

# Proyecto MongoDB

Uxía Taboada Nieto

Máster Big Data & Business Analytics 2022

## 1. Introducción

En este trabajo analizaremos algunos datos interesantes en relación a dos colecciones de fórmula 1 desde el año 1950 hasta el 2020. Las colecciones las hemos obtenido por el usuario **Deep Saha** en la página de **Kaggle**.

Antes de utilizar los datos, los he limpiado con Excel. Separando el nombre de los pilotos de sus abreviaturas en columnas diferentes y eliminando puntos y números de las localizaciones, al fin y al cabo dejando las colecciones lo más manejables posibles.

Una de ellas será la colección “drivers” que contiene los siguientes campos:

- Id : Identificador de cada piloto que ha corrido en la fórmula 1.
- Pos: Hace referencia a la posición en la que ha finalizado el campeonato mundial de pilotos en determinado año.
- Driver: Contiene Nombre y apellidos del piloto.
- Short: Abreviatura del piloto.
- Nationality: Nacionalidad del piloto.
- Car: Marca del coche de cada piloto.
- PTS: Total de puntos obtenido por año de cada piloto.
- Year: Año de disputa del mundial.

La otra colección es “races”, que contiene información de todas las carreras disputadas a lo largo de la historia de fórmula1, y contiene los siguientes campos:

- Location: País donde se disputó el GP.
- Date: Año en el que se corrió dicha carrera.
- P1,P2,P3,P5,P6,P7,P8,P9,P10: Contiene el nombre de los pilotos que terminaron en las posiciones de la 1 a la 10 respectivamente en cada GP.
- ShortP1, ShortP3,.....,ShortP10: Abreviaturas de los nombres de los piloto que han finalizado cada carrera en la posición 1 hasta la 10.

Después del proceso de limpieza, he importando los datos desde MongoDB Compass:



## 2. Queries

Procedí utilizando las dos colecciones indistintamente para las primeras queries, ya que ambas aportan información muy relevante. En las últimas “uno” las colecciones mediante un \$lookup a modo de simular el comportamiento de las bases de datos relacionales, ya que me parece muy interesante la manera que tiene esta función de unir las colecciones mediante arrays.

1. Lo primero que nos podríamos preguntar es quien ha ganado más títulos mundiales a lo largo de la historia desde la colección “drivers”.

```
db.drivers.aggregate([
  { $match: { Pos: "1" } },
  { $group: { _id: "$Driver", total: { $sum: 1 } } },
  { $sort: { total: -1 } } ])
```

	_id	total
1	Michael Schumacher	7
2	Lewis Hamilton	7
3	Juan Manuel Fangio	5
4	Sebastian Vettel	4
5	Alain Prost	4

2. También podríamos interesarnos para ver quién ha ganado más GP para comparar los resultados desde la colección “races”.

```
db.races.aggregate([
  { $group: { "_id": "$P1", "cuenta": { $sum: 1 } } },
  { $sort: { "cuenta": -1 } } ])
```

	_id	cuenta
1	Michael Schumacher	90
2	Lewis Hamilton	89
3	Sebastian Vettel	53
4	Alain Prost	51
5	Ayrton Senna	41
6	Fernando Alonso	32

3. Veamos cuantas carreras ha ganado Fernando Alonso.

```
db.races.find({"P1":"Fernando Alonso"}).count()
```

4. Como soy bastante fan de Lewis Hamilton, me interesó saber el total de victorias en su palmarés de cada circuito. Además es una buena forma de estudiar que circuitos se le dan mejor.

```
db.races.aggregate([
  { $match: { P1: "Lewis Hamilton" } },
  { $group: { _id: "$Location", total: { $sum: 1 } } },
  { $sort: { total: -1 } }
])
```

	_id	total
1	Hungary	8
2	Canada	7
3	Great Britain	7
4	China	6
5	Japan	6
6	United States	6
7	Italy	5
8	Spain	5

5. ¿Cuántas carreras ganó Hamilton en 2007?

```
var filtro1 = { P1: "Lewis Hamilton", "Date": "2007" }
```

```
var filtro2 = { "Location": 1, "Date": 1, "P1": 1 }
```

```
db.races.find(filtro1, filtro2)
```

Location	Date	P1
Canada	2007	Lewis Hamilton
United States	2007	Lewis Hamilton
Hungary	2007	Lewis Hamilton
Japan	2007	Lewis Hamilton

