

Escuela Técnica Superior de Ingeniería Informática

Grado en Ingeniería de Computadores Curso 2021-2022

Trabajo Fin de Grado

COMPARATIVA ENTRE LAS API DE SPARK EN SCALA Y PYTHON

Autor: Oscar Nydza Nicpoñ

Tutor: Juan Manuel Serrano Hidalgo

Agradecimientos

Breves agradecimientos o dedicatoria.

Resumen

Breve resumen del Trabajo de Fin de Grado (TFG). Recomendable entre 250-300 palabras, conteniendo los principales objetivos y resultados derivados del mismo.

Palabras clave:

- Python
- Ciberseguridad
- Aprendizaje automático (pueden ser varias)
- . . .

Índice de contenidos

Ín	lice de figuras	IX
Ín	lice de códigos	XI
1.	Introducción	III
	1.1. Contexto y alcance	1
	1.2. Estructura del documento	1
	1.2.1. Trabajos de grados en informática	1
	1.2.2. Trabajos del grado en matemáticas	2
2.	Objetivos	3
3.	Descripción Informática	6
	3.1. Fuentes de datos	7
	3.1.1. Tablas de entrada	7
	3.1.2. Diagramas de la estructura de las tablas	10
	3.2. Programación de queries en PySpark	10
	3.3. Programación de queries en Scala/Spark	10
	3.3.1. Piloto más consistente de la temporada 2012	10
	3.3.2. Dominio de fabricantes en la década de los 90	12
	3.4. Despliegue en AWS EMR	13
4.	Experimentos / Validación	14
	4.1. Análisis de requisitos no funcionales	15
5.	Conclusiones y trabajos futuros	16
	5.1. Texto de relleno	17
Bi	oliografía	22
Aı	éndices	24
Α.	Apéndice de figuras	26
	A.1. Tablas de entrada	26

Índice de figuras

A.1.	Tabla c	circuits .							 				•			27
A.2.	Tabla c	constructc	r_re	sult	S											27
A.3.	Tabla c	constructc	r_st	and	lin	gs										27
A.4.	Tabla c	constructo	rs													28
A.5.	Tabla d	driver_star	ndin	gs												28
A.6.	Tabla la	ap_times														28
A.7.	Tabla p	oit_stops														29
A.8.	Tabla c	qualifying														29
A.9.	Tabla r	aces														29
A.10	.Tabla r	esults														30
A.11	.Tabla s	seasons .														30
A.12	.Tabla s	status														30
A.13	.Tabla d	drivers .														31
A.14	.Tabla r	esults														31

Índice de códigos

Introducción

Se puede añadir texto antes de empezar la primera sección.

1.1. Contexto y alcance

Contexto. Situar al lector. Objetivo general y alcance del trabajo.

1.2. Estructura del documento

La estructura del TFG no es fija. El tutor indicará una estructura adecuada dependiendo del trabajo concreto.

Se puede incluir dentro de cada apartado secciones adicionales. La copia en papel de la memoria del TFG será encuadernada en pasta dura de color azul (p.e. encuadernación tipo chanel). La portada, que puede ser una pegatina transparente, seguirá el modelo que se adjunta, que incluye el escudo y nombre de la URJC, la titulación cursada por el alumno, el curso académico, el título del TFG, el autor y el o los directores/tutores.

1.2.1. Trabajos de grados en informática

Una posible estructura de la memoria final asociada con cada TFG podría ser la siguiente (leed la normativa de TFG):

- 1. Introducción
- 2. Objetivos (incluyendo descripción del problema, estudio de alternativas y metodología empleada)
- 3. Descripción informática (puede incluir especificación, diseño, implementación y pruebas).
- 4. Experimentos / validación
- 5. Conclusiones (incluyendo los logros principales alcanzados y posibles trabajos futuros)
- 6. Bibliografía
- 7. Apéndices

1.2.2. Trabajos del grado en matemáticas

Una posible estructura de la memoria final asociada con cada TFG podría ser la siguiente:

- 1. Introducción
- 2. Objetivos (incluyendo descripción del problema, estudio de alternativas y metodología empleada)
- 3. Material y métodos / Metodología / Cuerpo del trabajo (describir las metodologías empleadas en el desarrollo del TFG o el desarrollo del mismo en caso de ser un trabajo de recopilación bibliográfica sobre un tema).
- 4. Resultados (opcional, dependiendo del tipo de trabajo desarrollado)
- 5. Conclusiones (incluyendo los logros principales alcanzados y posibles trabajos futuros)
- 6. Bibliografía
- 7. Apéndices

2 Objetivos

El principal objetivo de este Trabajo de Fin de Grado realizar una comparativa entre las API de Spark de Scala y de Python. Sin embargo, también existen otros objetivos que irán surgiendo mientras se avanza sobre el principal. Entre ellos estarían:

- 1. Aprender Scala como lenguaje funcional: para ello utilizaré distintos recursos bibliográficos, pero principalmente "Programming In Scala", 4ª edición, de Martin Odersky y "Functional Programming in Scala" 1ª edición, de Paul Chiusano. Se aprenderán conceptos como las funciones de orden superior, tipos de datos algebráicos, currificación de funciones y recursividado la correspondencia de Curry-Howard.
- 2. Aprender Spark como API tanto para Scala como para Python: para ello principalmente usaré el libro "Spark: The Definitive Guide", de Bill Chambers. Aprenderé conceptos como la API de DataFrames, Spark SQL en profundidad cómo utilizar la Spark UI para obtener métricas del proceso.
- 3. Visualizar de los resultados de las queries realizadas usando Plotly.
- 4. Migrar queries desde PySpark a Scala Spark, centrando la explicación en las diferencias entre ambas APIs y en detalles a tener en cuenta al hacer una migración de este estilo.
- 5. Medir y comparar el rendimiento de ambas API utilizando la Spark UI, que proporciona métricas de rendimiento en tiempo y memoria.
- 6. Realizar queries a un cluster AWS EMR.

Como objetivo adicional y debido a que se usará un dataset de la Fórmula 1, se buscará encontrar la siguiente información:

- Piloto más consistente en un periodo de tiempo concreto: se calculará la diferencia entre el tiempo medio de todas las vueltas de cada piloto ese periodo de tiempo en concreto y la media de sus vueltas más rápidas.
- Piloto más dominante en un periodo de tiempo concreto calculando valores estadísticos como el total de carreras ganadas, el total de títulos, el número de vueltas lideradas, el número de primeras posiciones en clasificación, número de vueltas rápidas, etc. Todo ello relativo a su periodo de actividad.
- Similar al punto anterior, pero con fabricantes. Normalmente cada fabricante tiene varios pilotos, así que se tomarán como valor la media de todos los pilotos en cada métrica.
- En base a lo anterior, cuál ha sido el peor año de esa marca en ese periodo de tiempo teniendo en cuenta resultados de carrera, problemas de fiabilidad y paradas en boxes.

- Análisis de temporada por pilotos y constructores: se calcularán diversas medidas estadísticas para cada piloto o fabricante (utilizando la media de los valores de los pilotos en caso del fabricante). Por ejemplo, el total de podios, el porcentaje de carreras en las que se ha acabado en podio, la media de posiciones perdidas y ganadas por carrera, el número de vueltas lideradas, etc.
- Temporada más interesante para el espectador, teniendo en cuenta métricas como el número de adelantamientos, accidentes, retiradas de pilotos, más cambios de líder en la clasificación general, etc.

Descripción Informática

3.1. Fuentes de datos

Como se mencionó brevemente en el apartado de Objetivos, se ha utilizado un conjunto de datos de la Fórmula 1 que fue obtenido del siguiente enlace: click aquí. Concretamente, este dataset tiene 13 tablas que proporcionan información sobre distintos aspectos de esta competición. Estas tablas son:

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

3.1.1. Tablas de entrada

En este apartado se proporcionará una descripción de los datos más útiles que contiene cada una de las tablas de entrada además de una visualización de las mismas.

Tabla circuits

Esta tabla contiene información sobre todos los circuitos en los que se ha llevado a cabo un Gran Premio. Las columnas más interesantes son el nombre del circuito, una referencia textual y la localización. A.1

Tabla constructor_results

Esta tabla nos proporciona información sobre los resultados de las carreras en base a los constructores. A.2

Tabla constructor_standings

Esta tabla contiene información sobre la clasificación de constructores. Como particularidad, tiene una entrada por carrera y constructor participante. Por tanto, podríamos ver cómo ha ido cambiando la clasificación de constructores a lo largo del campeonato.

Las columnas más interesantes son el identificador de la carrera, identificador del constructor, los puntos, la posición en la clasificación y las victorias hasta ese punto. A.3

Tabla constructors

Esta tabla contiene información sobre los distintos constructores que han participado en algún campeonato mundial de Fórmula 1. Las columnas más interesantes son el id de constructor, la referencia, el nombre del constructor y la nacionalidad. A.4

Tabla driver_standings

Similar a la tabla de clasificación de constructores, pero para pilotos. Tenemos las mismas columnas, salvo que en lugar de tener un id de constructor, lo tenemos de piloto. A.5

Tabla lap_times

Esta tabla es una de las más interesantes, ya que nos da todos los tiempos de vuelta de todos los pilotos desde que hay registros. Esto es, desde parte de 1996 y 1997 al completo.

