

Obligatorio 2024

Propuesta

Se plantea la creación de un Data Warehouse basados en los archivos subidos por la cátedra respecto a la historia de la Fórmula 1. Para ello deberán analizar los distintos archivos que existen:

Circuitos / Resultados de constructores / Constructores / Pilotos / Clasificación de Constructores / Clasificación de pilotos / Tiempos de vuelta / Paradas / Clasificaciones / Carreras / Resultados / Temporadas / Status.

A partir de dichos archivos, los mismos deberán subirse a un área de Staging y luego cargar dichos datos a un data warehouse para responder a diferentes cruzamientos, como ser:

- Pilotos con más carreras
- Pilotos con más carreras ganadas en un período de tiempo
- Pilotos con más paradas en un período de tiempo
- Comparación de paradas vs carreras ganadas
- Resultados finales de campeonatos
- Resultados de constructores
- Si fuera el director de equipo y debe planificar las estrategias de parada, ¿cuál sería la mejor estrategia para cada circuito?
- Análisis que permita decidir cuestiones del estilo "Aquellos pilotos que siempre ganaron en Canadá, han terminado ganando el campeonato mundial", o más complejo "Aquellos pilotos que han ganado en Canadá, San Marino y Nürburgring, pero no en Hockenheim, han ganado el campeonato mundial"
- Los pilotos con mejor tiempo, ¿terminaron ganando la carrera? ¿terminan en el top 3?
- Dividiendo las vueltas de una carrera en tercios, ¿en que tercio se da la mayoría de los récords de vuelta?
- Dentro de los pilotos que no clasificaron en las primeras 3 posiciones y terminaron la carrera en las primeras 3 posiciones, ¿en qué posición de récord de vuelta estuvo?
- Si fuera director de equipo, ¿de qué nacionalidad buscaría sus pilotos para la próxima temporada?
- Alguno más que les resulte interesante

Es importante aplicar los diferentes conceptos que han visto en el curso.

Para ello deberán utilizar las herramientas del curso, para ETL Pentaho PDI y para la visualización de datos PowerBI.

Tareas

- Creación del modelo de staging y data warehouse creado a partir de los archivos de datos otorgados por la cátedra y teniendo en cuenta los requerimientos de información explicitados en el presente documento. **Entregable: Scripts de creación de las bases de datos y sus correspondientes estructuras (tablas, etc).**
- Carga de los datos usando la herramienta PDI. **Entregable: Jobs y Transformations de PDI para la carga de datos.**
- Presentación de los datos en PowerBI, cumpliendo al menos los siguientes requisitos:
 - Un reporte que contenga mapas
 - Un reporte que permita cruzar pilotos con escuderías por año
 - Un reporte que permita cruzar aquellos pilotos que han ganado más veces los campeonatos mundiales, con las respectivas escuderías
 - Un reporte que muestre las escuderías más ganadoras de la historia y de los últimos 10 años
 - Top 10 de pilotos que han ganado más veces en circuitos en Europa vs circuitos en América y viceversa.