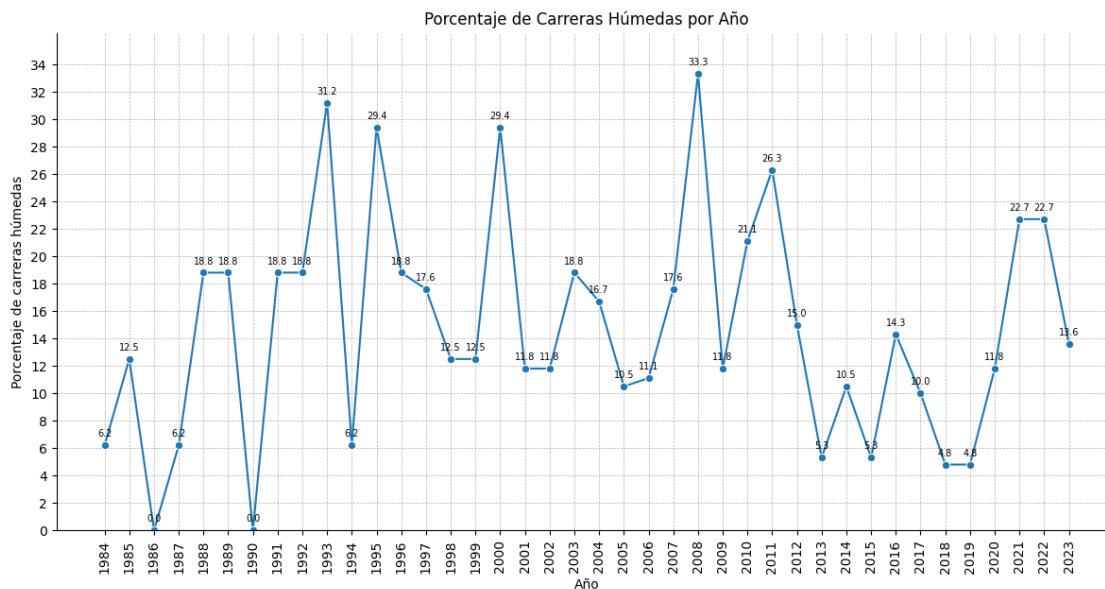
Los puntos altos de la línea de adelantamientos en húmedo se pueden deber a que, según la intensidad de la lluvia, la rapidez con que esta aparece y el grado de intermitencia que esta pueda tener, pueden determinar momentos de adelantamientos o cambios de posición, que en una carrera en seco sería prácticamente imposible de ver.



Aquí un gráfico que muestra más claramente esa tendencia alcista de los adelantamientos y como debido a los grandes picos del grafico anterior (recordemos que las carreras en húmedo son más imprevisibles) la pendiente de la recta de adelantamientos en húmedo es mayor.



Se va a analizar si el número de carreras declaradas en húmedo y seco también ha cambiado mucho durante el transcurso de las temporadas, por ver si quizá este hecho también es determinante en el aumento de adelantamientos.



1984-1993 → 21 carreras en húmedo → 2,1 carreras en húmedo por temporada

1994-2009 → 55 carreras en húmedo → 3,4 carreras en húmedo por temporada

2010-2023 → 38 carreras en húmedo → 2.7 carreras en húmedo por temporada

Es destacable ver como el periodo con menos adelantamientos fue el que más carreras en húmedo tuvo.

Por lo tanto, el factor lluvia aumenta los adelantamientos dentro de un mismo año, pero un mayor número de carreras con lluvia por temporada no aumentan el número total de adelantamientos.

Por lo tanto, hay otros factores más determinantes, causantes del bello arte de adelantar.

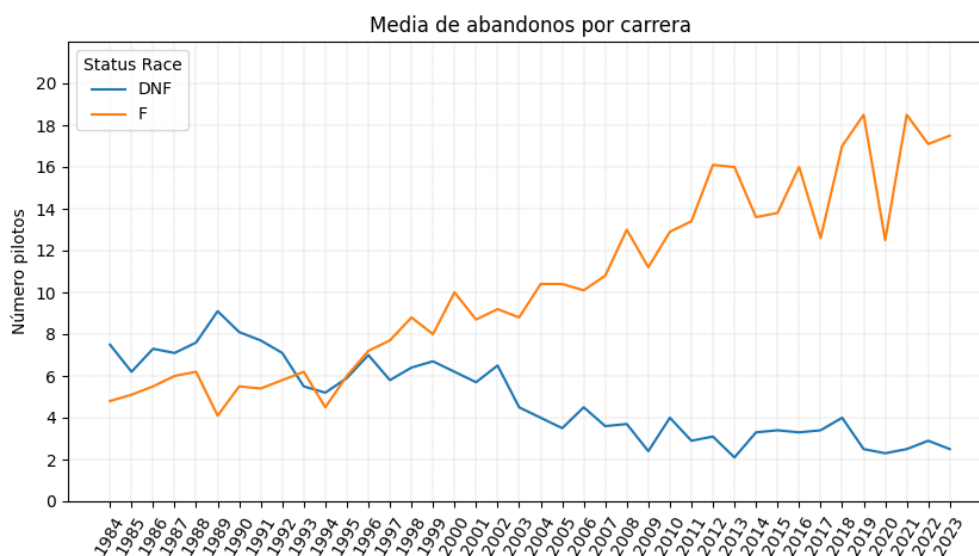
3.4| Influencia de los abandonos (DNF – Do Not Finish) en los adelantamientos