Las columnas más llamativas podrían ser el id de carrera, el de piloto, la vuelta en cuestión, la posición y el tiempo en milisegundos. A.6

Tabla pit_stops

Esta tabla contiene información de las paradas en boxes. Las columnas más interesantes son los id de carrera y piloto, el índice de parada (si es la primera, segunda, etc), la vuelta en la que se hace y la duración en milisegundos. A.7

Tabla qualifying

Esta tabla nos da información sobre los resultados de todas las rondas de clasificación. La columnas más interesantes son la posición final y los tiempos en Q1, Q2 y Q3. A.8

Tabla races

Esta tabla contiene información sobre todas las carreras celebradas en la historia de la competición. Contiene columnas como el id del circuito, el nombre del Gran Premio, la fecha y el año en el que se celebró. Esta última quizá sea la más útil de todo el dataset, ya que es la única forma de filtrar las carreras o los resultados por temporada. A.9

Tabla results

Esta tabla es similar a la de resultados por constructor, pero para pilotos. Es la tabla más completa de todas, ya que nos proporciona una entrada por piloto y carrera con información relevante de cómo se ha desarrollado la misma. Las columnas más interesantes pueden ser la posición de salida y la posición final, los puntos ,las vueltas dadas, la vuelta más rápida, la velocidad más rápida y, en el caso de que haya habido algún incidente, el id del estado. A.10

Tabla seasons

Quizá se trate de la tabla menos útil, ya que solamente contiene una columna con el año y otra con una url a un artículo de Wikipedia para cada entrada. A.11

Tabla status

Esta tabla nos da información sobre los estados en los que ha podido acabar la carrera un piloto determinado. Contiene un identificador y el estado en cuestión. A.12

Tabla drivers

Contiene información sobre todos los pilotos que han competido a lo largo de la historia. En concreto la información más relevante puede ser el nombre y apellido, el código, la fecha de nacimiento y la nacionalidad. A.13

3.1.2. Diagramas de la estructura de las tablas

En la siguiente figura se puede apreciar el diagrama Entidad-Relación del conjunto de datos: A.14

3.2. Programación de queries en PySpark

3.3. Programación de queries en Scala/Spark

3.3.1. Piloto más consistente de la temporada 2012

En esta query intentaremos averiguar cuál ha sido el piloto más consistente de la temporada 2012. Ya que este término puede resultar ambigüo, en concreto intentaremos averiguar qué piloto tuvo una menor diferencia entre la media de sus vueltas rápidas del campeonato y la media de todas las vueltas de todos los Grandes Premios de la temporada.

Necesitaremos cruzar varias fuentes de datos para esto:

- races.csv
- lap_times.csv
- drivers.csv
- results.csv

Para filtrar cualquier query por temporada o por rango de temporadas, tenemos que filtrar la tabla races según la columna year. En nuestro caso, filtramos la tabla para la temporada 2012.

Una vez tenemos todas las carreras de la temporada en cuestión, necesitamos hallar el tiempo medio por vuelta de cada piloto en cada Gran Premio. En la tabla lap_times.csv existe una columna llamada milliseconds, que denota los milisegundos de la vuelta dada. Haciendo un right join, podemos obtener todas las vueltas dadas en una temporada. Usando una ventana que particione los datos por piloto, haremos la media de la columna milliseconds para hallar la media de toda la temporada.

Para obtener la media de las vueltas rápidas de cada piloto, debemos usar la tabla results, que tiene una entrada por cada piloto y Gran Premio que contiene además su vuelta más rápida según el formato MM:ss.mmm, siendo MM los minutos, ss los segundos y mmm los milisegundos. Para convertir este formato a milisegundos y viceversa, he creado dos UDFs llamadas lapTimeToMs y msTolapTime.

```
val msToLapTime = (time: Long) => {
  val mins = time / 60000
  val secs = (time - mins*60000)/1000
  val ms = time - mins*60000 - secs*1000

val formattedSecs = if((secs / 10).toInt == 0) "0" + secs else secs
  // if ms = 00x -> "0"+"0"+x . if ms = 0xx -> "0"+ms
  val formattedMs = if((ms / 100).toInt == 0) "0" + (if((ms / 10).toInt == 0) "0" + ms else ms) else ms
  mins + ":" + formattedSecs + "." + formattedMs
}: String
```

La función lapTimeToMs debe tener en cuenta también que hay entradas que no tienen un tiempo válido, sino que contienen el string "\\N". Esto se debe a que hay ocasiones en los que un piloto no termina la vuelta. Un ejemplo sería si se retirase de la carrera en esa vuelta. Se ha decidido sumar 3 minutos (180000 ms) como penalización.

Una vez tenemos estas dos funciones, ya podemos el tiempo medio de vueltas rápidas. De nuevo particionamos los datos según el piloto y hacemos la media de la columna fastestLapTime. Este valor en milisegundos se convierte al formato de tiempo de vuelta con la UDF correspondiente y tras esto se ordena por el diferencial hallado de forma ascendente.

