



Data Analytics CoderHouse

Proyecto : FÓRMULA 1

Autor MUÑOZ FACUNDO

Profesor Leandro Araque, Tutor: Leywill Mañez

La Fórmula 1 es el deporte motor más importante del mundo, tanto por su rica historia, la innovación, la cantidad de aficionados alrededor del mundo, como por su impacto global.

El Campeonato de Fórmula 1 se desarrolla desde mitad del siglo pasado y tenemos en Argentina grandes exponentes como el recordado múltiple campeón de la categoría Juan Manuel Fangio, pasando por pilotos como Reuteman y Froilan Gonzales, otros grandes animadores del circo de la F1.

El Dataset fue seleccionado de <https://www.kaggle.com/> y tiene el título de Formula 1 World Championship (1950 - 2023), F1 race data from 1950 to 2023. Está conformado por 14 entidades o tablas y 120 atributos.

[Formula_1.zip](#)

Hipótesis.

Los adelantos tecnológicos en los autos de fórmula generan asimetrías en las diferentes escuderías según el presupuesto disponible para los desarrollos. Estas asimetrías disminuyen o cambian el paradigma cuando cambian los lineamientos técnicos generales de la categoría.

De la mano de la hipótesis anterior con el paso de los años estas asimetrías generan carreras más estáticas donde disminuyen los adelantamientos.

Objetivo

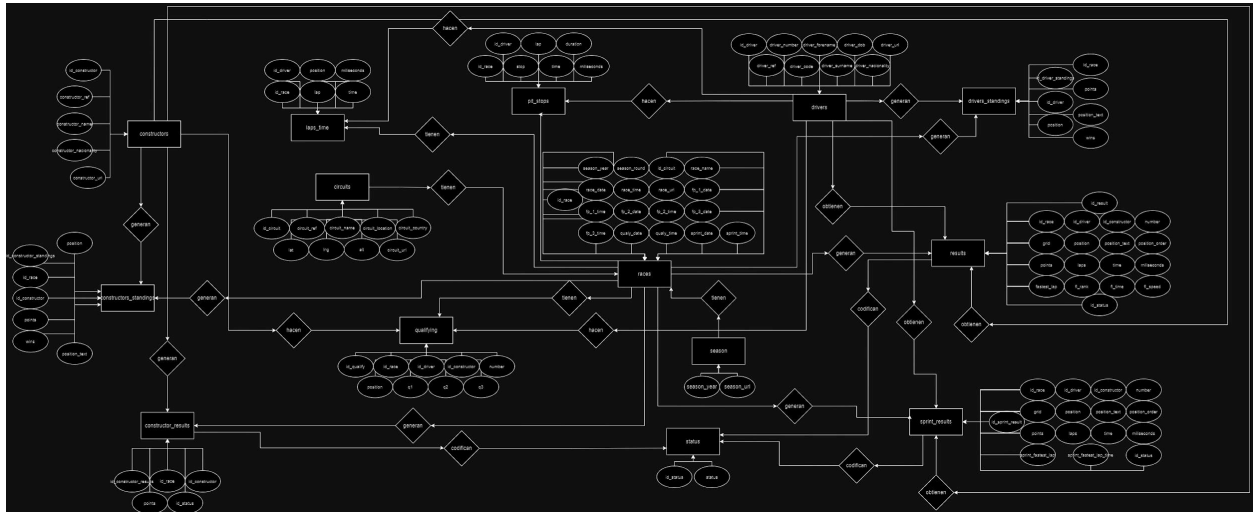
Se abordará la información desde un punto de vista descriptivo con la intención de buscar insights que permitan detectar los hitos históricos, según pilotos, escuderías y trazados. Y que nos permitan confrontar las hipótesis planteadas.

Usuario final



Este análisis, según nuestra hipótesis y objetivo está pensado para un público con mediana o alta capacidad de acción, está pensado para poder tomar decisiones tácticas y estrategias que permitan mejorar el show del campeonato mundial de Fórmula 1.

Diagrama Entidad Relación



F1 entidad relacion.jpg

Listado de Tablas de la Base de Datos

1. Races: Tabla de los eventos carrera con sus características
 - a. PK id_race
 - b. FK id_circuit
 - c. FK season_year
2. Drivers: Tabla de los pilotos
 - a. PK id_driver
3. Drivers_standings: Tabla de posiciones de los pilotos
 - a. PK id_driver_standing
 - b. FK id_race
 - c. FK id_driver
4. Constructors: Tabla de constructores
 - a. PK id_constructor
5. Constructors_standings Tabla de posiciones de los constructores
 - a. PK id_constructor_standing
 - b. FK id_race



