

Analisis Exploratorio de Datos sobre Fórmula 1

Aliendro Maite

Introducción

En este estudio se analiza un conjunto de datos referido a campeonatos de la *Fórmula 1*, comprendidos entre los años 2018 y 2023. Se cuenta con las siguientes cuatro tablas.

Tabla 1. `carreras.csv`

Variable	Descripción
<code>idCarrera</code>	Identificador de la carrera.
<code>año</code>	Año de realización de la carrera.
<code>ronda</code>	Orden de la carrera dentro del campeonato (año).
<code>nombre</code>	Nombre de la carrera.
<code>fecha</code>	Fecha de realización de la carrera.
<code>hora</code>	Hora de realización de la carrera.
<code>tempAire</code>	Temperatura atmosférica media durante la carrera (en grados Celsius).
<code>tempPista</code>	Temperatura media de la pista durante la carrera (en grados Celsius).
<code>lluvia</code>	Indicador de presencia de lluvia durante la carrera.

Tabla 2. `conductores.csv`

Variable	Descripción
<code>idConductor</code>	Identificador del conductor.
<code>codigo</code>	Código asociado al conductor.
<code>nombre</code>	Nombre de pila del conductor.
<code>apellido</code>	Apellido del conductor.
<code>fechaNac</code>	Fecha de nacimiento del conductor.

Variable	Descripción
<code>nacionalidad</code>	Nacionalidad del conductor.

Tabla 3. `constructores.csv`

Variable	Descripción
<code>idConstructor</code>	Identificador de la escudería.
<code>nombre</code>	Nombre de la escudería.
<code>nacionalidad</code>	Nacionalidad de la escudería.

Tabla 4. `resultados.csv`

Variable	Descripción
<code>idResultado</code>	Identificador del resultado.
<code>idCarrera</code>	Identificador de la carrera.
<code>idConductor</code>	Identificador del conductor.
<code>idConstructor</code>	Identificador de la escudería.
<code>posInicial</code>	Posición del conductor al inicio de la carrera.
<code>posFinal</code>	Posición del conductor al final de la carrera.
<code>milisegundos</code>	Tiempo hasta la finalización de la carrera (en milisegundos).
<code>estado</code>	Estado de finalización (completada, accidente, descalificado, otro).

En base a estos datos, se responden los interrogantes enunciados en la sección *Desarrollo*.

Desarrollo

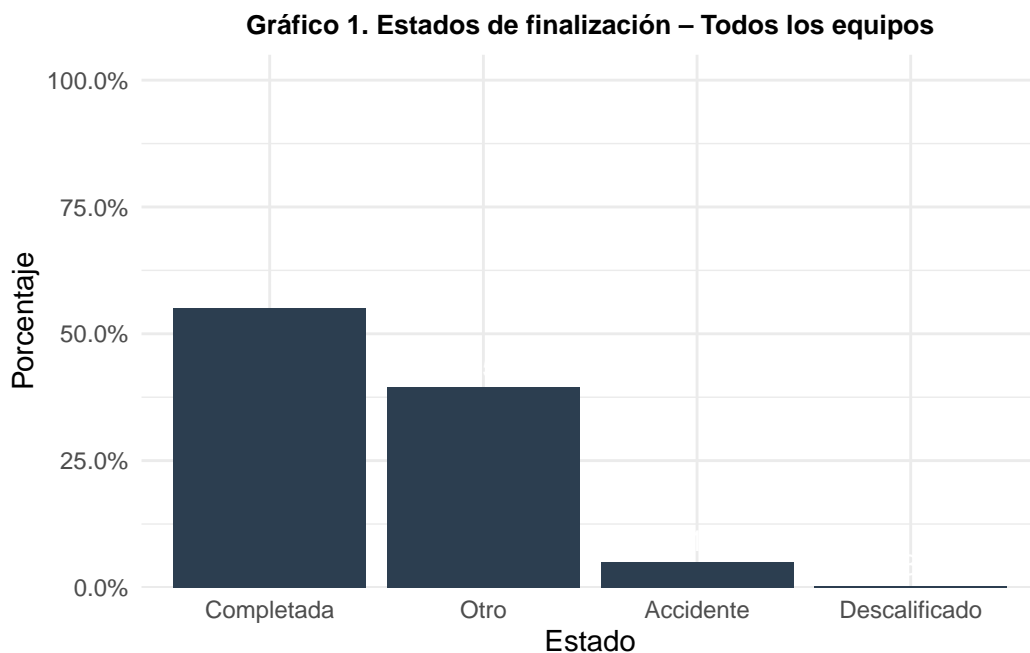
Estado de finalización de los corredores.