Llegamos al punto donde otro factor debe de ser estudiado para entender si tiene o no relación con el hecho de aumentar los adelantamientos.

El gráfico de abajo muestra la evolución en tanto por cien de los pilotos que no (DNF) acabaron la carrera contra los pilotos que si (F) acabaron la carrera.

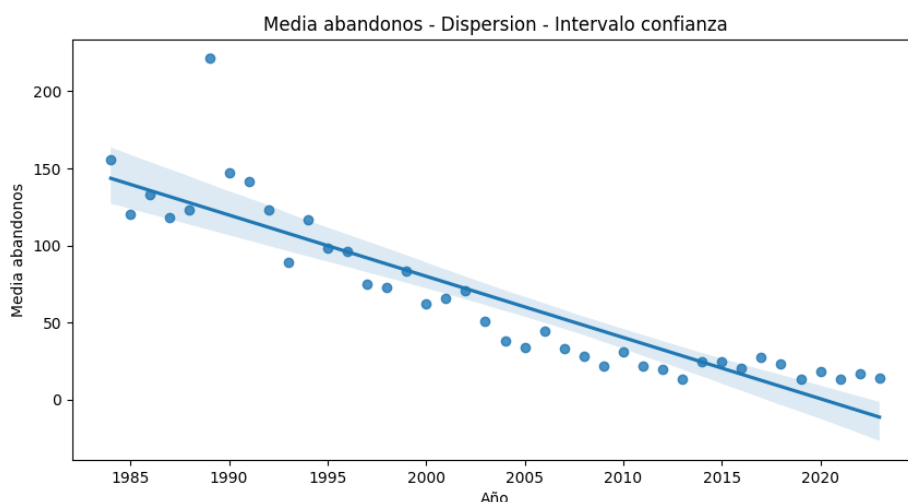
Es bastante alarmante ver la situación de la F1 donde hasta el periodo que comienza en el año 1995 hasta el 2023, había más pilotos que no acaban la carrera, que pilotos que si lo hacían.

A remarcar el año 1989 donde hay la mayor diferencia en este aspecto



Se aprecia muy claramente como la tendencia de abandonos si que tiene relación directa con el avance de los años.

Tiene toda la lógica pues la fiabilidad de los coches de la Formula 1 va de la mano de la tecnología y la tecnología del avance de los años.



Conforme han ido pasando los años, los procesos productivos más capaces, la introducción de nuevos materiales, la introducción de nuevas tecnologías de fabricación, la introducción de la

telemetría y el control en tiempo real de los parámetros óptimos de funcionamiento han conseguido que a día de hoy un Formula 1 tenga la misma fiabilidad que un coche de calle, si nos atenemos a cuál es el uso y el desgaste para el cual fueron diseñados ambos.

Análisis estadístico - Abandonos vs. temporada

Para comprobar si estadísticamente hablando hay relación entre los abandonos y el avance temporal de los años se ha usado el Test de Pearson para tal efecto:

Variables comparadas:

- % Abandonos
- Temporadas F1

Coeficiente de correlación: -0.901165743941513 , Valor p: $2.272636985071103e-15$

Aquí sí que se ve que se rechaza la hipótesis nula, es decir, SI que hay relación entre las variables.

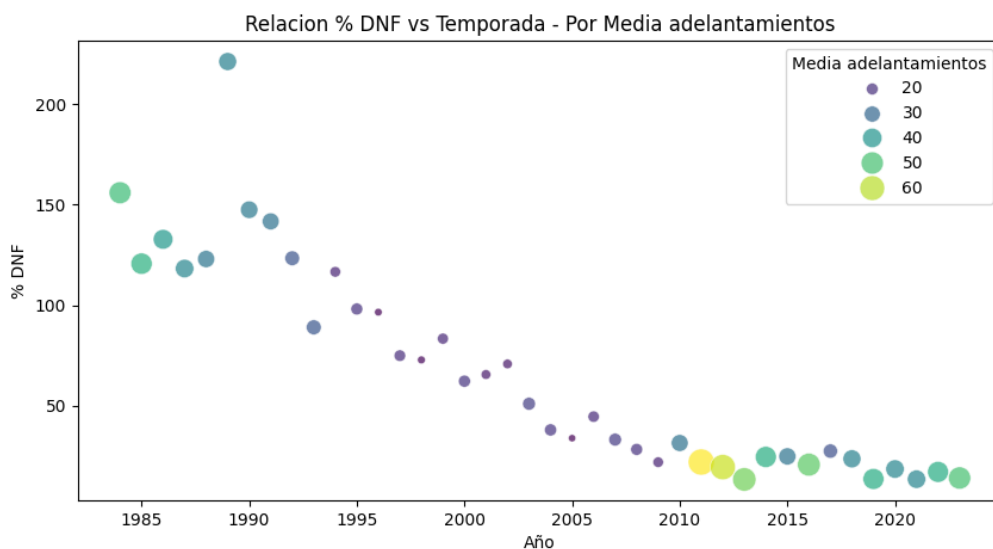
El eje temporal define claramente la reducción en DNF.

