

CODERHOUSE

Data Analytics

Comisión: 29690. Alumno: Matias Mananian.



TOUR
DE
FRANCE

Contenido

1. Introducción	3
2. Objetivo.....	3
3. Alcance.....	3
4. Usuario final y nivel de aplicación del análisis.....	3
5. Herramientas tecnológicas implementadas	3
6. Base de datos relacionada a Power BI	4
7. Diagrama entidad-relación de las tablas seleccionadas.	4
8. Listado de tablas, con definición de clave primaria y/o foránea, según corresponda.	5
9. Listado de columnas por tablas, con definiciones de tipos de datos.	5
10. Segmentaciones elegidas.....	6
11. Transformaciones realizadas.	6
12. Visualización de los datos	8
13. Futuras actualizaciones.....	12

1. Introducción

El TDF es una carrera ciclista anual de etapas múltiples que se lleva a cabo principalmente en Francia. Aunque ocasionalmente pasa por países cercanos.

Es de las competiciones deportivas más vistas del mundo. Emitida en 180 países y vista por 3.500 millones de personas cada año.

El Tour se creó en 1903 y apenas se detuvo durante las 2 Guerras Mundiales.

2. Objetivo

Se quiere explorar la historia del TDF desde una perspectiva de datos.

La idea es sacar afirmaciones sobre los corredores y sobre la carrera en general.

3. Alcance

Para realizar el estudio se tomó en cuenta toda la información encontrada desde 1930 a 2016.

El 22 de octubre de 2012, Lance Armstrong es despojado formalmente de los siete títulos del Tour de Francia que ganó entre 1999 y 2005. Se le prohibió de por vida competir en ciclismo después de comprobarse que había usado drogas ilegales.

Para este trabajo no se tuvo en cuenta los datos de esos años.

4. Usuario final y nivel de aplicación del análisis.

El mismo apunta al público en general, que quizás no conoce de la carrera, y a personas que tienen una noción de la misma y quieren conocer aún más.

5. Herramientas tecnológicas implementadas

Para el presente trabajo se utilizaron los siguientes programas:

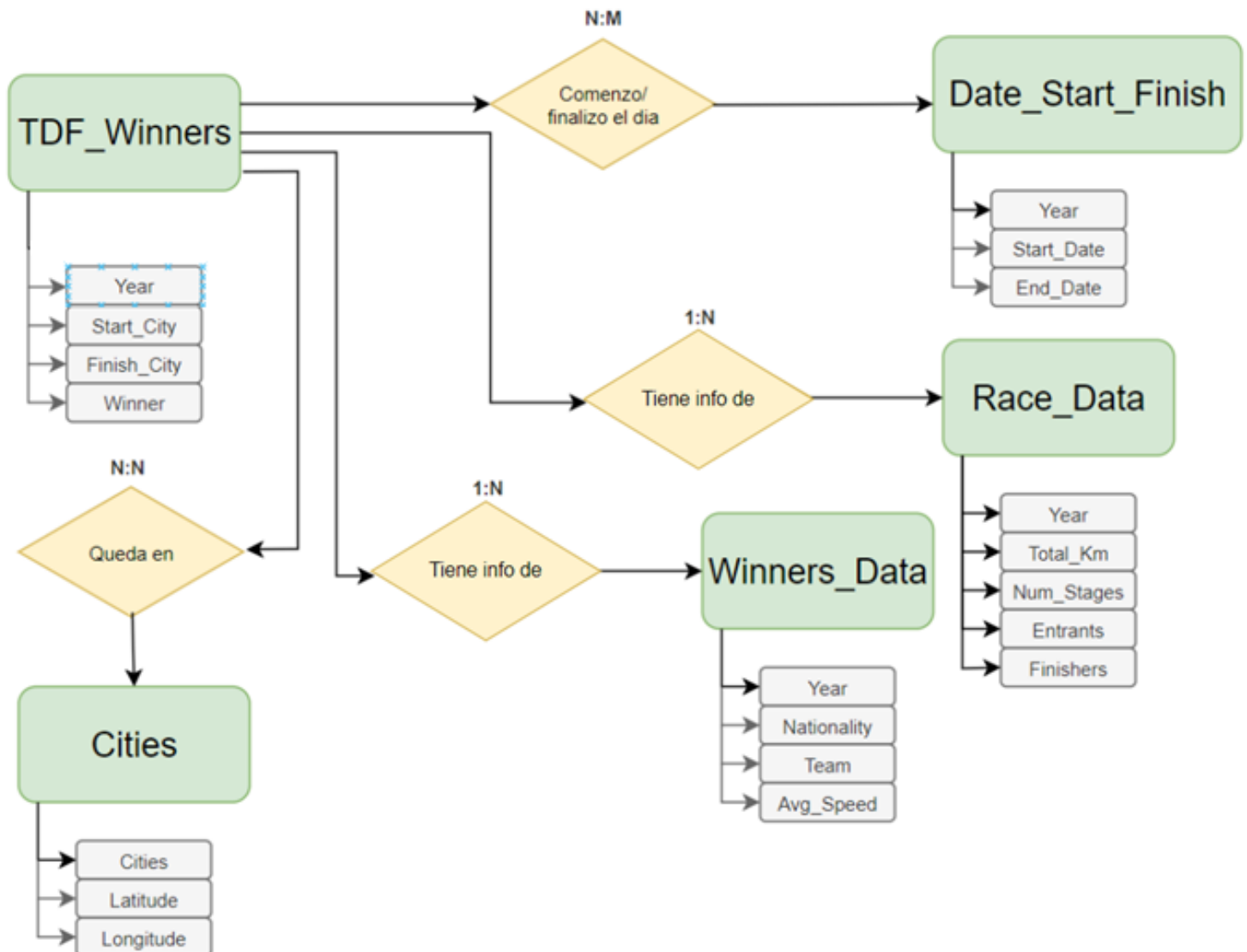
- Excel para la lectura y limpieza de los datasets.
- Draw.io para la creación del diagrama entidad-relación (<https://app.diagrams.net>).
- Power BI Desktop para la creación del tablero de control.