- c. FK id_constructor
- 6. Constructors_results: Tabla de resultados de los constructores
 - a. PK id_constructor_result
 - b. FK id_race
 - c. FK id_constructor
 - d. FK id_status
- 7. Results: Tabla de resultados de las carreras
 - a. PK id_result
 - b. FK id_race
 - c. FK id_driver
 - d. FK id_constructor
 - e. FK id_status
- 8. Qualifying: Tabla de la clasificación a cada carrera
 - a. PK id_qualify
 - b. FK id_race
 - c. FK id_driver
 - d. FK id_constructor
- 9. Lap_time: Tabla de tiempos de vuelta de cada carrera
 - a. FK id_race
 - b. FK id_driver
- 10. Pit_stop: Tabla de Paradas en boxes en cada carrera
 - a. FK id_race
 - b. FK id_driver
- 11. Sprint_results: Tabla de resultados de las carreras sprint
 - a. PK id_sprint_result
 - b. FK id_race
 - c. FK id_driver
 - d. FK id_constructor
 - e. FK id_status
- 12. Circuits: Tabla de circuitos del campeonato
 - a. PK id_circuit
- 13. Seasons: Tabla de temporadas
 - a. PK season_year
- 14. Status: Tabla de estado particulares de cada resultado.
 - a. PK id_status



Listado de Columnas por Tabla

RACES		
PK	id_race	INT not null
FK	season_year	INT
	round	INT
FK	id_circuit	INT
	race_name	VARCHAR (50)
	race_date	DATE
	race_time	TIME
	race_url	VARCHAR (250)
	fp_1_date	DATE
	fp_1_time	TIME
	fp_2_date	DATE
	fp_2_time	TIME
	fp_3_date	DATE
	fp_3_time	TIME
	qualy_date	DATE
	qualy_time	TIME
	sprint_date	DATE
	sprint_time	TIME



RESULTS		
PK	id_result	INT not null
FK	id_race	INT
FK	id_driver	INT
FK	id_constructor	INT
	number	INT
	grid	INT
	position	INT
	position_text	VARCHAR (50)
	position_order	INT
	points	FLOAT
	laps	INT
	time	VARCHAR (50)
	milliseconds	INT
	fastest_lap	TIME
	fl_rank	INT
	fl_time	TIME
	fl_speed	FLOAT
FK	id_status	INT

SPRINT_RESULTS		
PK	id_sprint_result	INT not null
FK	id_race	INT



FK	id_driver	INT
FK	id_constructor	INT
	number	INT
	grid	INT
	position	INT
	position_text	VARCHAR (50)
	position_order	INT
	points	FLOAT
	laps	INT
	time	VARCHAR (50)
	milliseconds	INT
	sprint_fastest_lap	INT
	sprint_fastest_lap_time	TIME
FK	id_status	INT

DRIVERS		
PK	id_driver	INT not null
	driver_ref	VARCHAR (50)
	driver_number	INT
	driver_code	VARCHAR (50)
	driver_forename	VARCHAR (50)
	driver_surname	VARCHAR (50)



	driver_dob	DATE
	driver_nacionalidad	VARCHAR (50)
	driver_url	VARCHAR (250)

DRIVER_STANDINGS		
PK	id_driver_standing	INT not null
FK	id_race	INT
FK	id_driver	INT
	points	FLOAT
	position	INT
	position_text	VARCHAR (50)
	wins	INT

CONSTRUCTORS		
PK	id_constructor	INT not null
	constructor_ref	VARCHAR (50)
	constructor_name	VARCHAR (50)
	constructor_nacionalidad	VARCHAR (50)
	constructor_url	VARCHAR (250)

CONSTRUCTOR_STANDINGS		
PK	id_constructor_standing	INT not null
FK	id_race	INT



FK	id_constructor	INT
	points	FLOAT
	position	INT
	position_text	VARCHAR (50)
	wins	INT

CONSTRUCTORS_RESULTS		
PK	id_constructor_results	INT not null
FK	id_race	INT
FK	id_constructor	INT
	points	FLOAT
FK	id_status	INT

CIRCUITS		
PK	id_circuits	INT not null
	circuit_ref	VARCHAR (50)
	circuit_name	VARCHAR (50)
	circuit_location	VARCHAR (50)
	circuit_country	VARCHAR (50)
	lat	FLOAT
	lng	FLOAT
	alt	INT
	circuit_url	VARCHAR (250)



LAPS_TIME		
FK	id_race	INT
FK	id_driver	INT
	lap	INT
	position	INT
	time	TIME
	milliseconds	INT

PIT_STOP		
FK	id_race	INT
FK	id_driver	INT TM
	stop	INT
	lap	INT
	time	TIME
	duration	TIME
	milliseconds	INT

QUALIFYING		
PK	id_qualify	INT not null
FK	id_race	INT



FK	id_driver	INT
FK	id_constructor	INT
	number	INT
	position	INT
	q1	TIME
	q2	TIME
	q3	TIME

STATUS		
PK	id_status	INT not null
	status	VARCHAR (50)

SEASON		
PK	season_year	INT not null
	season_url	VARCHAR (250)