Después del análisis de la evolución de los adelantamientos a lo largo de los años, es necesario vincularlos al número de adelantamientos, para ver si hay o no relación entre ellos.

La primera gráfica de abajo muestra claramente como no hay una relación directa.

Si nos fijamos sobre todo en la zona media del gráfico, vemos como el numero de abandonos disminuyen notablemente, pero no hay un cambio en el tamaño/color de los puntos (adelantamientos).

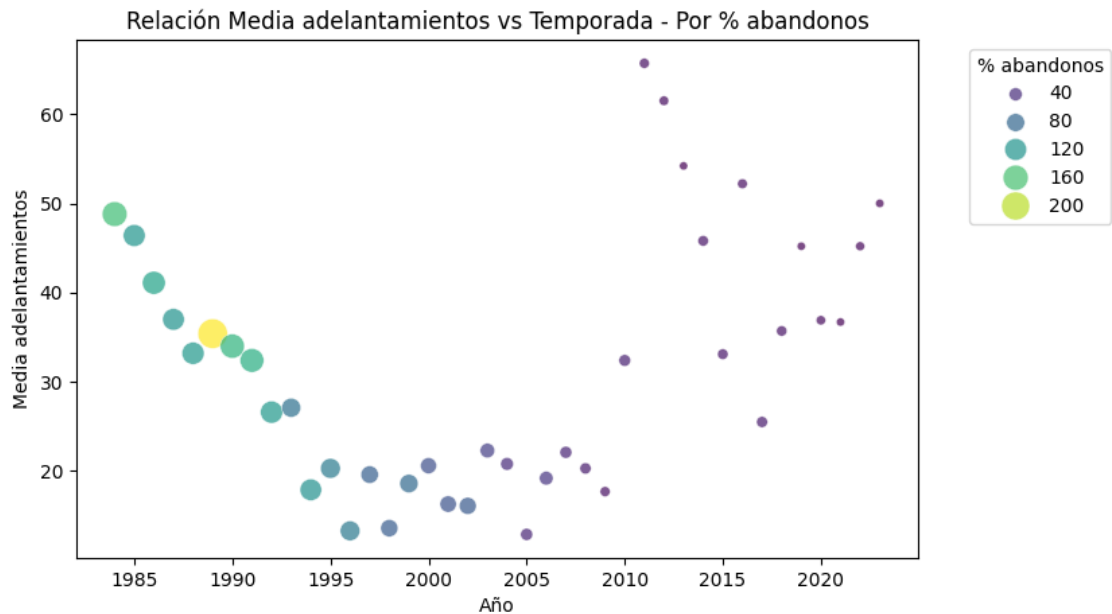
Luego, tanto el primer tercio como el último, presentan puntos de color y tamaño, que sin ser semejantes son muchos mayores que la parte central del gráfico, por lo que el numero de abandonos no tiene ningún tipo de relación con los adelantamientos, tiene que haber otra variable que tengas más peso en esa ecuación.



También se ha calculado poniendo como tercera variable a los DNF para ver como se comportaba la gráfica de adelantamientos.

Se observa cómo hay puntos de todos los tamaños/colores en todo el rango de “Media de adelantamientos”.

Por lo tanto, tener más o menos adelantamientos no es relativo al número de pilotos que no acaban la carrera, sino que depende del momento histórico en el que se mire.



Análisis estadístico - Adelantamientos vs. Abandonos

Para comprobar si estadísticamente hablando hay relación entre los abandonos y el avance temporal de los años se ha usado el Test de Pearson para tal efecto:

Variables comparadas:

- Media adelantamientos
- % Abandonos

Coefficiente de correlación: -0.14536166736537515 , Valor p: 0.3708074036471338

No hay relación estadística entre la media de adelantamientos con el % de abandonos.

3.5 | Influencia del circuito en los abandonos

Siguiendo el hilo del análisis, se va estudiar como de decisivo es el circuito donde se corre sobre el número de abandonos.