En esta sección se analiza cómo finalizan los corredores de todas las carreras, de todos los equipos y luego; puntualmente, al equipo Mercedes . Se estudia la variable cualitativa “estado de finalización”, clasificada en categorías como “completada”, “accidente”, “descalificado”, “otro”.

Estado de finalización general.

Tabla 1. Estado de finalización – Todos los equipos

Estado	Frecuencia	Porcentaje (%)
Completada	1378	55.1
Otro	987	39.5
Accidente	127	5.1
Descalificado	8	0.3
Total	2500	100.0

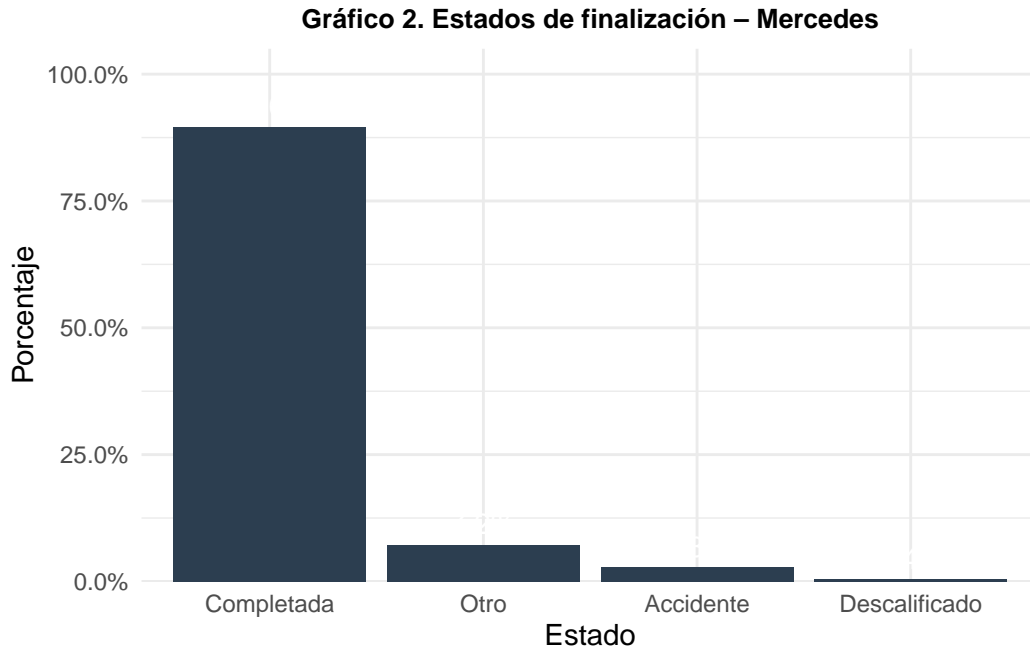


En el análisis general, se observa que aproximadamente el 55% de las carreras finalizan en forma exitosa, es decir, que el corredor completa la carrera. El resto de los casos presenta diversas situaciones que impiden la finalización, destacándose un 39.5% en la categoría “Otro”, que incluye abandonos por fallos mecánicos, problemas técnicos, entre otros motivos distintos a accidentes o descalificaciones. Los accidentes representan el 5.1% del total. Por otro lado, las descalificaciones constituyen un porcentaje pequeño (0.3%), por lo que tienen poco impacto en el panorama general.

Estado de finalización del equipo Mercedes.