Paleta de colores



#401920
RGB 64, 25, 32

#1B1C26
RGB 27, 28, 38

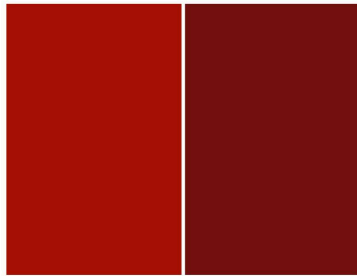
#F21905
RGB 242, 25, 5



#F2F1F0
RGB 242, 241, 240

#F21905
RGB 242, 25, 5

#BF1304
RGB 191, 19, 4



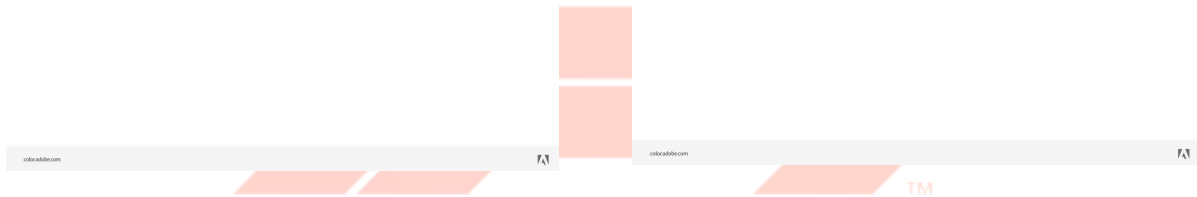
#A61103
RGB 166, 17, 3

#731010
RGB 115, 16, 16



#F25D50
RGB 242, 93, 80

#F28BB6
RGB 242, 185, 182





Modificaciones

Creación de Tablas

*Tabla Calendario

```
calendar (MIN(races[race_date]),max(races[race_date]))  
mes = MONTH(TablaCalendario[Date])  
anio = YEAR(TablaCalendario[Date])
```

* Tabla Ganadores

Duplique la tabla resultados y elimine todos los registros con la posición distinta a 1

* Tabla Anexar1

Anexe las tablas Results y Sprint_Results para poder sumar la cantidad de puntos por cada temporada

* **Tabla Países** a partir del Archivo plano World_Population de Kaggle para poder usar la segmentacion por continentes

* Tabla TopDriverTable1

```
TopDriverTable1 = TOPN(1, SUMMARIZE( Anexar1, Anexar1[id_driver], "TotalPoints",  
SUM(Anexar1[points]) ), [TotalPoints], DESC )
```

* Tabla Medidas

```
*CantidadCarreras = DISTINCTCOUNT(races[id_race])
```

```
*CantidadCarreras_sprint = DISTINCTCOUNT(sprint_results[id_race])
```

```
*CantidadCarrerasDriver = CALCULATE (COUNT (races [id_race]), DISTINCT  
(results[id_driver]))
```

```
*CantidadCircuitos = DISTINCTCOUNT(circuits[id_circuit])
```

```
*CantidadConstructores = DISTINCTCOUNT(results[id_constructor])
```



```
*CantidadPilotos = DISTINCTCOUNT(drivers[id_driver])

*Constructor_Rank = SUM(constructor_standings[position])

*Driver_Rank = SUM(driver_standings[position])

*TasaCompetencia = IF(COUNT(results[id_result])=0,0, 1-( [NoCompetencia]
/COUNT(results[id_result])))

* TasaVictoria = CALCULATE(
COUNT('#tabla_ganadores_carreras'[position_order])/ + COUNT(results[id_race]),
results[id_driver])

*PolePositionCount = CALCULATE(COUNT(qualifying[id_race]), DISTINCT
(qualifying [id_driver]),qualifying[position]=1)

*VueltaRapidaQualy = MINX('qualifying', IF('qualifying'[position]=1, (IF (ISBLANK
('qualifying'[q3]), IF(ISBLANK('qualifying'[q2]), 'qualifying'[q1], 'qualifying'[q2]),
'qualifying'[q3])))

*PolePosition_id_driver = SUMX(qualifying, IF('qualifying'[position]=1, 'qualifying'
[id_driver]))

*TasaCompetencia F1 = 0.29

*Competencia =
CALCULATE(DISTINCTCOUNT(results[id_result]),results[id_status] = 1)

*NoCompetencia =
CALCULATE(DISTINCTCOUNT(results[id_result]),results[id_status] <> 1)

*Campeon =

VAR TopDriverTable = TOPN( 1, SUMMARIZE( Anexar1, Anexar1[id_driver],
    "TotalPoints", SUM(Anexar1[points]) ), [TotalPoints],DESC )
RETURN MAXX(TopDriverTable, Anexar1[id_driver])

*CampeonEquipo =

VAR TopEquipoTable = TOPN( 1, SUMMARIZE( Anexar1, Anexar1 [id_constructor],
    "TotalPoints", SUM(Anexar1[points]) ), [TotalPoints], DESC )
RETURN MAXX(TopEquipoTable, Anexar1[id_constructor])
```

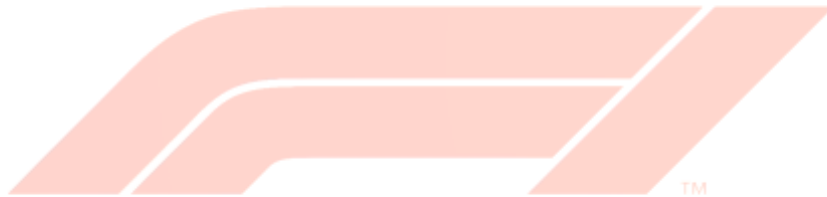


*NombrePilotoCampeon = LOOKUPVALUE(drivers[piloto], drivers[id_driver],
[Campeon])