Para ello se han cogido 4 grupos de 3 años consecutivos de la siguiente manera:

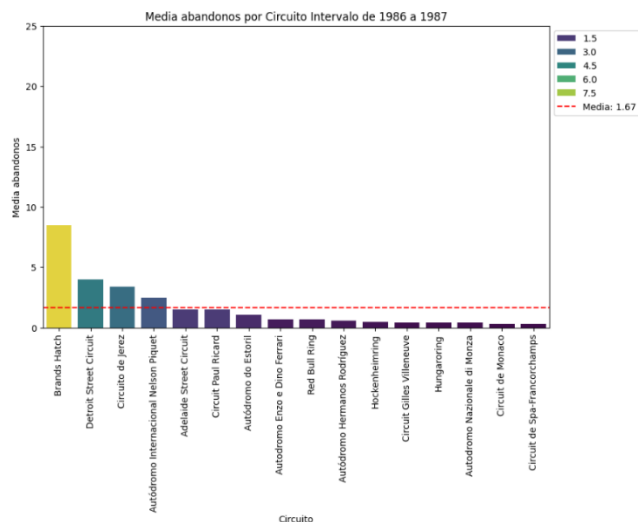
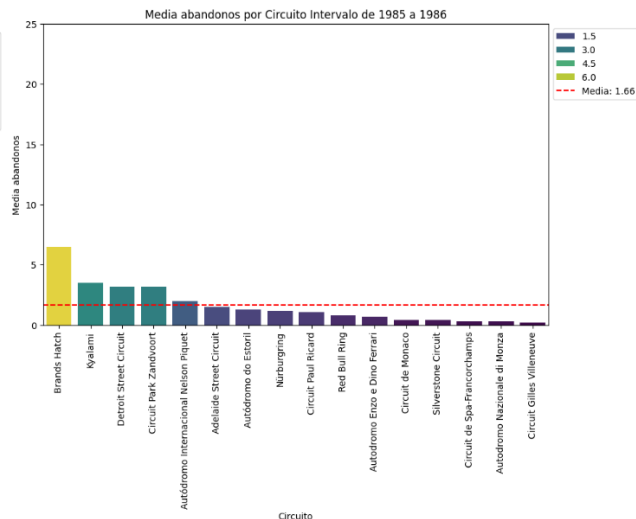
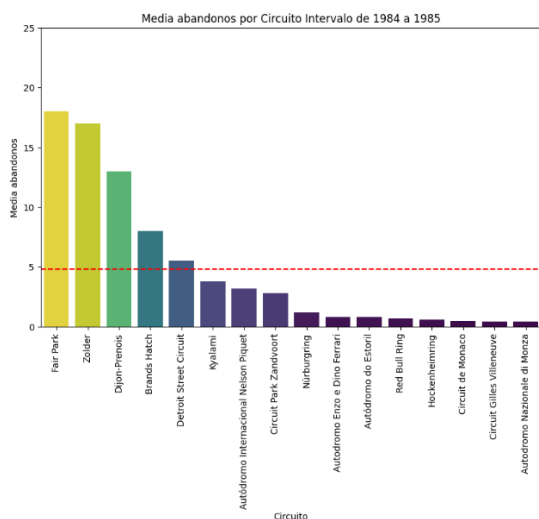
De 1984 a 1986 / De 1994 a 1996 / De 2004 a 2006 / De 2014 a 2016 / De 2021 a 2023

Con ello se quiere identificar si son siempre los mismos circuitos los que se encuentran en la zona más alta en cuenta a abandonos.

Periodo 1984 a 1986:

En 1984, los 3 circuitos con más abandonos no aparecen en 1985, pero si quitamos esos 3 circuitos los otros 3 circuitos siguientes son de 1984 son los 3 primeros circuitos de 1985.

En 1986, dos de los 3 circuitos con más abandonos repiten respecto a 1985, por lo que se puede confirmar que el circuito sí que tiene influencia en el número de abandonos.