Tabla 2. Estado de finalización – Mercedes

Estado	Cantidad	Porcentaje (%)
Completada	224	89.6
Otro	18	7.2
Accidente	7	2.8
Descalificado	1	0.4
Total	250	100.0



Al focalizar el análisis en el equipo Mercedes, se aprecia un desempeño superior en términos de finalización de carrera. En este caso, el 89.6% de las carreras terminan exitosamente, casi duplicando el porcentaje observado en el análisis general. Asimismo, el porcentaje de accidentes para Mercedes es mucho menor (2.8%). La categoría “Otro” también presenta un porcentaje reducido (7.2%), lo que indica que Mercedes experimenta menos problemas técnicos o abandonos por causas diversas en comparación con el promedio general. La tasa de descalificaciones permanece casi nula (0.4%).

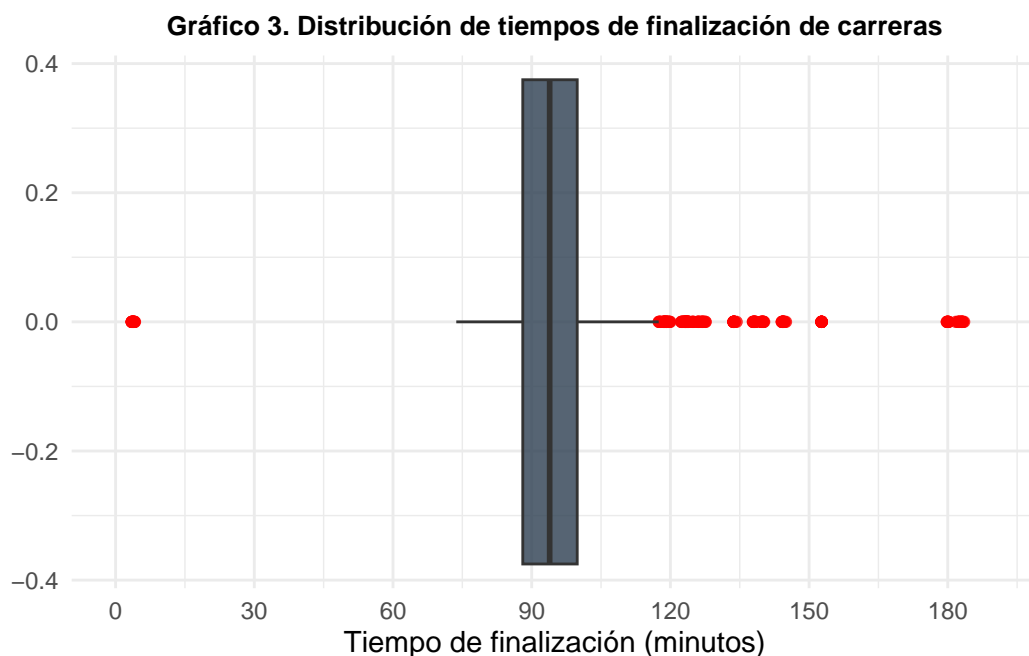
En conclusión, el análisis comparativo evidencia que Mercedes supera el desempeño promedio en cuanto a finalización exitosa de carreras, accidentes y abandonos técnicos, consolidándose como un equipo con una tasa elevada éxito en la Fórmula 1.

Tabla 1. Estadísticas para detección de valores atípicos en tiempos de finalización

Estadística	Valor (minutos)
Q1 (Primer cuartil)	88.08
Q3 (Tercer cuartil)	99.84
Rango intercuartílico (RI)	11.77
Límite inferior	70.42
Límite superior	117.50
Tiempo mínimo observado	3.45
Tiempo máximo observado	183.53

Distribución de tiempos de finalización y carrera atípica

En esta sección se analiza la variable cuantitativa continua, duración de las carreras, de Fórmula 1 entre 2013 y 2023; tomando el tiempo total expresado en minutos. Para explorar la distribución y detectar posibles valores atípicos, se utiliza un diagrama de caja (boxplot).



El gráfico de boxplot muestra la distribución de los tiempos de finalización de carrera, donde se identifican varios valores atípicos.