Además de esto, haciendo un análisis previo se observó que había pilotos que habían completado pocas carreras en comparación al resto, ya que a que sustituían a algún otro piloto debido a algún accidente. Se decidió filtrar estos outliers de forma que, independientemente del periodo de tiempo sobre el que queramos lanzar esta query, el resultado fuese significativo. Para ello se usó la tabla obtenida con todas las carreras en la temporada y, a partir de la columna lap de la tabla lap_times se realizó un conteo de todas las vueltas dadas por cada piloto utilizando, de nuevo, una ventana de datos particionada por piloto.

De este DataFrame queremos obtener tanto la cuenta de vueltas por piloto como el numero total de pilotos que han participado en la temporada, que haremos de la siguiente manera:

```
val (distinctDrivers, allLaps) = lapCount
   .agg(
      countDistinct("driverID"),
      count(col("lap"))
).as[(BigInt, BigInt)]
   .collect()(0)
val avgLapsThisPeriod = allLaps.toInt / distinctDrivers.toInt
```

Habiendo obtenido la media de vueltas por piloto de la temporada, podemos filtrar aquellos cuyo conteo sea inferior.

Solo quedaría formatear los datos de salida. Para ello primero se hace un join con la tabla de pilotos, lo cual nos permite concatenar nombre y apellido de cada uno de los participantes en una nueva columna. En esta caso, como solo nos interesa el piloto y su diferencial, estas serán las únicas columnas que dejaremos en el DataFrame.

3.3.2. Dominio de fabricantes en la década de los 90

Con esta query se pretende hallar qué fabricante ha sido el más dominante en la década de los 90. En concreto intentaremos hallar el número de mundiales ganados y el número de carreras ganadas.

Se usarán las siguientes fuentes de datos:

- races.csv
- constructor_standings.csv
- constructors.csv

Al igual que en la query anterior, si queremos fijar nuestra atención en un periodo de tiempo, tenemos que hacerlo filtrando la columna year de la tabla races.csv. En este caso, necesitamos todas las carreras entre el año 1990 y el año 1999.

Una vez obtenidas todas las carreras de la década, tenemos que obtener la última carrera de cada temporada. Esto es debido a que en constructor_standings.csv tenemos la clasificación resultante al final de cada carrera. Para ello, usaremos crearemos una ventana en la que particionaremos los datos por año y que usaremos con la función max() sobre la columna round, que nos indica el índice de la carrera, es decir, la primera carrera de la temporada tendrá round === 1, para crear una columna llamada max en la que guardaremos el índice de la última carrera de la temporada. Finalmente, filtraremos los datos

para quedarnos con aquellos en los que la columna round === max.

Tras esto último, unimos las tabla constructor_standings.csv con la recién obtenida para quedarnos con los resultados en las últimas carreras y filtramos según la columna position === 1 para quedarnos con los ganadores. Teniendo esto, podemos ver también que la columna wins nos proporciona el número de victorias de cada escudería en esa temporada, así que, creando una ventana en la que particionemos por fabricante podemos hallar tanto la suma de victorias como el conteo de apariciones de cada una.

Pasando ya a la presentación de los datos, se filtrarían los constructores duplicados y se ordenarían los datos según el total de campeonatos ganados primero y, en caso de empate, por número de victorias. Además, se hace un join con la tabla de constructores para obtener su nombre.

Lo interesante de esta query es que se puede usar para cualquier periodo de tiempo. Podemos averiguar por ejemplo qué fabricante ha sido el más dominante en toda la historia de la competición y qué constructor ha dominado ciertos años concretos.

3.4. Despliegue en AWS EMR

Experimentos / Validación

4.1. Análisis de requisitos no funcionales

Conclusiones y trabajos futuros

En este capítulo se detallan las conclusiones derivadas del TFG y la propuesta de posibles trabajos futuros.

Las citas del texto Autor [1], Autor [2], Autor [3], Autor [4] y Autor [5] deben ir referenciadas en la bibliografia.

5.1. Texto de relleno

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc

vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis portitior. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit

ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus

quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Bibliografía

- [1] M. Giaquinta and S. Hildebrandt, *Calculus of variations II*. Springer Science and Business Media, 2013, vol. 311.
- [2] S. Fortune and C. J. Van Wyk, "Efficient exact arithmetic for computational geometry," in Proceedings of the Ninth Annual Symposium on Computational Geometry, 1993, pp. 163– 172.
- [3] S. Fortune, "Voronoi diagrams and delaunay triangulations," Computing in Euclidean geometry, pp. 225–265, 1995.
- [4] J. C. Mitchell, "Social networks," *Annual review of anthropology*, vol. 3, no. 1, pp. 279–299, 1974.
- [5] C. B. Morrey Jr, *Multiple integrals in the calculus of variations*. Springer Science and Business Media, 2009.