6. Y como 2007 fue un año muy polémico entre Alonso y Hamilton, decidí buscar el resultado total de puntos de ese mundial de ambos pilotos para hacer una comparativa. El resultado fue sorprendente.

```
var filtro1 = { $or: [ { "Driver": "Fernando Alonso" },
```

```
{ "Driver": "Lewis Hamilton" } ], "Year": "2007" }
```

```
var filtro2 = { "Driver": 1, "Year": 1, "PTS": 1 }
```

```
db.drivers.find(filtro1, filtro2)
```

Driver	PTS	Year
Lewis Hamilton	1090	2007
Fernando Alonso	1090	2007

**7. ¿Cuántos mundiales ha Ganado Alonso?, ¿En que años? Y ¿con que coche?**

```
db.drivers.find({ "Driver": "Fernando Alonso", "Pos": "1" }).count()
```

```
var query = { "Driver": "Fernando Alonso", "Pos": "1" }
```

```
var project1 = { "Pos": 1, "Year": 1, "Car": 1, }
```

```
db.drivers.find(query, project1)
```

Pos	Car	Year
1	Renault	2005
1	Renault	2006

**8. Podríamos preguntarnos que nacionalidad ha predominado más a lo largo de la historia.**

Ojo. Aquí se muestran los pilotos repetidos porque están agrupados por año.

```
var query = { "Driver": "Fernando Alonso", "Pos": "1" }
```

```
var project1 = { "Pos": 1, "Year": 1, "Car": 1, }
```

```
db.drivers.find(query, project1)
```

_id	total
GBR	285
ITA	192
FRA	177
USA	133
GER	130
BRA	110
FIN	60
AUS	45
AUT	44
ESP	43

9. Nombre y abreviatura de todos los pilotos con nacionalidad ESP por año.

```
var query = { "Nationality": "ESP" }
```

```
var project1 = { "Driver":1, "Short":1, "Pos": 1, "Year":1, "Car":1 }
db.drivers.find(query,project1)
```

Pos	Driver	Short	Car	Year
7	Paco Godia	GOD	Maserati	1956
15	Alfonso de Portago	DEP	Ferrari	1956
20	Alfonso de Portago	DEP	Ferrari	1957
26	Luis Perez-Sala	PER	Minardi Ford	1989
17	Pedro de la Rosa	DLR	Arrows	1999
18	Marc Gene	GEN	Minardi Ford	1999
16	Pedro de la Rosa	DLR	Arrows Supertec	2000
19	Marc Gene	GEN	Minardi Fondmetal	2000
16	Pedro de la Rosa	DLR	Jaguar Cosworth	2001
21	Pedro de la Rosa	DLR	Jaguar Cosworth	2002

10. Para ver que coche y por ello el motor es el mejor de la historia, hagamos una query para ver el coche que más mundiales ha ganado en la historia.

```
db.drivers.aggregate([
  { $match: { Pos: "1" } },
  { $group: { _id: "$Car", total: { $sum: 1 } } },
  { $sort: { total: -1 } }])
```

_id	total
Ferrari	15
Mercedes	7
Lotus Ford	4
Williams Renault	4
McLaren Honda	4
McLaren TAG	3
McLaren Mercedes	3
Red Bull Racing Renault	3

11. Visto que el coche que mas veces ganó es Ferrari, ¿Qué pilotos han ganado con ese coche?

```
var query = { "Car": "Ferrari", "Pos": "1" }
```

```
var project1 = { "Pos": 1, "Year": 1, "Car": 1, "Driver": 1 }
```

```
db.drivers.find(query, project1)
```

Pos	Driver	Car	Year
1	Alberto Ascari	Ferrari	1952
1	Alberto Ascari	Ferrari	1953
1	Juan Manuel Fangio	Ferrari	1956
1	Mike Hawthorn	Ferrari	1958
1	Phil Hill	Ferrari	1961
1	John Surtees	Ferrari	1964
1	Niki Lauda	Ferrari	1975
1	Niki Lauda	Ferrari	1977
1	Jody Scheckter	Ferrari	1979
1	Michael Schumacher	Ferrari	2000
1	Michael Schumacher	Ferrari	2001
1	Michael Schumacher	Ferrari	2002
1	Michael Schumacher	Ferrari	2003
1	Michael Schumacher	Ferrari	2004
1	Kimi Räikkönen	Ferrari	2007