El valor mínimo, cercano a 3.5 minutos, se ubica muy por debajo del límite inferior de 70.42 minutos, según la regla de los cuartiles. Esto confirma estadísticamente su carácter atípico extremo inferior. Este valor corresponde al Gran Premio de Bélgica 2021, una carrera con

Tabla 3. Distribución de podios por piloto (puestos 1º, 2º y 3º)

Piloto	1º	2º	3º
VER	51	23	13
HAM	41	24	15
BOT	7	21	17

duración excepcionalmente corta. La razón detrás de esta anomalía fue la suspensión de la carrera debido a condiciones climáticas adversas: la lluvia intensa provocó que la competencia se disputara casi toda detrás del Safety Car y se otorgaran solo medio punto, sin reinicio efectivo de la carrera. La información oficial puede consultarse en el sitio de [Formula 1](#).

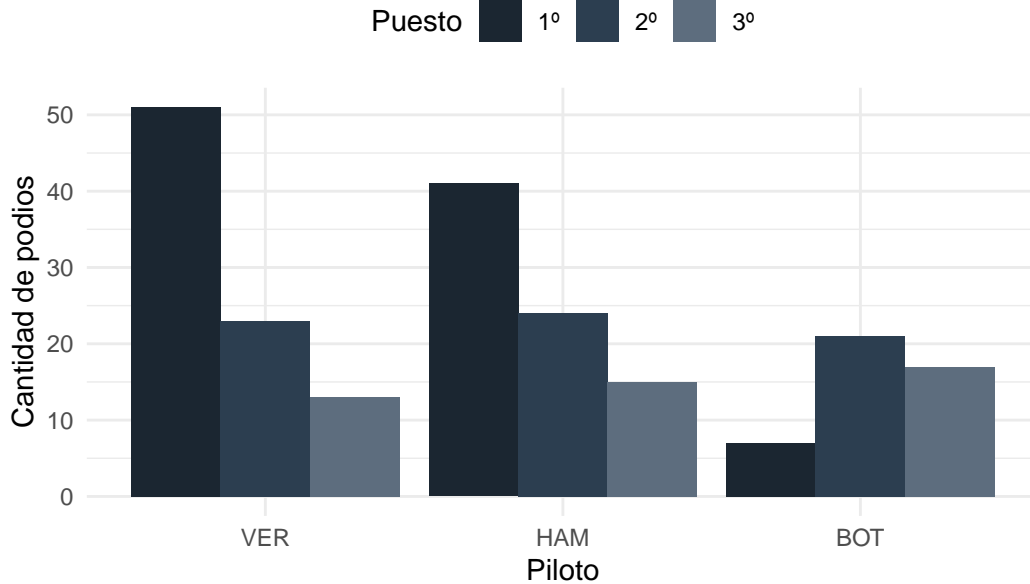
También se identifican valores atípicos superiores, es decir, tiempos que superan el límite superior de 117,50 minutos. Estas carreras, aunque no excepcionales como Spa 2021, probablemente se vieron afectadas por interrupciones, condiciones climáticas variables o accidentes.

Distribución de podios entre los tres corredores con más victorias

En esta sección se analiza la variable categórica “posición final”, acotada a los tres primeros puestos (1º, 2º y 3º), correspondientes a los corredores que integraron el podio en cada carrera. El análisis se centra exclusivamente en los tres pilotos con mayor cantidad de victorias: Max Verstappen (VER), Lewis Hamilton (HAM) y Valtteri Bottas (BOT).

Se busca describir la distribución de los puestos del podio entre estos corredores, con el objetivo de identificar patrones de desempeño y comparar el tipo de protagonismo que cada uno ha tenido a lo largo de las competencias.