Apéndice de figuras

A.1. Tablas de entrada

Diagrama Entidad-Relación

+		+		+				++
circuitId	circuitRef	name	location	country	lat	lng	alt	url
1	albert park	Albert Park Grand	Melbourne	Australia	-37.8497	144.968	10	http://en.wikiped
j 2	sepang	Sepang Internatio	Kuala Lumpur	Malaysia	2.76083	101.738	18	http://en.wikiped
j 3	bahrain	Bahrain Internati	Sakhir	Bahrain	26.0325	50.5106	7	http://en.wikiped
4	catalunya	Circuit de Barcel	Montmeló	Spain	41.57	2.26111	109	http://en.wikiped
j 5	istanbul	Istanbul Park	Istanbul	Turkey	40.9517	29.405	130	http://en.wikiped
j 6	monaco	Circuit de Monaco	Monte-Carlo	Monaco	43.7347	7.42056	7	http://en.wikiped
7	villeneuve	Circuit Gilles Vi	Montreal	Canada	45.5	-73.5228	13	http://en.wikiped
j 8	magny cours	Circuit de Nevers	Magny Cours	France	46.8642	3.16361	228	http://en.wikiped
j 9	silverstone	Silverstone Circuit	Silverstone	į uki	52.0786	-1.01694	153	http://en.wikiped
10	hockenheimring	Hockenheimring	Hockenheim	Germany	49.3278	8.56583	103	http://en.wikiped
11	hungaroring	Hungaroring	Budapest	Hungary	47.5789	19.2486	264	http://en.wikiped
12	valencia	Valencia Street C	Valencia	Spain	39.4589	-0.331667	4	http://en.wikiped
13	spa	Circuit de Spa-Fr	Spa	Belgium	50.4372	5.97139	401	http://en.wikiped
14	monza	Autodromo Naziona	Monza	Italy	45.6156	9.28111	162	http://en.wikiped
15	marina bay	Marina Bay Street	Marina Bay	Singapore	1.2914	103.864	18	http://en.wikiped
16	fuji	Fuji Speedway	0yama	Japan	35.3717	138.927	583	http://en.wikiped
17	shanghai	Shanghai Internat	Shanghai	China	31.3389	121.22	5	http://en.wikiped
18	interlagos	Autódromo José Ca	São Paulo	Brazil	-23.7036	-46.6997	785	http://en.wikiped
j 19	indianapolis	Indianapolis Moto	Indianapolis	USA	39.795	-86.2347	223	http://en.wikiped
20	nurburgring	Nürburgring	Nürburg	Germany	50.3356	6.9475	578	http://en.wikiped

only showing top 20 rows

Figura A.1: Tabla circuits

+				
constructorResultsId	raceId	constructorId	points	status
+				+
1	18	1	14	\N
2	18	2	8	\N
3	18	3	9	\N
4	18	4	5	\N
5	18	5	2	\N
6	18	6	1	\N
7	18	7	0	\N
8	18	8	0	\N
9	18	9	0	\N
10	18	10	0	\N
11	18	11	0	\N
12	19	6	10	\N
13	19	2	11	\N
14	19	1	10	\N
15	19	7	5	\N
16	19	9	2	i (Ni
17	19	4	1	\N
18	19	11	0	ŃΝ
19	19	10	0	\N
20	19	3	0	\N
<u>+</u>				

only showing top 20 rows

Figura A.2: Tabla constructor_results

wins	sitionText	position p	points	constructorId	raceId	constructorStandingsId
1	1	1	14	1	18	1
0	3	3	8	2	18	j 2
0	2	2	9	3	18	j 3
0	4	4	5	4	18	4
0	5	5	2	5	18	5
0	6	6	1	6	18	6
1	1	1	24	1	19	7
0	2	2	19	2	19	8
0	4	4	9	3	19	9
0	5	5	6	4	19	10
0	8	8	2	5	19	11
1	3	3	11	6	19	12
0	6	6	5	7	19	13
0	7	7	2	9	19	14
0	9	9	0	11	19	15
0	10	10	0	10	19	16
0	11	11	0	8	19	17
1	3	3	28	1	20	18
0	1	1	30	2	20	19
0	4	4	10	3	20	20