12. Ahora veamos algo más curioso, hagamos la misma pregunta de que coche ganó mas mundiales pero en los últimos 20 años.

```
db.drivers.aggregate([
  { $match: { Pos: "1", Year: { $gt: "2000" } } },
  { $group: { _id: "$Car", total: { $sum: 1 } } },
  { $sort: { total: -1 } } ])
```

_id	total
Mercedes	7
Ferrari	5
Red Bull Racing Renault	3
Renault	2
Brawn Mercedes	1
RBR Renault	1
McLaren Mercedes	1

### 13. ¿Quién gano con Ferrari (y Mercedes) en los últimos 20 años?

```
var query = { "Car": "Ferrari", "Pos": "1", "Year": { $gt: "2000" } }
var project1 = { "Pos": 1, "Year": 1, "Car": 1, "Driver": 1 }
db.drivers.find(query, project1)
```

Pos	Driver	Car	Year	Pos	Driver	Car	Year
1	Michael Schumacher	Ferrari	2001	1	Lewis Hamilton	Mercedes	2014
1	Michael Schumacher	Ferrari	2002	1	Lewis Hamilton	Mercedes	2015
1	Michael Schumacher	Ferrari	2003	1	Nico Rosberg	Mercedes	2016
1	Michael Schumacher	Ferrari	2004	1	Lewis Hamilton	Mercedes	2017
1	Michael Schumacher	Ferrari	2007	1	Lewis Hamilton	Mercedes	2018
1	Kimi Räikkönen	Ferrari	2007	1	Lewis Hamilton	Mercedes	2019
				1	Lewis Hamilton	Mercedes	2020

### 14. Como sabemos, en la parrilla pueden haber coches de escuderías distintas pero que compartan el mismo motorista. Y en vista del dominio del motor Mercedes en los últimos años, querría saber que pilotos han corrido con el motor Mercedes.

```
busqueda= { "Car": { $regex: /Mercedes/ }, "Year": { $gt: "2000" } }
var project1 = { "Driver": 1, "Pos": 1, "Year": 1, "Car": 1 }
db.drivers.find(busqueda, project1)
```

Pos	Driver	Car	Year
17	Adrian Sutil	Force India Mercedes	2009
22	Vitantonio Liuzzi	Force India Mercedes	2009
4	Lewis Hamilton	McLaren Mercedes	2010
5	Jenson Button	McLaren Mercedes	2010
7	Nico Rosberg	Mercedes	2010
9	Michael Schumacher	Mercedes	2010
11	Adrian Sutil	Force India Mercedes	2010
15	Vitantonio Liuzzi	Force India Mercedes	2010

15. Veamos unos datos muy interesantes. Estaría bien calcular el número de mundiales disputados por piloto, la peor posición en la que terminaron, la mejor y la media entre esas dos posiciones. Esta es una buena manera de analizar en una sola tabla, la calidad de cada piloto.

Como los valores de “Pos” y “Year” estaban en formato *string*, para poder hacer estas cuentas, he tenido que convertirlos a *Int*.

```
PosQtyConversionStage = {
  $addFields: {

    Pos_converted: { $convert:
      {
        input: "$Pos",
        to: "int",
        onError: { $concat:
          [
            "Could not convert ",
            { $toString: "$qty" },
            " to type integer."
          ]
        },
        onNull: Int32("0")
      }
    },
  },
};
```



```

YearQtyConversionStage1 = {
  $addFields: {

    Year_converted: { $convert:
      {
        input: "$Year",
        to: "int",
        onError: { $concat:
          [
            "Could not convert ",
            { $toString: "$qty" },
            " to type integer."
          ]
        },
        onNull: Int32("0")
      } },
  }
};

```

```

db.drivers.aggregate( [
  PosQtyConversionStage,
  YearQtyConversionStage1,
  { $match: { Year_converted: { $gt: 1990 } } }
  { $group: { "_id": "$Driver", "conteo": { $sum: 1 }, "media": { $avg :
"$Pos_converted" } "peor": { $max: "$Pos_converted" },
"mejor": { $min: "$Pos_converted" } } }
  { $sort: { "media": 1 } }
])