Gráfico 4. Distribución de podios entre VER, HAM y BOT

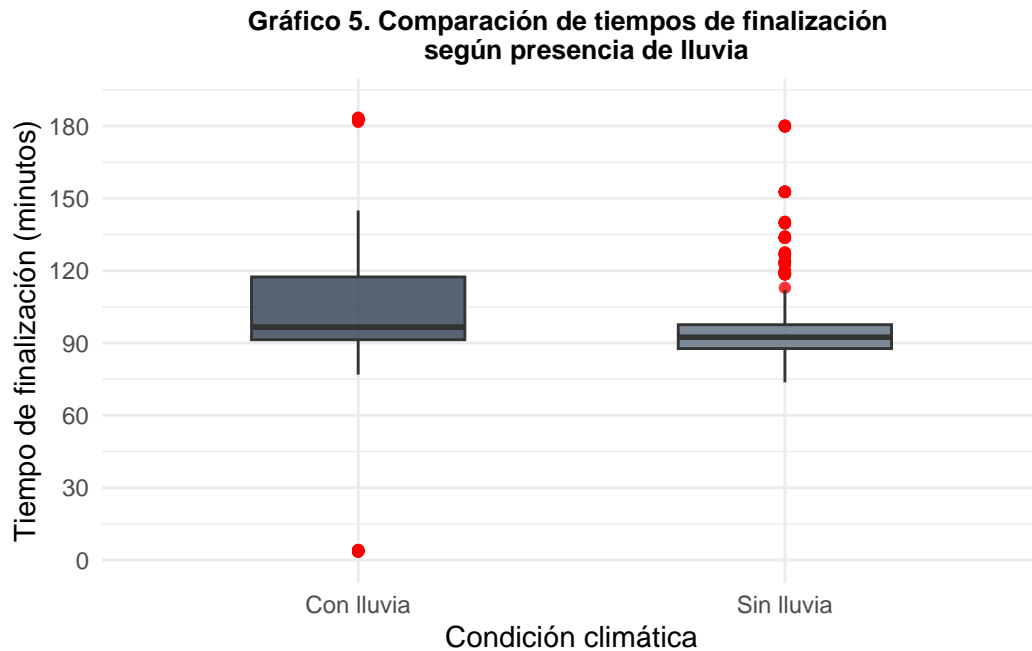


El análisis de los podios muestra cómo se distribuyen los primeros, segundos y terceros puestos entre los tres corredores con más victorias. Max Verstappen logró una gran cantidad de primeros puestos. Lewis Hamilton, en cambio, presenta un mayor equilibrio entre primeros y segundos lugares. Valtteri Bottas muestra una presencia destacada en el tercer puesto. Esta distribución permite visualizar el tipo de protagonismo que cada corredor ha tenido en la Fórmula 1 durante el período analizado.

Tiempos de finalización según presencia de lluvia

En esta sección se analiza la relación entre los tiempos de finalización de las carreras (expresados en minutos) y las condiciones climáticas presentes durante su desarrollo. Para ello, se combinan los datos de tiempos de llegada (provenientes de la tabla resultados) con la información climática registrada en la tabla carreras.

La variable de estudio es cuantitativa continua (tiempo_min) y se agrupa según una variable categórica, lluvia, con categorías “Con lluvia” y “Sin lluvia”). El objetivo es explorar si existen diferencias notables en los tiempos según las condiciones climáticas.



