

Proyecto:

Visualización histórica de la Formula 1



Alejandro Senabre

Introducción:

Proyecto personal combinando dos de mis pasiones, el análisis y visualización de datos junto con la Formula 1, deporte automovilístico que disfruto desde 2010 cuando Vettel se alzaba con su primer campeonato mundial de F1. Lo que intenté reflejar en este proyecto fue un resumen de la temporada 2023, mostrando cada escudería que la compuso y la competencia interna de sus pilotos, además de poder seleccionar a gusto el año en particular para ver datos históricos de como vario la pelea por el mundial de constructores y datos interesantes, como la cantidad de escuderías o los puntos repartidos de cada año.

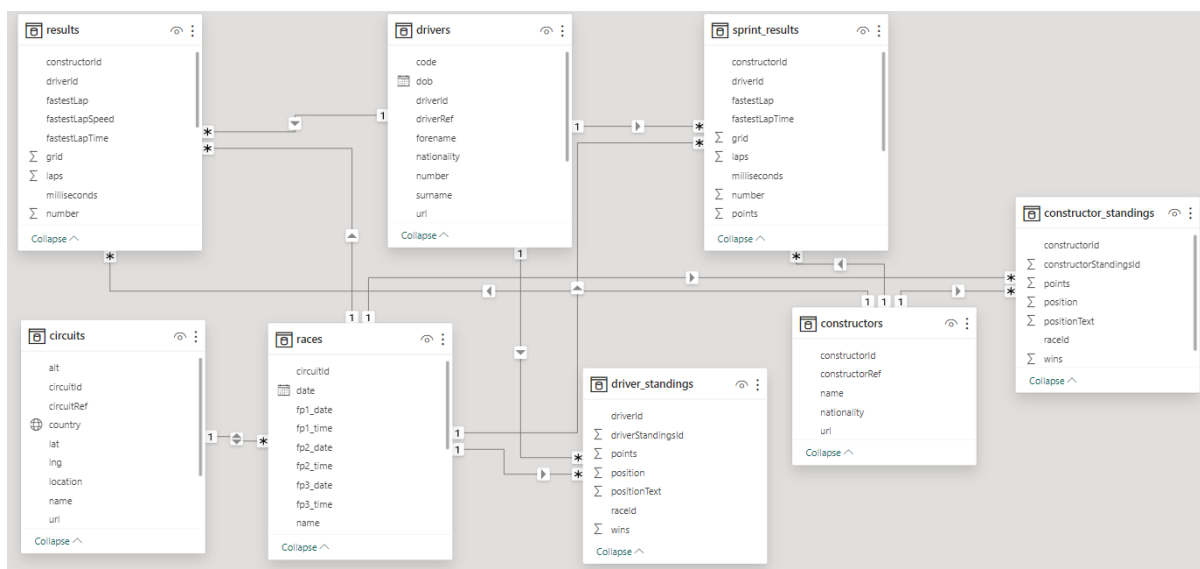
Obtención de los datos:

La datos se obtuvieron de [Kaggle](https://www.kaggle.com/datasets/ergast/ergast) que a su vez lo obtiene de la API de www.ergast.com/mrd donde recopila la información de cada carrera de F1, teniendo datos que van desde 1950 a la actual temporada 2024.

Los datos están compuestos por 14 tablas en archivos CSV:

- circuits.csv
- constructor_results.csv
- constructor_standings.csv
- constructors.csv
- driver_standings.csv
- drivers.csv
- lap_times.csv
- pit_stops.csv
- qualifying.csv
- races.csv
- results.csv
- seasons.csv
- sprint_results.csv
- status.csv

Para este proyecto se utilizaron las siguientes ocho tablas que se insertaron a un modelo relacional en Power BI, cada una relacionada mediante claves como “circuitid” y “driverid” para poder realizar consultas y visualizaciones cruzadas.