```

_id ↕	conteo ↕	media ↕	peor ↕	mejor ↕
Lewis Hamilton	14	2,3571	5	1
Juan Manuel Fangio	8	2,875	14	1
Ayrton Senna	10	3,1	9	1
Jackie Stewart	9	3,4444	9	1
Alain Prost	13	3,4615	15	1
Stirling Moss	8	3,75	13	2
Nino Farina	6	3,8333	8	1
Michael Schumacher	19	3,9444	13	1
Rudi Fischer	1	4	4	4
Sebastian Vettel	14	4,4286	14	1

16. Con el mismo procedimiento, podríamos visualizar el nombre de todos los pilotos que hayan quedado en el top 3 del mundial desde el año 1990.

```
PosQtyConversionStage = {
  $addFields: {
    Pos_converted: { $convert:
      {
        input: "$Pos",
        to: "int",
        onError: { $concat:
          [
            "Could not convert ",
            { $toString: "$qty" },
            " to type integer."
          ]
        },
        onNull: Int32("0")
      }
    },
  },
};
YearQtyConversionStage1 = {
  $addFields: {
    Year_converted: { $convert:
      {
        input: "$Year",
        to: "int",
        onError: { $concat:
          [
            "Could not convert ",
            { $toString: "$qty" },
            " to type integer."
          ]
        },
      }
    },
  },
};
```

```

    ]
    },
    onNull: Int32("0")
  }},
}
};
db.drivers.aggregate( [
  PosQtyConversionStage, YearQtyConversionStage1,
  { $match: { Pos_converted:{$lte: 3}, Year_converted:{$gt:1990}} }
  { $group: { "_id": "$Driver" } }
])

```

_id
David Coulthard
Lewis Hamilton
Alain Prost
Felipe Massa
Damon Hill
Nigel Mansell
Rubens Barrichello
Kimi Räikkönen
Mark Webber

17. Como sabemos, no todos los años se corren el mismo números de Grandes Premios en cada mundial. ¿Cuántos GP se han disputado cada año? (Ordenados de mayor a menor)

```

db.races.aggregate( [
  { $group: { _id: "$Date", total: { $sum: 1 } } }
  {$sort: {total:-1}} ])

```

_id	total
2018	21
2019	21
2016	21
2012	20
2017	20
2014	19

**18. Concatenemos el nombre de cada piloto con su abreviación “Short”.**

```
var concatenar = { $concat: [ "$Driver", "->", "$Short" ] }

var nombre = { $toUpper: concatenar }

var fase1 = { $project: { _id: 0, "Nombre" : nombre } }

var etapas = [ fase1 ]

db.drivers.aggregate( etapas )
```



Nombre
NINO FARINA-> FAR
JUAN MANUEL FANGIO-> FAN
LUIGI FAGIOLI-> FAG
LOUIS ROSIER-> ROS
ALBERTO ASCARI-> ASC
JOHNNIE PARSONS-> PAR

**19. Por ultimo, quería saber quien ganó el GP tan polémico de Hungría 2007?**

Se podría hacer directamente desde la colección “races”, pero también me interesaría saber el coche con el que se ganó ese GP y el resto de información personal que nos proporciona la colección “drivers”.

```
db.drivers.aggregate( [
  {
    $lookup:
    {
      from: "races",
      let:{driver_qty: "$Year",driver_item: "$Driver" },
      pipeline: [
        { $match:

          { $expr:
            { $and:
              [
                {$eq :["$Date", "$$driver_qty"]},
                {$eq :["$Date", "2007"]},
              ]
            }
          }
        }
      ]
    }
  }
]
```

```

        { $eq: [ "$P1", "$$driver_item" ] },
        { $eq :["$Location", "Hungary"]},

    ]
}

}

},

{ $project: { driver_qty: 0, _id: 0,}


},
as: "data"
}

},

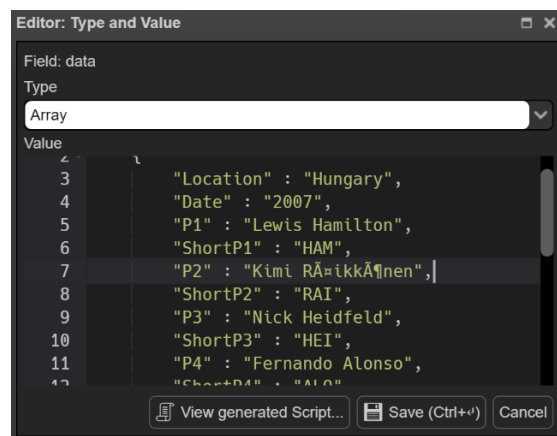
{ $match: { $expr: { $eq: [{ $size: "$data" }, 1]} } }},

])