*NombreEquipoCampeon = LOOKUPVALUE(constructors[constructor_name],
constructors [id_constructor], [CampeonEquipo])

Creacion de Columna en Tabla Resultados

Agrupación Position (grupos)



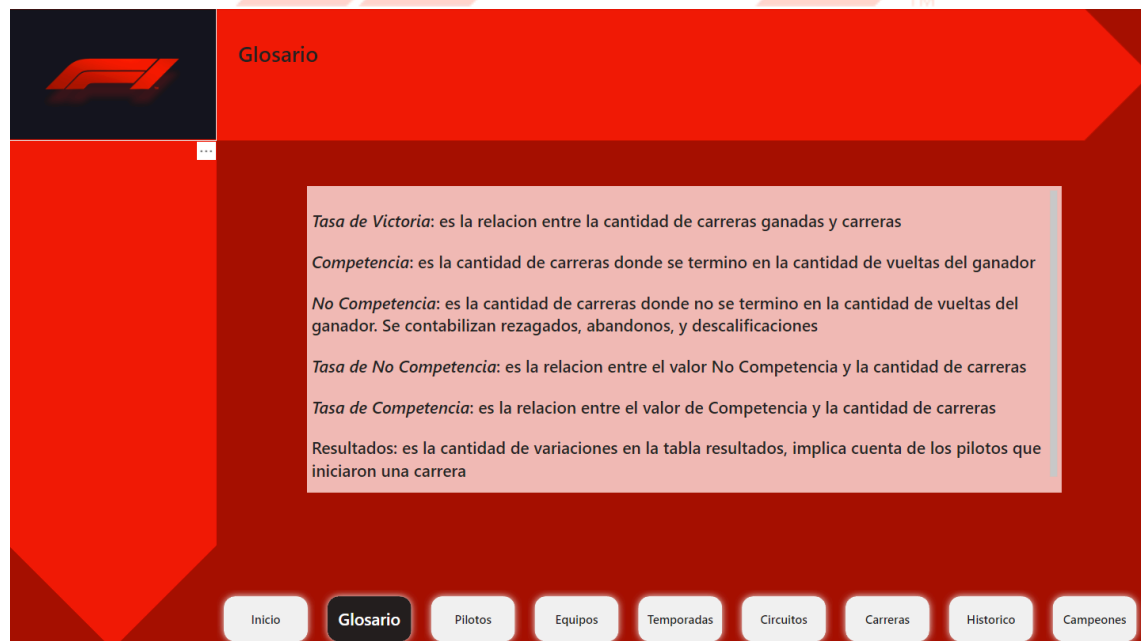


Tablero



Página de Inicio del Tablero , Botones de navegación por página y paleta de colores de la temática.

Glosario



Página con conceptos claves que se utilizaron para los análisis



Pilotos



Análisis histórico de pilotos que participaron del campeonato mundial de F1 desde su creación en 1950.

Filtros por Nacionalidad y por Piloto

Gráfico 1 Puntos por Piloto, top 10 pilotos con mayor cantidad de puntos. Segmentado por posiciones

Gráfico 2 Victorias y Carreras por Piloto, top 10 pilotos con mayor cantidad de carreras y las victorias de cada uno

Gráfico 3 Tasa de Victoria y Tasa de Competencia, Top 10 Pilotos con mejor tasa de Victorias (filtro pilotos con más de 10 victorias) y su tasa de competencia.

Gráfico 4 Pole Position (1994-act) por piloto, Top 15 de pilotos con mayor cantidad de Pole Positions.

Tarjeta 1 Cantidad de Pilotos

Tarjeta 2 Cantidad de Carreras



Tarjeta 3 Cantidad de Pilotos Ganadores

Tarjeta 4 Tasa de Victoria

Equipos



Análisis histórico de Equipos que participaron del campeonato mundial de F1 desde su creación en 1950.

Filtros por Equipo y por Temporadas

Gráfico 1 Puntos por Equipo

Gráfico 2 Puntos por Equipo segmentado por posiciones, Top 10 Equipos con mayor cantidad de puntos

Gráfico 3 Carreras y Victorias por Equipo, Top 10 Equipos con mayor cantidad de carreras

Gráfico 4 Tasa de Competencia y Carreras por Equipo (filtro Equipos con más de 100 Carreras). Solo 29 equipos de 210 en toda la historia superaron las 100 carreras