Explicación paginas

Inicio

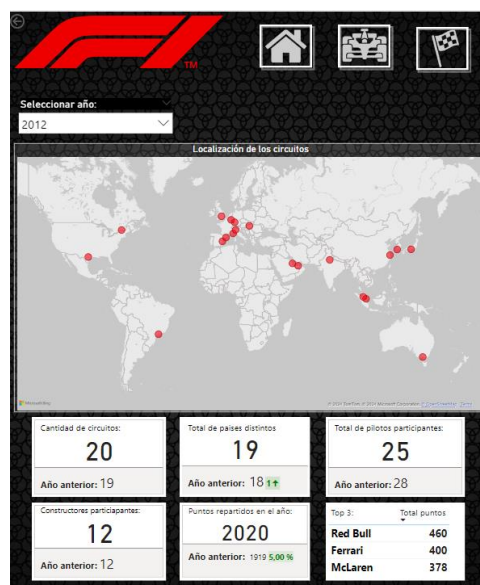


Lo primero que se aprecia es la personalización del tamaño de la pagina, se decidió hacer con una altura de 1200px y 1000px de ancho, con alineamiento vertical. El motivo de esta decisión es para tener un tamaño de pagina moderno que sea fácilmente exportable a una visualización móvil.

Todos los fondos fueron realizados mediante PowerPoint, inclusive los botones donde posteriormente se le agregó una botonera invisible en Power BI para poder interactuar con la misma.

La primera pagina solamente esta compuesto con botones que nos redirecciona a distintas pestañas con datos.

Pagina resumen



Esta segunda contiene mayor cantidad de detalles, lo primero son los accesos directos de la parte posterior derecha para ir a diferentes pestañas y volver al inicio. Lo siguiente

es el mapa, que para su creación se utilizó las coordenadas de latitud y longitud proporcionada por la tabla “circuits” para que nos ubique en el lugar exacto de la pista.

Lo más destacable son las 6 tarjetas de visualización de la parte inferior. Para estas se utilizaron las nuevas tarjetas que ofrece Power BI llamadas “card (new)” que nos da la posibilidad de crear características dinámicas en una misma tarjeta, sin tener la necesidad de crear 2 superpuestas. Las tarjetas donde se pueden apreciar mejor el funcionamiento de las nuevas visualizaciones son las siguientes:



Estas dos además de tener el visualizador clásico en la parte superior, podemos agregar un dato de referencia, además de incluir un detalle a este último, por lo que se aprovecha esta función para mostrar su relación con la temporada anterior, mostrando si hubo un aumento o disminución de los puntos o de los países donde visito la F1.

Para evitar el filtro de la página afecte a los datos del año anterior, se tuvo que agregar el valor mediante una función DAX

```
Total_Puntos_Anterior =  
CALCULATE(  
    SUM(results[points])+SUM(sprint_results[points]) ,  
    SAMEPERIODLASTYEAR(races[date]),  
    ALLEXCEPT(races, races[date])  
)
```

Se utilizó las siguientes funciones:

- **CALCULATE:** Evalúa una expresión en un contexto de filtro modificado. En este caso, la expresión es la suma de los puntos de las carreras normales y de las carreras Sprint (carreras similares a las normales pero con una calificación aparte y número de vueltas reducido, donde también se reparten puntos), luego se aplican los filtros con las otras dos funciones.
- **SAMEPERIODLASTYEAR:** Devuelve una tabla con las fechas del mismo periodo pero del año anterior. En este caso, se utilizó la columna del filtro.
- **ALLEXCEPT:** Elimina todos los filtros de contexto de una tabla, excepto los que se aplican a las columnas que se especifican.

Para la segunda visualización se usó algo similar pero con la suma de los circuitos.

```

Total_paises_año_anterior =
CALCULATE(
    DISTINCTCOUNT(circuits[country]),
    SAMEPERIODLASTYEAR(races[date]),
    ALLEXCEPT(races, races[date])
)

```

Este segundo caso, lo que tiene de interesante es la agregación del detalle, donde incluimos una fecha apuntando a la dirección dependiendo del valor del mismo, si es positivo da como resultado una flecha apuntado para arriba, en cambio para el negativo, apunta para abajo, en caso que sea igual, se agrega el signo del igual. El código para realizar esto fue el siguiente:

```

Dif_paises_años =
VAR __DIFERENCIA = DISTINCTCOUNT(circuits[country]) - [Total_paises_año_anterior]
VAR __UP_ARROW= UNICHAR(129033)
VAR __DO_ARROW= UNICHAR(129035)
VAR __IGUAL= UNICHAR(61)
RETURN
IF( __DIFERENCIA <0, __DIFERENCIA & __DO_ARROW,
IF ( __DIFERENCIA = 0 , __DIFERENCIA & __IGUAL , __DIFERENCIA & __UP_ARROW))

```

Se incluyeron las siguientes funciones:

- VAR: Define variables locales dentro de una formula. Variables para ser usado en otra parte de la formula.
- UNICHAR: Devuelve el carácter Unicode que corresponde a un valor numérico.
- RETURN: Indica el valor que devuelve la formula. Se usa el condicional IF para indicar si la diferencia es positiva, negativa o cero.