```

Id ▾	Pos ▾	Driver ▾	Short ▾	Nationality ▾	Car ▾	PTS ▾	Year ▾	data ▾
1247	2	Lewis Hamilton	HAM	GBR	McLaren Mercedes	1090	2007	 Array[1]

Si seleccionamos el Array[1], podremos ver que la consulta se ha realizado correctamente. Ya que nos ha buscado en la colección “races” la localización “Hungary”.



### 3. Conclusiones

Podemos ver que el piloto que más carreras ha ganado hasta la fecha de nuestro dataset, es Michael Schumacher, con 90 victorias, seguido de Hamilton, con 89. Aunque actualmente sabemos que ese récord ya ha sido batido por el piloto inglés en 2021. El récord que sigue manteniendo *El Káiser*, es el de más mundiales ganados, empatado con Lewis en un total de 7.

La consulta número 4, de que circuito ha ganado más veces determinado piloto, es particularmente interesante, sobre todo a la hora de realizar las estadísticas previas a las carreras y determinar que circuitos se le dan mejor a cada piloto.

Lo mismo ocurre con la query 7, de cuantas veces ha ganado el mundial cierto piloto y con que coche, ya que como sabemos, la fórmula 1 no solo es una competición de pilotos, si no de pura ingeniería, y por ello el coche que tenga un piloto entre las manos es fundamental. Con esta consulta, vemos de primera mano la calidad del piloto, ya que tendría menos mérito ganar el mundial con un coche claramente superior al del resto, que con uno que no sea el mejor de la parrilla. Como es el caso de Fernando Alonso, que ganó 2 veces el campeonato con su Renault R25, que claramente era un coche competitivo, pero sin duda no el mejor de la parrilla.

Este ejemplo es la antítesis de lo que ocurrió desde 2014 en la fórmula1 con el comienzo de la era híbrida. Año a partir del cual comenzó un dominio aplastante de Mercedes sobre el resto de competidores, razón por la cual Hamilton tiene que agradecer los últimos 6 mundiales que tiene en su palmarés. Esto se puede ver claramente en la query 7 cambiando el nombre del piloto. Vemos también que el único año que Lewis no ganó desde 2014, fue 2016, que ganó Nico Rosberg, compañero de equipo del inglés durante ese año. Analizando un poco más los datos, vemos que a partir de ese año, Rosberg se retiró de la fórmula 1, sospechoso, ya que es raro que un piloto se retire el mismo año que gana su primer título. Podríamos especular sobre su relación con Hamilton, pero a tanto no llegan los datos.

Otros análisis interesantes hacen referencia a la escudería de renombre Ferrari. Es cierto que se ha ganado la fama, ya que es el coche que más títulos mundiales ha ganado en toda la historia como podemos ver en la query 10. Pero, acotando la consulta a los últimos 20 años, vemos que ese título lo porta Mercedes (query 12). Y en la consulta 13 de quien ganó con Ferrari (y Mercedes) en los últimos años, vemos que la última vez que ganó un coche de la escudería italiana fue en el año 2007 gracias a Kimi Räikkönen. Es decir, que viendo los datos se ve que Mercedes les está ganando terreno. Pareciera que la escudería se haya estancado y que viva de la historia, y viendo las carreras, el pésimo rendimiento del equipo Ferrari es evidente.

La consulta que personalmente más interesante me resulta, es la 15. Si eliminamos el \$match con el filtro de la fecha, y analizamos todos los datos desde 1950. Basándonos en los datos de la tabla, podríamos decir que Hamilton es el mejor piloto de la historia porque en toda su carrera ha terminado el mundial de media en el segundo puesto, y resulta

impresionante a primera vista que su peor posición haya sido quinto , seguido de Fangio, corredor de f1 en los años 50, con una media de 2,8 y Ayrton Senna, con una media de 3,1. Es innegable que estos tres pilotos tenían o tienen una gran calidad, y además han podido competir con coches muy superiores. Aun así, sería muy interesante ser capaces de determinar el porcentaje de mérito que tiene el piloto frente a la relevancia de su coche.