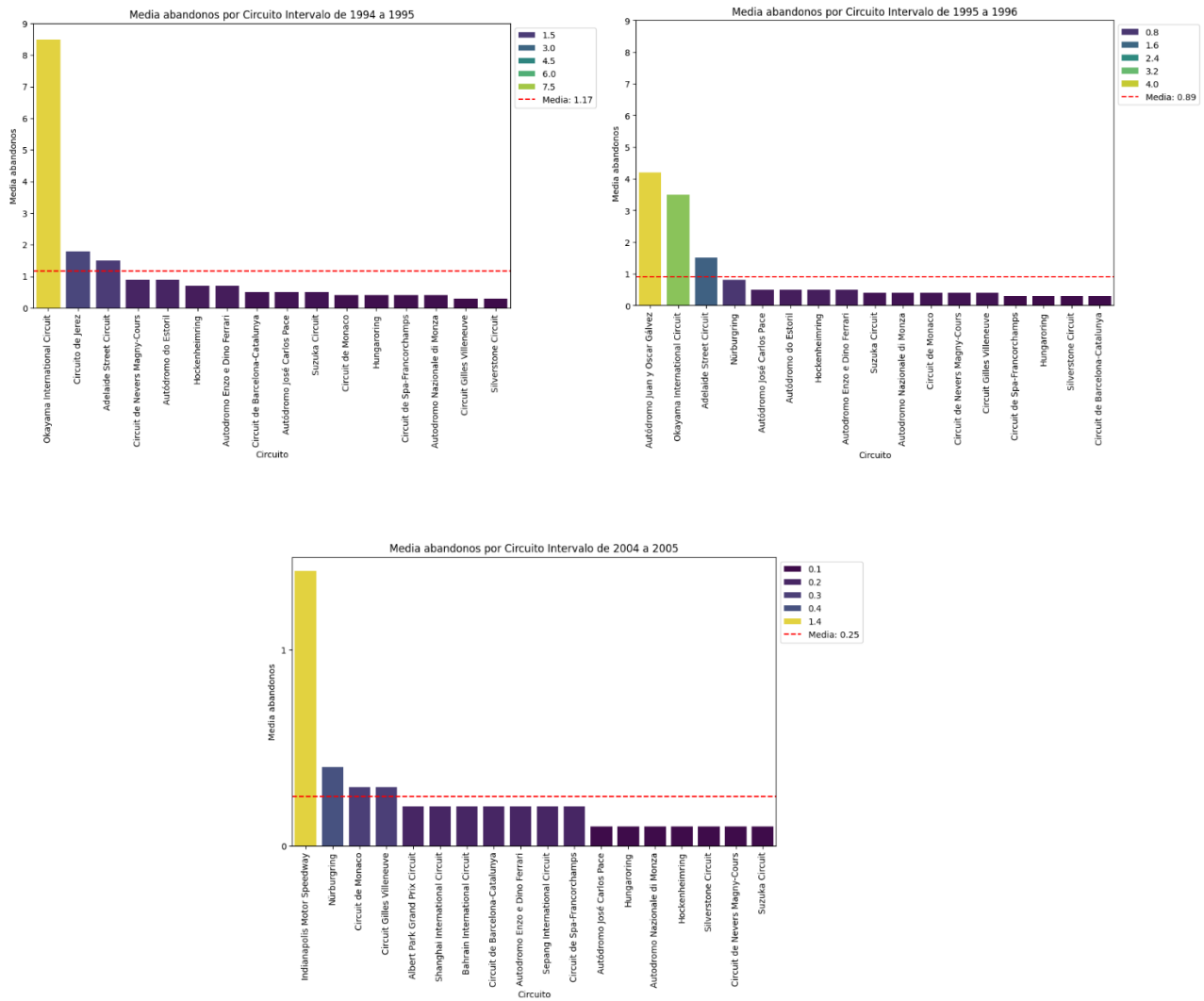


Periodo 1994 a 1996:

En 1994 y 1995, dos de los tres circuitos con mas abandonos son los mismos.

En 1996, aparece en primer lugar el mismo circuito que en 1995. Además, dos de los circuitos con mas abandonos de 1995 no están en la temporada 1996, pero el cuarto circuito con mas abandonos de 1995 es el segundo en el ranking de 1996.

Se confirma que los circuitos con mas abandonos lo son siempre.