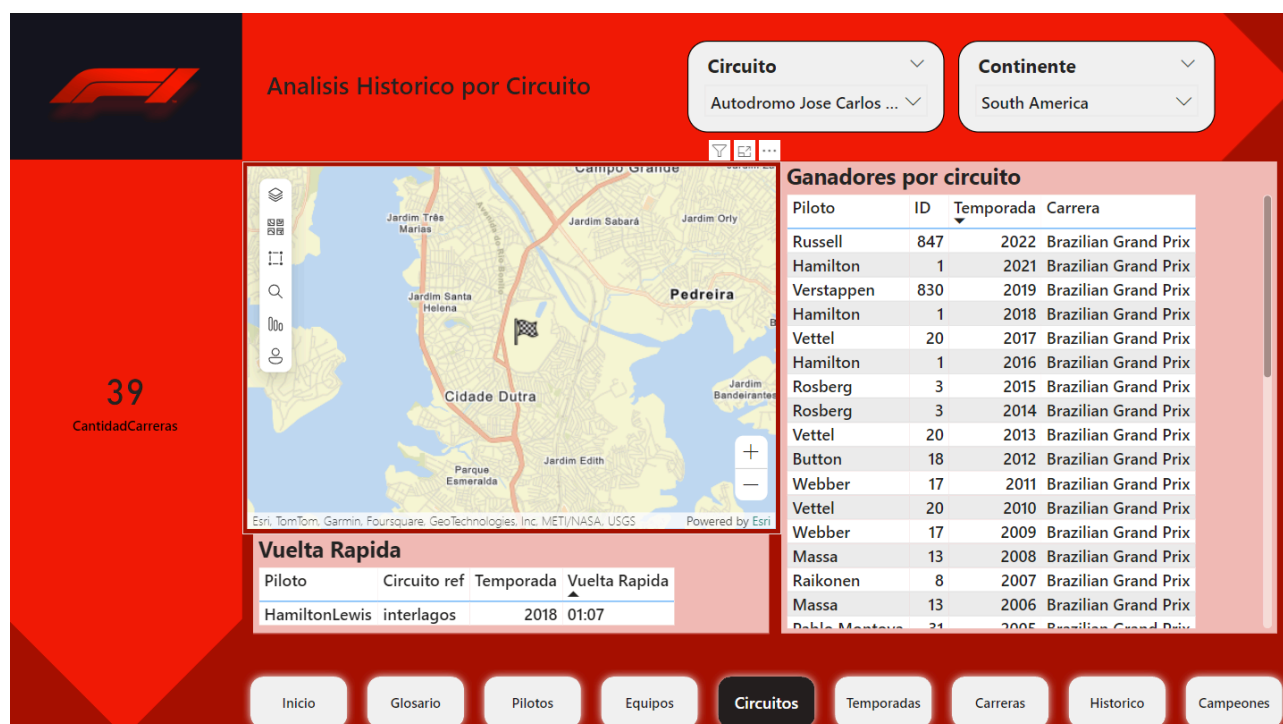
Tarjeta 1 Cantidad de Equipos



Tarjeta 2 Cantidad de Carreras

Tarjeta 3 Cantidad de Equipos Ganadores

Circuitos



Análisis histórico de los Circuitos que participaron del campeonato mundial de F1 desde su creación en 1950.

Filtros por Circuito y Continente

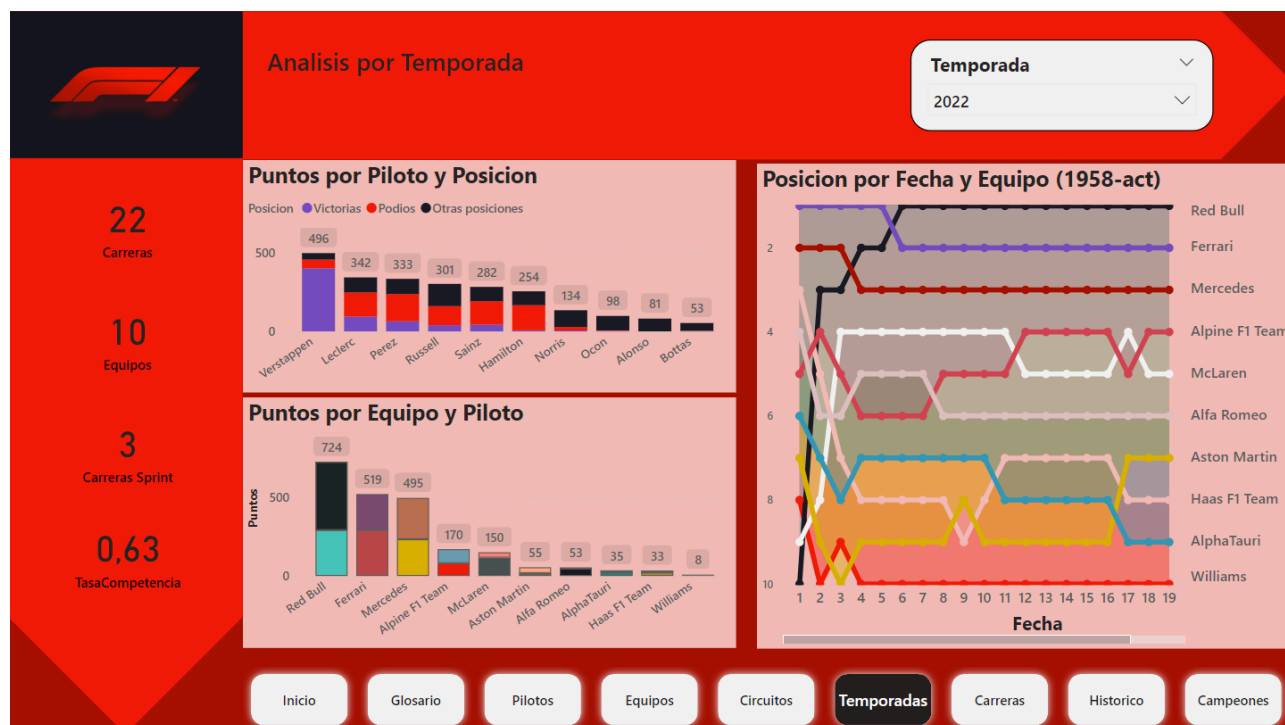
Gráfico 1 Ubicación del Circuito

Gráfico 2 Pilotos Ganadores del Circuito

Gráfico 3 Vuelta mas rapida de Clasificación del Circuito

Tarjeta 1 Carreras, Cantidad de Carreras según el filtro, Tooltip Porcentaje de Carreras por Continente