Figura A.3: Tabla constructor_standings

+				h+
constructorId	constructorRef	name	nationality	url
+		·		h+
1	mclaren	McLaren	British	http://en.wikiped
2	bmw sauber	BMW Sauber	German	http://en.wikiped
] 3	williams	Williams	British	http://en.wikiped
4	renault	Renault	French	http://en.wikiped
5	toro_rosso	Toro Rosso		http://en.wikiped
6	ferrari	Ferrari	Italian	http://en.wikiped
7	toyota			http://en.wikiped
8		Super Aguri		http://en.wikiped
9	red_bull			http://en.wikiped
10	_	Force India		http://en.wikiped
11	honda			http://en.wikiped
12	spyker			http://en.wikiped
13	mf1	MF1	Russian	http://en.wikiped
14				http://en.wikiped
15	sauber	Sauber		http://en.wikiped
16	bar	BAR	British	http://en.wikiped
17	jordan		Irish	http://en.wikiped
18	minardi	Minardi	Italian	http://en.wikiped
19	jaguar	Jaguar	British	http://en.wikiped
20	prost	Prost	French	http://en.wikiped

only showing top 20 rows

Figura A.4: Tabla constructors

+	+					++
driverStandingsId	raceId	driverId	points	position	positionText	lwinsl
+	-					
1	18	1	10	1	1	1
2	18	2	8	2	2	0
3	18	3	6	3	3	0
4	18	4	5	4	4	0
5	18	5	4	5	5	0
6	18	6	3	6	6	0
7	18	7	2	7	7	0
8	18	8	1	8	8	0
9	19	1	14	1	1	1
10		2	11	3	3	0
11	19	3	6	6	6	0
12		4	6	7	7	0
13		5	10	4	4	0
14	19	6	3	9	9	0
15		7	2	10	10	0
16		8	11	2	2	1
17	19	9	8	5	5	0
18		15	5	8	8	0
19		17	2	11	11	0
20	19	14	0	12	12	0
+	4					++

only showing top 20 rows

Figura A.5: Tabla driver_standings

raceId	driverId	lap	position	time	milliseconds
+				+	+
841				1:38.109	
841	20			1:33.006	
841	20			1:32.713	
841	20	4	1	1:32.803	92803
841	20	5	1	1:32.342	92342
841	20	6	1	1:32.605	92605
841	20	7	1	1:32.502	92502
841	20	8	1	1:32.537	92537
841	20	9	1	1:33.240	93240
841	20	10	1	1:32.572	92572
841	20	11	1	1:32.669	92669
841	20	12	1	1:32.902	92902
841	20	13	1	1:33.698	93698
841	20	14	3	1:52.075	112075
841	20	15	j 4	1:38.385	98385
i 841i	20	16	2	1:31.548	91548
841	20	17	1	1:30.800	i 90800 i
841	20	18		1:31.810	
841	20	19		1:31.018	
841	20	20		1:31.055	
+				+	+
only sho	owing top	20 1	rows		

Figura A.6: Tabla lap_times

+		++			+	++
raceId driv	erId	stop	lap	time	duration	milliseconds
+		++			+	++
841	153	1	1	17:05:23	26.898	26898
841	30	1	1	17:05:52	25.021	25021
841	17	1	11	17:20:48	23.426	23426
841	4	1	12	17:22:34	23.251	23251
841	13	1	13	17:24:10	23.842	23842
841	22	1	13	17:24:29	23.643	23643
841	20	1	14	17:25:17	22.603	22603
841	814	1	14	17:26:03	24.863	24863
841	816	1	14	17:26:50	25.259	25259
841	67	j 1	15	17:27:34	25.342	25342
841	2	j 1 j	15	17:27:41	22.994	22994
841	1	j 1 j	16	17:28:24	23.227	23227
841	808	j 1 j	16	17:28:39	24.535	24535
841	3	j 1	16	17:29:00	23.716	23716
841	155	j 1	16	17:29:06	24.064	24064
841	16	j 1	16	17:29:08	25.978	25978
841	15	i 1	16	17:29:49	24.899	24899
841	18	j 1 j	17	17:30:24	16.867	16867
841	153	į zį	17	17:31:06	24.463	24463
841	5	i 1	17	17:31:11	24.865	24865
·		4			+	·

only showing top 20 rows

Figura A.7: Tabla pit_stops

+	+	4		+	+		+	
qualifyId rac	eId	driverId	constructorId	number	position	q1	q2	q3
+	4	4			+		+	
1	18	1	1	22	1	1:26.572	1:25.187	1:26.714
2	18	9	2	4	2	1:26.103	1:25.315	1:26.869
j 3 j	18	5	1	23	j 3	1:25.664	1:25.452	1:27.079
4	18	13	6	2	j 4	1:25.994	1:25.691	1:27.178
5	18	2	2	3	j 5	1:25.960	1:25.518	1:27.236
6	18	15	7	11	j 6	1:26.427	1:26.101	1:28.527
j 7 j	18	3	3	7	j 7	1:26.295	1:26.059	1:28.687
8	18	14	9	9	j 8	1:26.381	1:26.063	1:29.041
j 9 j	18	10	7	12	j 9	1:26.919	1:26.164	1:29.593
10	18	20	5	15	10	1:26.702	1:25.842	[\N[
11	18	22	11	17	11	1:26.369	1:26.173	[\N[
12	18	4	4	5	12	1:26.907	1:26.188	[\N[
13	18	18	11	16	13	1:26.712	1:26.259	\N
14	18	6	3	8	14	1:26.891	1:26.413	\N
15	18	17	9	10	15	1:26.914	\ N	[\N
16	18	8	6	1	16	1:26.140	\ N	\N
17	18	21	10	21	17	1:27.207	\N	\N
18	18	7	5	14	18	1:27.446	\N	\N
19	18	16	10	20	19	1:27.859	\N	\N
20	18	11	8	18	20	1:28.208	\N	\N
+	+			+	+	+	+	++