Por ultimo a lo que respecta al color, se agrego un formato condicional al valor del detalle

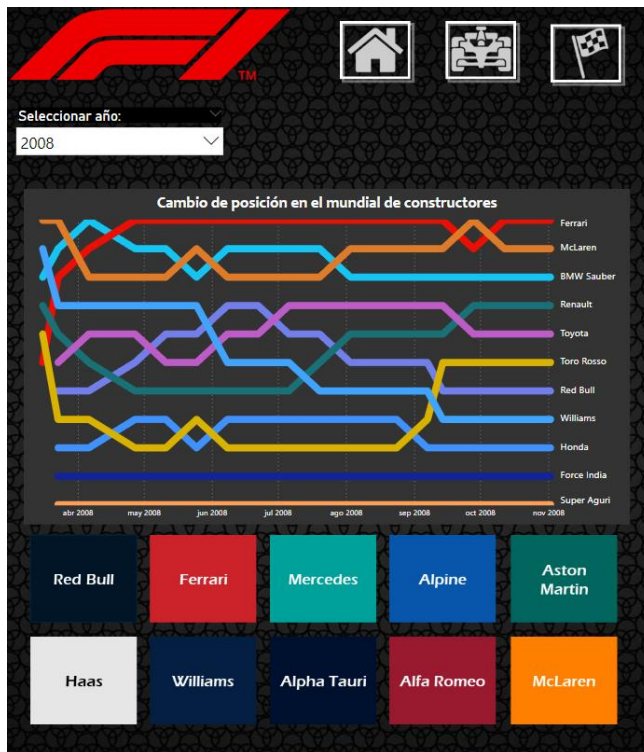
```

Color Country =
VAR __PEPE= DISTINCTCOUNT(circuits[country]) - [Total_paises_año_anterior]
RETURN
IF ( __PEPE <0, "Red",
IF( __PEPE =0 , "Blue", "Green"))

```

Este condicional, nos devuelve como valor el color que queremos utilizar a la hora de visualizar, dependiendo de la diferencia y su valor.