Temporadas



Análisis de las temporadas del Campeonato Mundial de F1

Filtros por Temporada (Selección única)

Gráfico 1 Puntos por Piloto y posición, Top 10 Pilotos por puntos segmentado por posiciones

Gráfico 2 Posición por Fecha del campeonato y Equipo (1958-act)

Gráfico 3 Puntos por Equipo y Piloto, Total de puntos por Equipo en la temporada segmentado por sus pilotos.

Tarjeta 1 Cantidad de Carreras

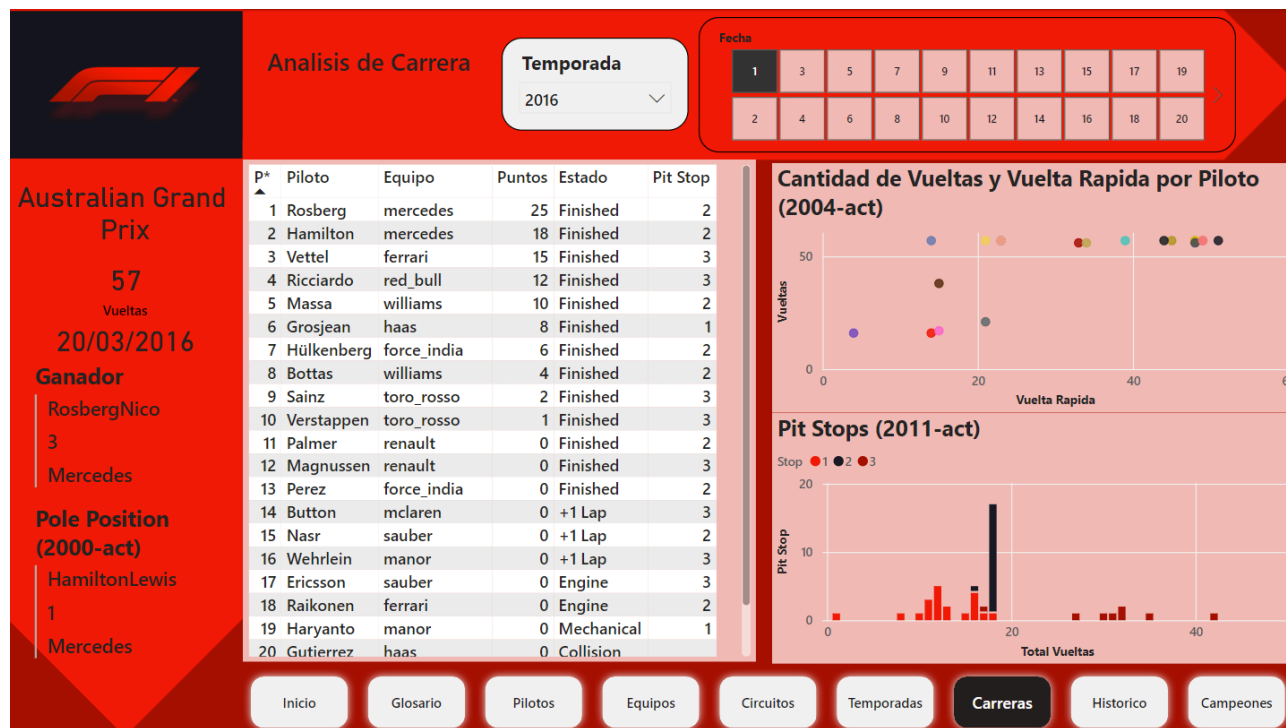
Tarjeta 2 Cantidad de Equipos

Tarjeta 3 Cantidad de Carreras Sprint

Tarjeta 4 Tasa de Competencia Temporada



Carreras



Análisis de las Carreras del Campeonato Mundial de Fórmula 1™

Filtro por Temporada y por Fecha

Tabla 1 Posiciones por Piloto, Equipo, Puntos, Estado y Pit Stop

Gráfico 1 Cantidad de Vueltas y Vuelta Rápida por Piloto (2004-act) gráfico de dispersión que relaciona esos 2 factores por piloto, Cuanto más alto en el eje y más vueltas completo de la carrera y el eje X muestra en qué momento de la carrera realizó su vuelta rápida.