only showing top 20 rows

Figura A.8: Tabla qualifying

raceId	year	round	circuitId			name		date		time		url
1	2009	1	1	Australian	Grand		2009-	03-29	06:0	90:00	http://en.w	ikiped
2	2009	2	2	Malaysian	Grand	Prix	2009-	04-05	09:0	90:00	http://en.w	ikiped
j 3	2009	3	17	Chinese	Grand	Prix	2009-	04-19	07:0	90:00	http://en.w	ikiped
4	2009	4	3	Bahrain	Grand	Prix	2009-	04-26	12:0	90:00	http://en.w	ikiped
5	2009	5	4	Spanish	Grand	Prix	2009-	05 - 10	12:0	90:00	http://en.w	ikiped
6	2009	6	6	Monaco	Grand	Prix	2009-	05-24	12:0	90:00	http://en.w	ikiped
7	2009	7	5	Turkish	Grand	Prix	2009-	06-07	12:0	90:00	http://en.w	ikiped
8	2009	8	9	British	Grand	Prix	2009-	06-21	12:0	90:00	http://en.w	ikiped
9	2009	9	20	German	Grand	Prix	2009-	07-12	12:0	90:00	http://en.w	ikiped
10	2009	10	11	Hungarian	Grand	Prix	2009-	07-26	12:0	90:00	http://en.w	ikiped
11	2009	11	12	European	Grand	Prix	2009-	08-23	12:0	90:00	http://en.w	ikiped
12	2009	12	13	Belgian	Grand	Prix	2009-	08-30	12:0	90:00	http://en.w	ikiped
13	2009	13	14	Italian	Grand	Prix	2009-	09-13	12:0	90:00	http://en.w	ikiped
14	2009	14	15	Singapore	Grand	Prix	2009-	09-27	12:0	90:00	http://en.w	ikiped
15	2009	15	22	Japanese	Grand	Prix	2009-	10-04	05:0	90:00	http://en.w	ikiped
16	2009	16	18	Brazilian	Grand	Prix	2009-	10-18	16:0	90:00	http://en.w	ikiped
17	2009	17	24	Abu Dhabi	Grand	Prix	2009-	11-01	11:0	90:00	http://en.w	ikiped
18	2008	1	1	Australian	Grand		2008-	03-16	04:3	30:00	http://en.w	ikiped
19	2008	2	2	Malaysian	Grand	Prix	2008-	03-23	07:0	90:00	http://en.w	ikiped
20	2008	3	3	Bahrain	Grand	Prix	2008-	04-06	11:3	30:00	http://en.w	ikiped
only sho	wing	top 20	rows								·	

Figura A.9: Tabla races

1	18	1	1	22	1	1	1	1	10	58 1	:34:50.616	5690616	39	2	1:27.452	218.300	
2	18	2	2	3	5	2	2	2	8	58	+5.478	5696094	41	3	1:27.739	217.586	
3	18	3	3	7	7	3	3	3	6	58	+8.163	5698779	41	5	1:28.090	216.719	
4	18	4	4	5	11	4	4	4	5	58	+17.181	5707797	58	7	1:28.603	215.464	
5	18	5	1	23	3	5	5	5	4	58	+18.014	5708630	43	1	1:27.418	218.385	
6	18	6	3	8	13	6	6	6	3	57	\N	\N	50	14	1:29.639	212.974	
7	18	7	5	14	17	7	7	7	2	55	\N	\N	22	12	1:29.534	213.224	
8	18	8	6	1	15	8	8	8	1	53	\N	\N	20	4	1:27.903	217.180	
9	18	9	2	4	2	\N	R	9	0	47	\N	\N	15	9	1:28.753	215.100	
10	18	10	7	12	18	\N	R	10	0	43	\N	\N	23	13	1:29.558	213.166	
11	18	11	8	18	19	\N	R	11	Θ	32	\N	\N	24	15	1:30.892	210.038	
12	18	12	4	6	20	\N	R	12	Θ	30	\N	\N	20	16	1:31.384	208.907	
13	18	13	6	2	4	\N	R	13	0	29	\N	\N	23	6	1:28.175	216.510	
14	18	14	9	9	8	\N	R	14	0	25	\N	\N	21	11	1:29.502	213.300	
15	18	15	7	11	6	\N	R	15	Θ	19	\N	\N	18	10	1:29.310	213.758	
16	18	16	10	20	22	\N	R	16	Θ	8	\N	\N	8	17	1:32.021	207.461	
17	18	17	9	10	14	\N	R	17	0	Θ	\N	\N	\N	\N	\N	\N	
18	18	18	11	16	12	\N	R	18	0	θ	\N	\N	\N	/N	\N	\N	
19	18	19	8	19	21	\N	R	19	Θ	0	\N	\N	\N	/N	\N	\N	
20	18	20	5	15	9	\N	R	20	0	Θ	\N	\N	\N	\N	\N	\N	