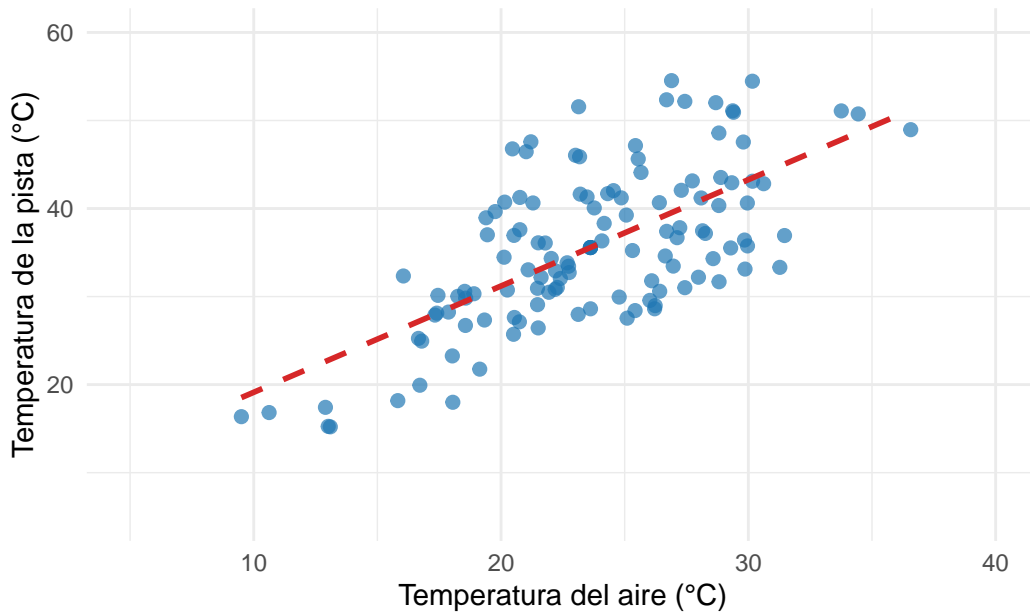
El gráfico compara la distribución de los tiempos de finalización entre carreras con y sin lluvia. Se observa que las carreras disputadas bajo condiciones de lluvia tienden a tener una mayor dispersión y, en algunos casos, mayores tiempos. Esto puede explicarse por la menor visibilidad, el menor agarre y la frecuencia de accidentes o neutralizaciones (como Safety Car), que son más comunes en condiciones climáticas adversas.

Relación entre temperatura del aire y temperatura de la pista

En esta sección se analiza la relación entre la temperatura atmosférica y la temperatura de la pista al momento de la carrera. Ambas variables son cuantitativas continuas y provienen de la tabla carreras.

Se espera encontrar una asociación positiva, dado que a mayor temperatura ambiente, generalmente se eleva también la temperatura del asfalto. Para evaluar esta relación, se emplea un gráfico de dispersión que permite observar la tendencia conjunta y posibles patrones lineales entre ambas variables.

Gráfico 6. Relación entre temperatura atmosférica y de pista



[1] 0.6675263

El gráfico de dispersión muestra una clara relación creciente entre la temperatura del aire y la temperatura de la pista. A medida que aumenta la temperatura atmosférica, también lo hace la del circuito.

Esta relación fue reforzada mediante una recta de ajuste lineal que evidencia una tendencia positiva. Además, se calculó la correlación de Pearson, que resultó en un valor de $r = 0.67$, lo cual indica una asociación moderada positiva entre ambas variables.

Los datos se concentran mayormente en el rango de 15 a 35°C en el aire y de 20 a 50°C en la pista. Esto confirma la influencia directa de las condiciones ambientales sobre el estado del circuito.

Sin embargo, aunque la relación es moderada y positiva, se observa una dispersión considerable de puntos alrededor de la recta de ajuste, lo que sugiere que podrían estar influyendo otras variables no consideradas en este análisis. Factores adicionales, como la humedad, la velocidad del viento, la radiación solar o incluso el calor generado por los propios autos en la pista, podrían afectar la temperatura y explicar la variabilidad observada en los datos.

Conclusión

En este trabajo se llevó a cabo un análisis exploratorio de datos (AED) sobre la Fórmula 1, abordando variables categóricas y numéricas vinculadas a los resultados de carrera, condiciones climáticas y desempeño de pilotos y equipos.

A través de gráficos, tablas y medidas estadísticas, se identificaron patrones clave como la alta proporción de finalizaciones completas, la duración típica de las carreras y la presencia de valores atípicos extremos, como el caso del **Gran Premio de Bélgica 2021**. Asimismo, se exploró la influencia de la **lluvia** en los tiempos de finalización, la distribución de **podios** entre los corredores más destacados, y la relación entre variables continuas como la **temperatura del aire** y la **temperatura de la pista**.

La aplicación de herramientas estadísticas y visuales permitió no solo describir los datos, sino también detectar comportamientos excepcionales y entender mejor el contexto competitivo y climático de la Fórmula 1. Estos hallazgos destacan la importancia del **análisis exploratorio de datos (AED)** como paso fundamental para interpretar grandes volúmenes de información y guiar futuras investigaciones o decisiones estratégicas.