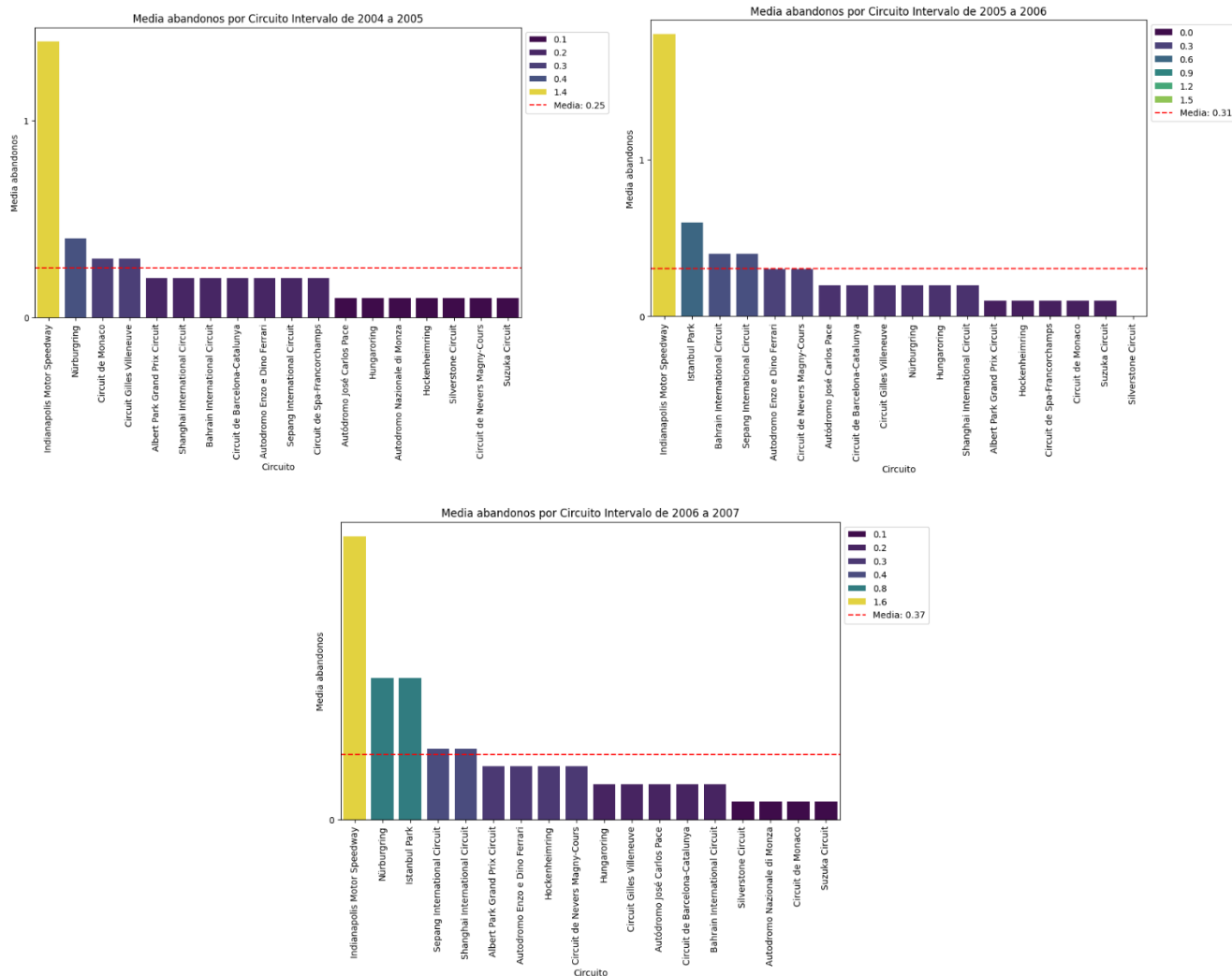
Periodo 2004 a 2006:

En 2004 y 2005 dos de los tres circuitos con mas abandonos, repiten en cada año.

Lo mismo pasa entre 2005 y 2006, donde dos circuitos repiten.

Además, añadir que en los tres años consecutivamente esta en la primera posición de abandonos el mismo circuito.

Se confirma de nuevo la teoría de que el circuito tiene relación directa con el número de abandonos.



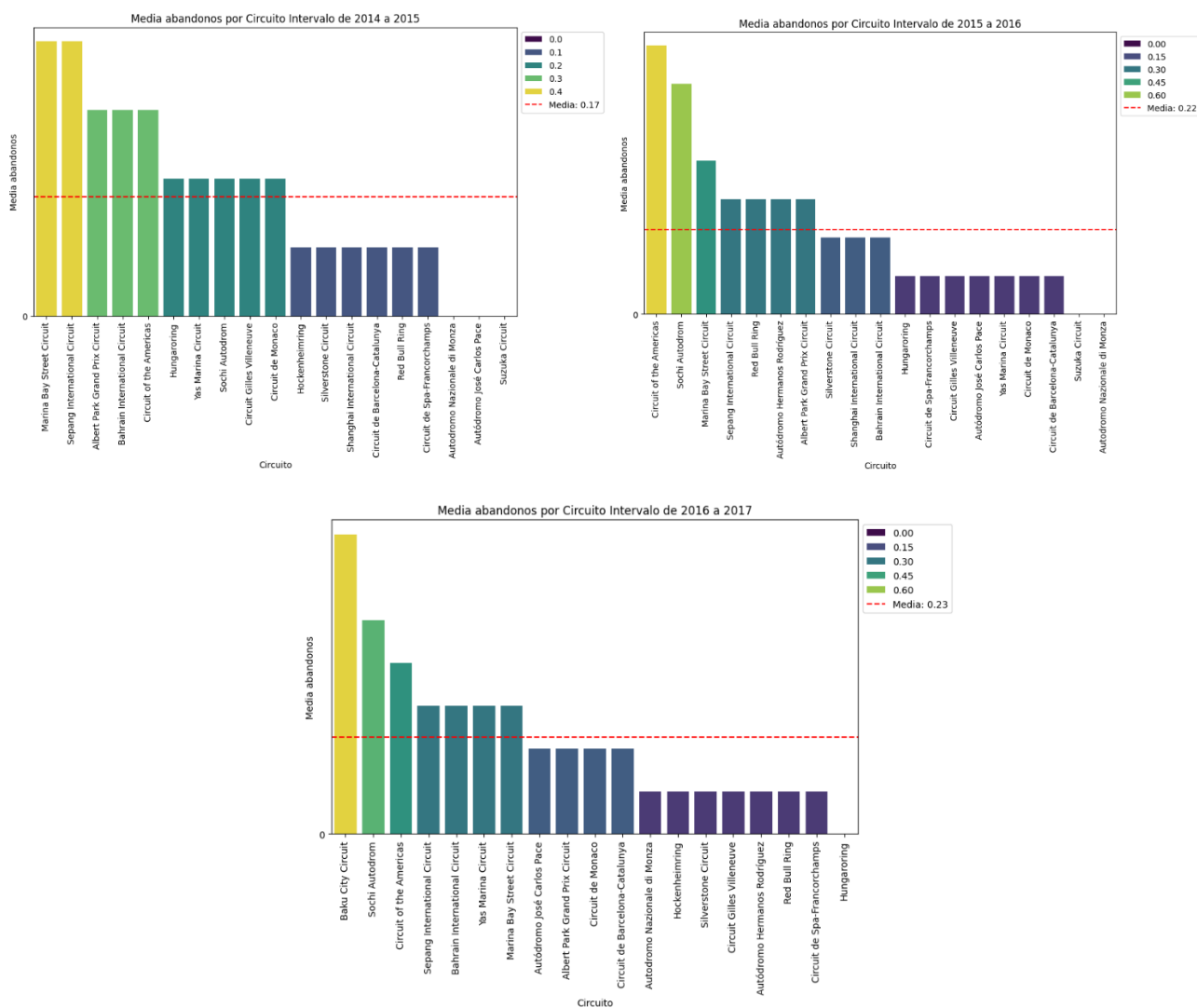
Periodo 2014 a 2016:

En 2014 el primer circuito por abandonos es el tercer circuito en 2015.

Además, en los 5 primeros puestos se van intercambiando 3 de estos circuitos entre el año 2014 y 2015.

En 2016 las 4 primeras posiciones son las mismas 4 posiciones que en 2015, solo hay intercambio entre el primero y el tercer circuito.

Esto reafirma todavía más la hipótesis de que el circuito es muy diferencial a la hora de haber abandonos.



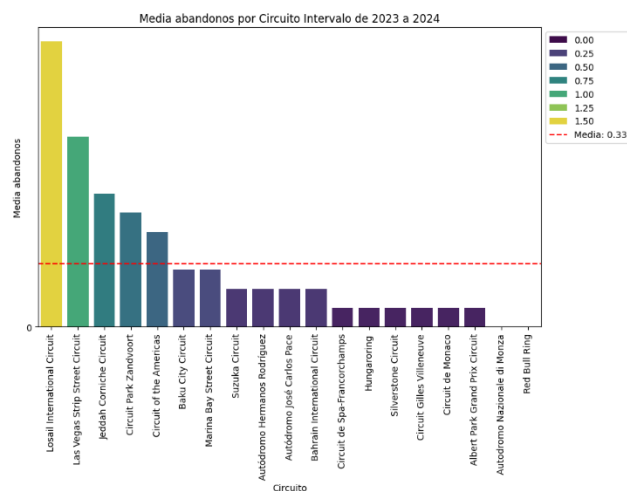
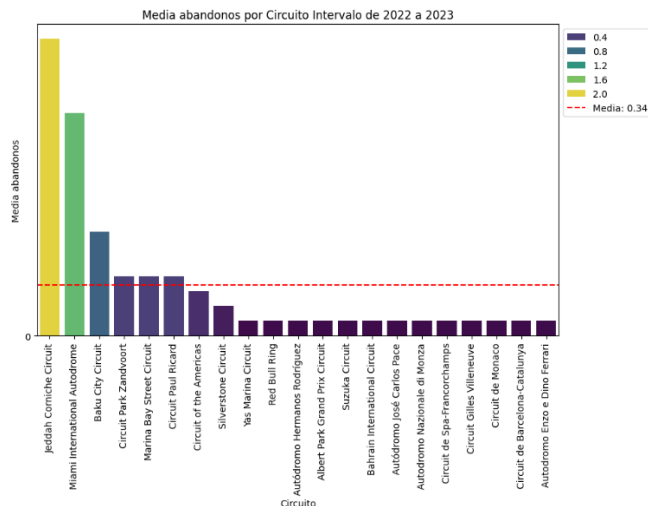
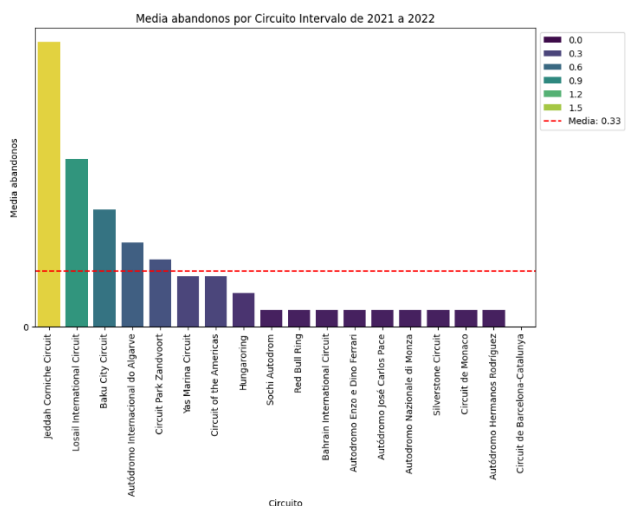
Periodo 2021 a 2023:

De 2021 hay dos circuitos en el top 3 compartidos con 2022.