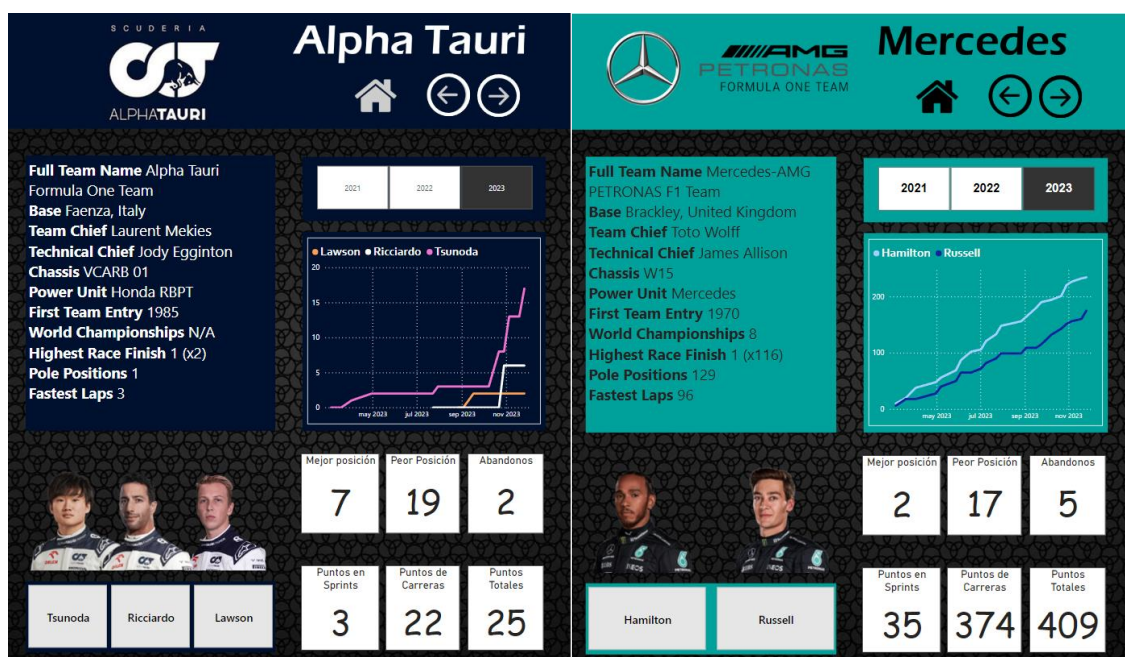
Página de escuderías



Simplemente nos encontramos con un grafico de líneas donde en la X se encuentra la fecha de los grandes premios y en el eje Y, la cantidad de puntos hasta el momento del mundial de pilotos, mostrando de inicio a fin de la temporada como variaron sus respectivas posiciones.

En la parte inferior se haya un acceso directo a cada una de las escuderías que compitieron en la temporada 2023.

Pagina de equipos



Tenemos 10 paginas distintas para cada escudería participante. En esta podemos encontrar información sobre la marca, los pilotos con los que compitió durante la temporada, datos numéricos sobre los objetivos alcanzados durante el año, y una grafica que muestra la evolución de los puntos a lo largo de las carreras para ambos compañeros de equipo.

Se aprecia dos tipos de filtro, uno para los pilotos, y otro para los años. El de los pilotos modifica los datos numéricos dependiendo de que piloto los obtuvo, mientras que el filtro de año, nos muestra como fue el rendimiento en años anteriores.

Futuras mejoras

Gráfico de red

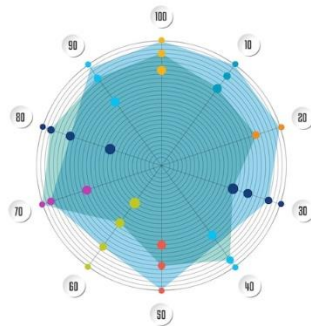
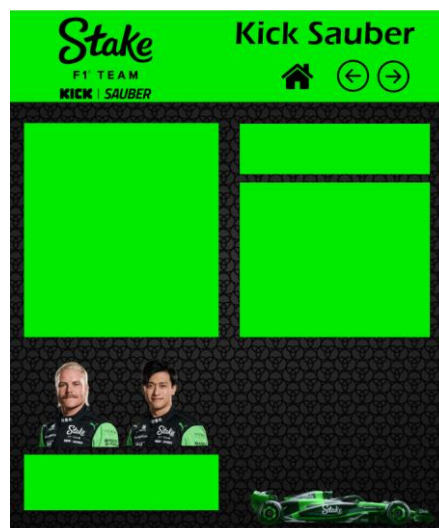


Gráfico de red que nos muestre el valor para cada escuderia o piloto, que este compuesto por la fiabilidad (veces que tuvieron que abandonar por avería), potencia (velocidad máxima alcanzada en las carreras a comparación de sus rivales), experiencia de los pilotos (temporadas que participaron y sus resultados), entre otros.

Temporada 2024



Crear una sección para que incluya a los equipos actualizados a la temporada 2024, aunque sea similar a la anterior temporada, se incluye dos equipos actualizados, Kick Sauber y RB Visa.

Otras mejoras

- **Actualización automática con la API:** Se planea conectar la API mediante Python para extraer los datos actualizados luego de cada carrera, subirlo a una base de datos y que Power BI lo actualice automáticamente.
- **Agregar mas datos:** Explotar al máximo la base de datos para incluir datos tales como, velocidad máxima alcanzada, cantidad de vueltas rápidas, etc.
- **Crear un dashboard:** Un Dashboard que resuma toda la información de una temporada, que incluya la tabla de pilotos y de constructores.
- **Datos sobre pistas:** Incluir una sección con la información del circuito, ubicación, records de vuelta, apariciones, etc.
- **Data Science:** Poder predecir el ritmo de carrera y su posterior ubicación con los datos de FP1, FP2, FP3 y clasificación.