

Fórmula 1

• Descripción

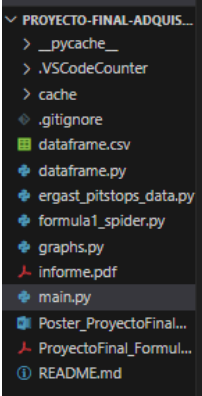
El proyecto consiste en adoptar el rol de analista de datos para un equipo de Fórmula 1, con el objetivo de compilar y analizar datos relevantes para evaluar la influencia de los pit-stops en los resultados del campeonato mundial. Mediante el uso de técnicas de web scraping y APIs, se recopila información detallada de cada carrera y los pit-stops de las temporadas 2012 a 2023. Luego, se realiza un cruce y análisis de los datos para identificar tendencias y elaborar estrategias basadas en los hallazgos, culminando con un informe que presenta conclusiones apoyadas por visualizaciones gráficas.

• Desarrollo

Este proyecto consta de los siguientes elementos de desarrollo de software:

- 563 líneas de código.
- 5 módulos.
- 25 funciones.
- 1 clases.
- 1 fichero de datos. (df que se crea)
- Librerías:
 - pandas, matplotlib.pyplot, numpy, scrapy, re, os, shutil, requests, json.

Hemos utilizado librerías fundamentales para la adquisición y manipulación de datos. Pandas, matplotlib han sido cruciales en la visualización de datos. Scrapy, por su parte, ha sido esencial para recolectar los datos directamente de la web, y requests nos ha permitido interactuar con APIs para obtener información actualizada. Juntas forman el soporte técnico para extraer y procesar.



PROYECTO-FINAL-ADQUIS...
 > __pycache__
 > .VSCodeCounter
 > cache
 > .gitignore
 dataframe.csv
 dataframe.py
 ergast_pitstops_data.py
 formula1_spider.py
 graphs.py
 informe.pdf
 main.py
 Poster_ProyectoFinal...
 ProyectoFinal_Formul...
 README.md

filename	language	code	comment	blank	total
README.md	Markdown	12	0	4	16
dataframe.py	Python	96	16	32	144
ergast_pitstops_data.py	Python	196	26	50	272
formula1_spider.py	Python	156	10	33	199
graphs.py	Python	93	0	18	111
main.py	Python	22	3	10	35

Figura 1. Estructura de módulos

• Resultados

En el primer apartado, a través de la página de Wikipedia pudimos obtener datos detallados de cada Gran Premio, incluyendo información sobre los pilotos, sus tiempos de vuelta y las estrategias de pit-stop.



Pos	Driver	DriverNumber	Driver	NPitstops	MedianPitStopDuration	Season	RaceNumber
1	bruno_senna	1	Bruno Senna	3	24.599	2012	1
2	ricciardo	3	Daniel Ricciardo	3	24.348	2012	1
3	massa	19	Felipe Massa	3	22.313	2012	1
4	rosberg	6	Nico Rosberg	2	23.11	2012	1
5	alonso	14	Fernando Alonso	2	21.9725	2012	1
6	kobayashi	10	Kamui Kobayashi	2	24.753999999999998	2012	1
7	resta	40	Paul di Resta	2	24.756	2012	1
8	webber	1	Mark Webber	2	23.1085	2012	1

Figura 2. Primer apartado Extracción de Datos F1 - Wikipedia

En el apartado 2 profundizamos en el procesamiento de los datos y aplicamos técnicas de manipulación de datos para limpiar, organizar y preparar los conjuntos de datos para un análisis más exhaustivo.



driverId	DriverNumber	Driver	NPitstops	MedianPitStopDuration	Season	RaceNumber
bruno_senna	1	Bruno Senna	3	24.599	2012	1
ricciardo	3	Daniel Ricciardo	3	24.348	2012	1
massa	19	Felipe Massa	3	22.313	2012	1
rosberg	6	Nico Rosberg	2	23.11	2012	1
alonso	14	Fernando Alonso	2	21.9725	2012	1
kobayashi	10	Kamui Kobayashi	2	24.753999999999998	2012	1
resta	40	Paul di Resta	2	24.756	2012	1
webber	1	Mark Webber	2	23.1085	2012	1

Figura 3. Resultados_Analiticos_TXT

En el apartado 3 se integran las fuentes de datos para sintetizar una visión unificada del rendimiento en las carreras. Este paso crítico nos permitió cruzar la información obtenida de Wikipedia con la extraída mediante web scraping y APIs, resultando en un único y robusto DataFrame.

	Pos	DriverNumber_x	Driver	...	DriverNumber_y	NPitstops	MedianPitStopDuration
0	1	9	Kimi Räikkönen	...	7.0	1.0	20.278
1	2	5	Fernando Alonso	...	14.0	1.0	19.368
2	3	1	Sebastian Vettel	...	5.0	2.0	22.440
3	4	3	Jenson Button	...	22.0	1.0	19.455
4	5	18	Pastor Maldonado	...	13.0	1.0	20.723
...
5038	Ret	14	Fernando Alonso	...	14.0	2.0	24.939
5039	Ret	81	Oscar Piastri	...	NaN	NaN	NaN
5040	Ret	31	Esteban Ocon	...	NaN	NaN	NaN
5041	DSQ	44	Lewis Hamilton	...	44.0	2.0	24.782
5042	DSQ	16	Charles Leclerc	...	16.0	1.0	23.963

[5043 rows x 15 columns]

Figura 4. Integración de Datos F1

El apartado 4 saca conclusiones sustentadas por los datos. La creatividad en la presentación de estos resultados, junto a la teoría permite no solo validar estrategias sino también descubrir oportunidades de mejora para el equipo.

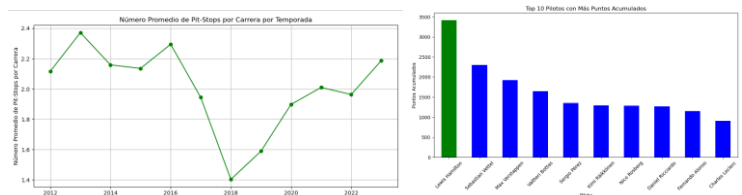


Figura 5. Visualización de Resultados Finales