Gráfico 2 Pit Stop (2011-act) Cantidad de Pit Stops por Vuelta Segmentado por el N de Pit Stop de cada Piloto

Tarjeta 1 Nombre de la Carrera

Tarjeta 2 Cantidad de Vueltas

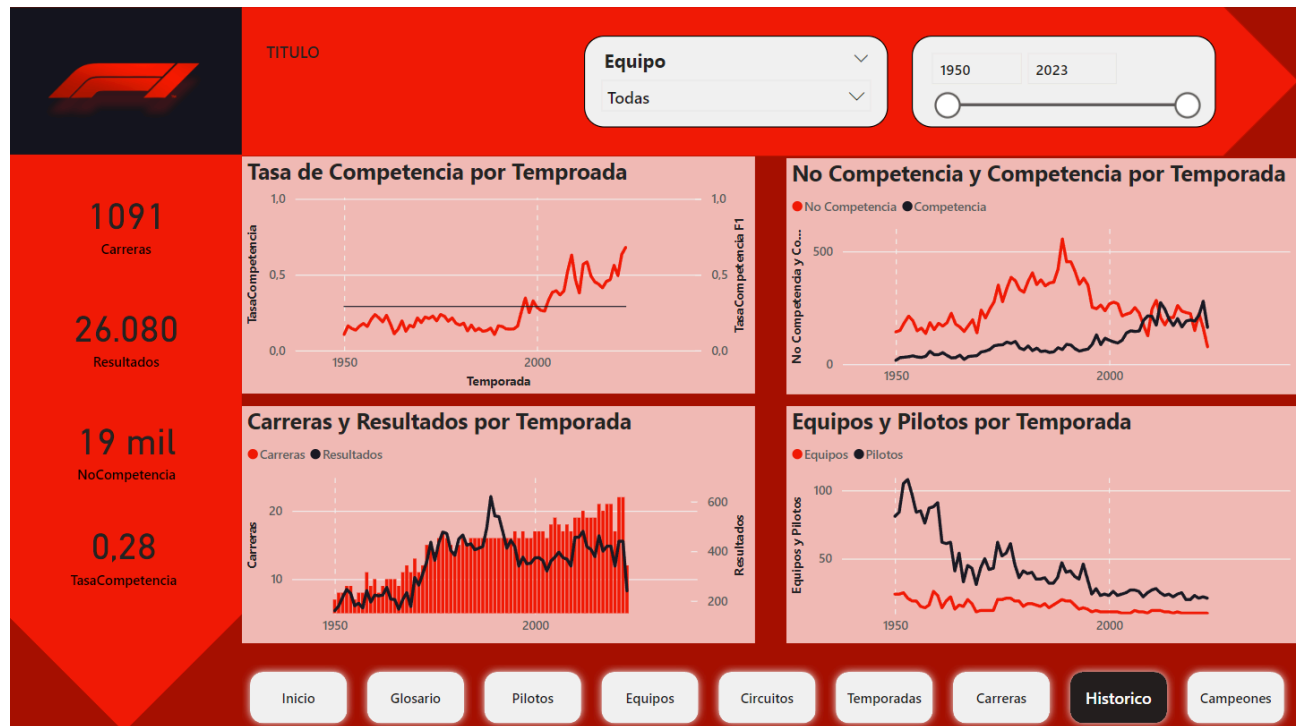
Tarjeta 3 Fecha de la Carrera

Tarjeta 4 Piloto y Equipo Ganador

Tarjeta 5 Piloto y Equipo Pole Position



Histórico



Análisis del Campeonato Mundial de F1

Filtro por Equipo y por Temporada

Gráfico 1 Tasa de Competencia por Temporada

Gráfico 2 No Competencia y Competencia por Temporada

Gráfico 3 Cantidad de Carreras y Resultados por Temporada

Gráfico 4 Cantidad de Equipos y Pilotos por Temporada

Tarjeta 1 Cantidad de carreras

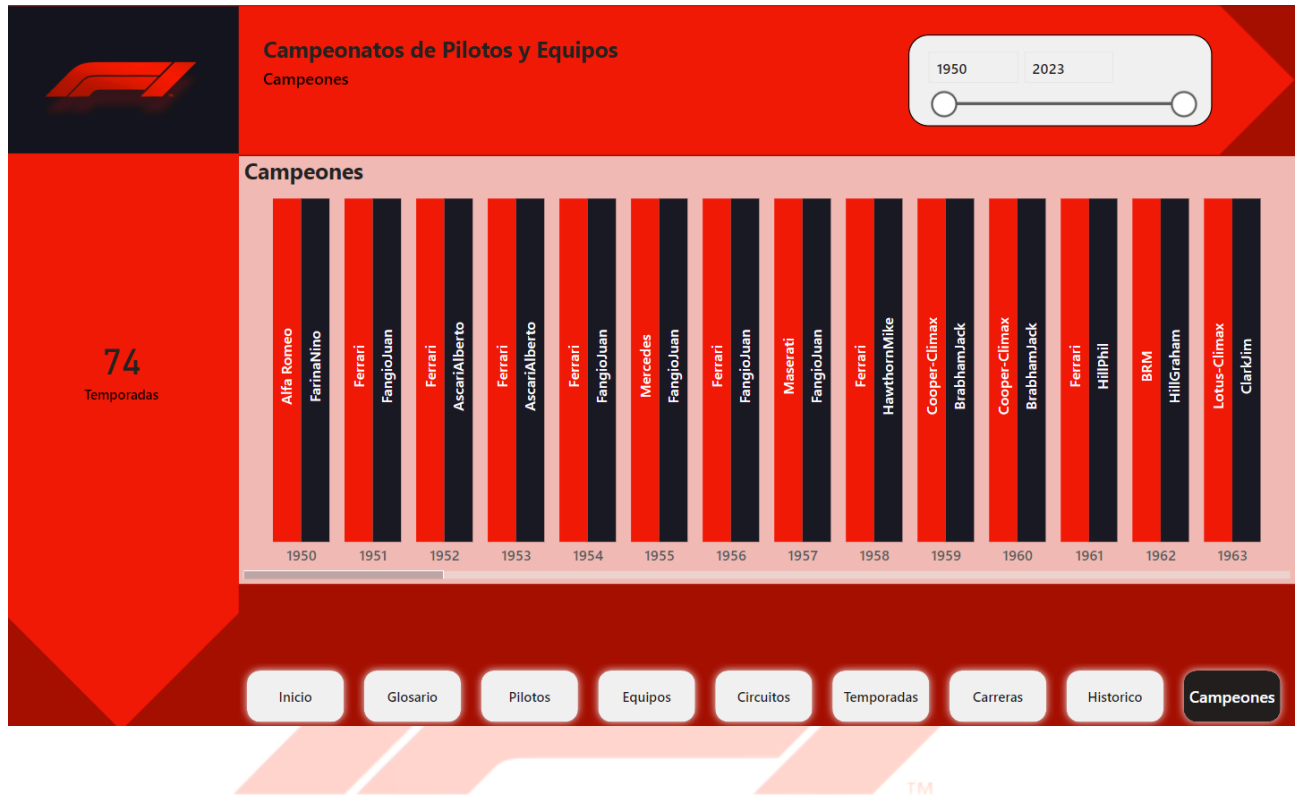
Tarjeta 2 Cantidad de Resultados

Tarjeta 3 Cantidad de No Competencia

Tarjeta 4 Tasa de Competencia



Campeones



Campeones del Campeonato Mundial de F1

Filtro por Temporada

Gráfico 1 Pilotos y Equipos Campeones del Campeonato Mundial de Fórmula 1 por Temporada

Tarjeta 1 Cantidad de Temporadas



Discusión

La complejidad del Dataset condiciono los resultados de este análisis, podemos destacar varios hitos que se desprenden del análisis de los datos

1- a partir de la temporada de 1995 cambia abruptamente la relación entre cantidad de resultados positivos y la totalidad, y empieza a correr al ritmo de la media histórica y por encima.

2- recién en 2008, 58 temporadas después de su creación, la cantidad de pilotos con resultados positivos es mayor que los negativos, y si bien ha ido fluctuando luego de este hito la tendencia, luego de las últimos tres campeonatos es que se mantenga en ascenso.

3- Cantidad de Carreras en aumento, hemos pasado de una media de 10 carreras por temporada durante los primeros 20 años del campeonato a estar en la actualidad consistentemente por encima de las 20 , con un máximo de 22 durante el campeonato del 2022. Sin embargo la cantidad de Resultados ha disminuido desde su pico máximo en 1989, donde la combinación de cantidad de pilotos y carreras ha saturado la participación en cada una.