Luego 2021 comparte dos circuitos también con 2023.

Es decir, en los 3 años, estudiados, hay repetición de los mismos circuitos en posiciones altas.

Tras el análisis de 5 fragmentos de historia de la F1 formada por 3 años cada una, se confirma sin lugar a dudas de que el circuito es el factor mas determinándote para los abandonos.



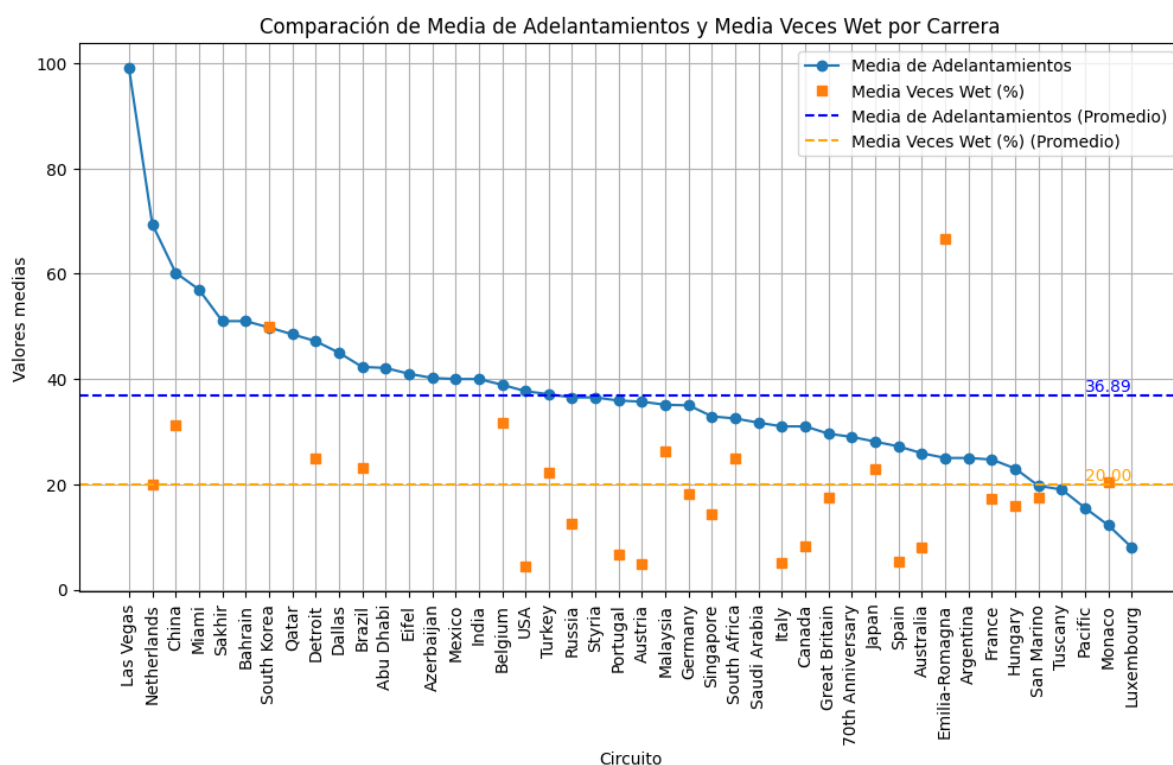
3.6| Estudio Circuitos – Propuesta de Calendario F1

Como último punto de estudio, se han analizado los circuitos donde más adelantamientos ha habido, teniendo en cuenta las participaciones de cada uno.

De esta manera obtenemos un listado de circuitos más propensos a los adelantamientos, y a la vez se ha adjuntado al gráfico, la probabilidad de que llueva.

Esto nos da una guía de cuál sería el calendario de F1 más atractivo en cuanto a acción en pista y a la probabilidad de lluvia.

Se han sacado los valores medios, para tener una referencia de que circuitos están por encima de la media, tanto en adelantamientos como en probabilidad de que llueva.



Propuesta calendario F1

Las Vegas	Mexico
Netherlands	India
China	Belgium
Miami	USA
Sakhir	Turkey
Bahrain	Malaysia
South Korea	Germany
Qatar	South Africa
Detroit	Japan
Dallas	
Brazil	
Abu Dhabi	
Eifel	
Azerbaijan	