Figura A.10: Tabla results

Figura A.11: Tabla seasons

+	
statusId	status
+	·+
1	Finished
1 2	Disqualified
j 3	
j 4	
j 5	Engine
j 6	Gearbox
7	Transmission
8	Clutch
9	
10	Electrical
11	+1 Lap
12	+2 Laps
13	+3 Laps
14	+4 Laps
15	+5 Laps
16	+6 Laps
17	+7 Laps
18	+8 Laps
19	+9 Laps
20	Spun off
+	<u> </u>
only shows	ing top 20 rows

Figura A.12: Tabla status

l u	nationality	dob	surname	forename	code	number	driverRef	driverId
http://en.wikiped.	British	1985-01-07	Hamilton	Lewis	HAM	44	hamilton	1
http://en.wikiped.	German	1977-05-10	Heidfeld	Nick	HEI	\N	heidfeld	2
http://en.wikiped.		1985-06-27		Nico	R0S	6	rosberg	3
http://en.wikiped.	Spanish	1981-07-29	Alonso	Fernando	ALO	14	alonso	4
http://en.wikiped.	Finnish	1981-10-19	Kovalainen	Heikki	KOV	\N	kovalainen	5
http://en.wikiped.	Japanese	1985-01-11	Nakajima	Kazuki	NAK	\N	nakajima	6
http://en.wikiped.	French	1979-02-28	Bourdais	Sébastien	BOU	\N	bourdais	7
http://en.wikiped.	Finnish	1979-10-17	Räikkönen	Kimi	RAI	7	raikkonen	8
http://en.wikiped.	Polish	1984-12-07	Kubica	Robert	KUB	88	kubica	9
http://en.wikiped.	German	1982-03-18	Glock	Timo	GLO	\N	glock	10
http://en.wikiped.	Japanese	1977-01-28	Sato	Takuma	SAT	\N	sato	11
http://en.wikiped.	Brazilian	1985-07-25	Piquet Jr.	Nelson	PIQ	\N	piquet_jr	12
http://en.wikiped.	Brazilian	1981-04-25	Massa	Felipe	MAS	19	massa	13
http://en.wikiped.	British	1971-03-27	Coulthard	David	COU	\N	coulthard	14
http://en.wikiped.	Italian	1974-07-13	Trulli	Jarno	TRU	\N	trulli	15
http://en.wikiped.	German	1983-01-11	Sutil	Adrian	SUT	99	sutil	16
http://en.wikiped.	Australian	1976-08-27	Webber	Mark	WEB	\N	webber	17
http://en.wikiped.	British	1980-01-19	Button	Jenson	BUT	22	button	18
http://en.wikiped.	British	1979-04-18	Davidson	Anthony	DAV	\N	davidson	19
http://en.wikiped.	German	1987-07-03	Vettel	Sebastian	VET	5	vettel	20

Figura A.13: Tabla drivers

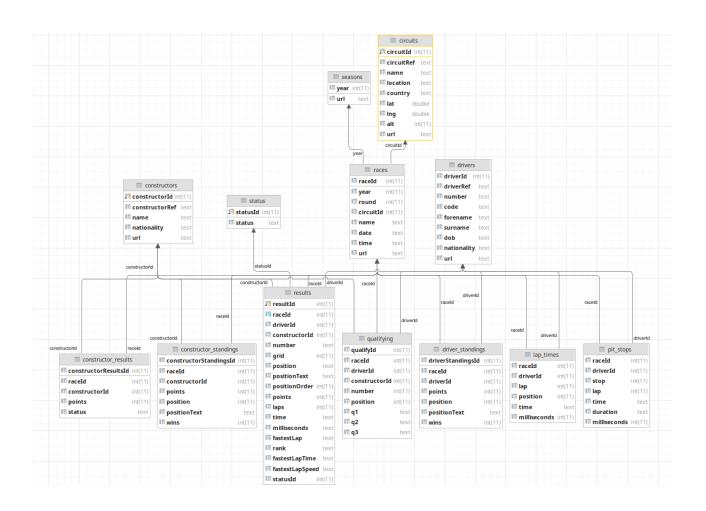


Figura A.14: Tabla results