En la Temporada 1989 hubo 20 equipos, 47 pilotos en 16 carreras lo que produjo un máximo de 620 autos saliendo a pista durante ese año. De esos 620 solo el 10% logró completar la carrera en la cantidad de vueltas del ganador, casi 3 veces menos que el promedio histórico del campeonato.

4- la cantidad de equipos viene en disminución desde un máximo de 26 equipos en 1959 hasta los 10 actuales, que hoy están acotados por los reglamentos y por la repartición de los derechos de imagen y transmisión del campeonato.

5- la cantidad de Pilotos viene disminuyendo, desde un máximo de 108 pilotos en 1953 a los actuales 20 por temporada, como en el punto anterior hoy por reglamento hay 2 pilotos por equipo titulares,



Estos últimos 2 puntos de análisis hacen inferir que con el tiempo hay una mayor profesionalización tanto de los equipos y pilotos, es de destacar que los pilotos con mayor cantidad de carreras son los que han competido en los últimos 20 años.

También inferimos que al haber una menor cantidad de equipos los presupuestos de desarrollo de los autos ha ido en aumento, hoy por hoy hay límites presupuestarios en los equipos para poder disminuir la brecha de rendimiento de cada uno. Aun así es claro que esta sumatoria de factores (equipos con más presupuesto, pilotos más profesionales) ha aumentado la eficiencia de los autos, bajando la cantidad de resultados negativos.

Este conjunto de factores hace que hoy el campeonato Mundial de Fórmula 1 sea muy competitivo, y los resultados están muy ligados a los cambios tecnológicos que puede presentar cada equipo en un determinado periodo de tiempo.

Por esto cada temporada el campeonato desarrolla un reglamento técnico para poder limitar esos avances y permitir una mayor paridad. Asimismo cada tantas temporadas se modifica sustancialmente el paradigma de los autos (la próxima gran modificación se va a dar en la temporada 2026, para poder agitar un poco más los esquemas establecidos en pos de la paridad y el show.