6. Base de datos relacionada a Power BI

Con el presente documento, se adjunto el archivo

Tour_De_France_Matias_Mananian_Entrega_Final.xlsx en donde se encuentra la bdd.

7. Diagrama entidad-relación de las tablas seleccionadas.



8. Listado de tablas, con definición de clave primaria y/o foránea, según corresponda.

- **TDF_Winners:** Contiene el año, el país de inicio, la ciudad donde inicio/finalizo la carrera y el nombre del ganador.
PK: Year.
FK: Start_City.
FK: Finish_City.
FK: Winner.
- **Date_Start_Finish:** Contiene el día de inicio y final según el año.
PK: Year.
- **Race_Data:** Contiene el año, el total de km recorridos, el número de etapas, la cantidad de corredores que comenzaron y que finalizaron.
PK: Year.
- **Winners_Data:** Contiene el año, la nacionalidad, el nombre del equipo y la velocidad promedio del corredor.
PK: Year.
- **Race_Data:** Contiene el nombre de las ciudades con su
FK: Cities.

9. Listado de columnas por tablas, con definiciones de tipos de datos.

TDF_Winners		
Campo	Tipos de campo	Tipos de clave
Year	int	PK
Start_City	varchar(20)	FK
Finish_City	varchar(20)	FK
Winner	varchar(20)	FK

Date_Start_Finish		
Campo	Tipos de campo	Tipos de clave
Year	int	PK
Start_Date	datetime	-
End_Date	datetime	-

Race_Data		
Campo	Tipos de campo	Tipos de clave
Year	int	PK
Total_Km	int	-
Num_Stages	int	-
Entrants	int	-
Finishers	int	-

Winners_Data		
Campo	Tipos de campo	Tipos de clave
Year	int	PK
Nationality	varchar(20)	-
Team	varchar(20)	-
Avg_Speed	int	-

Cities		
Campo	Tipos de campo	Tipos de clave
City	int	FK
Latitude	int	-
Longitude	int	-

10. Segmentaciones elegidas.

Se agregó un slicer para poder filtrar los datos por años. Con el mismo se puede elegir un año en particular o un grupo de años.

11. Transformaciones realizadas.

- Se creó la tabla **Table_Calendar**.
- Debido que hay resultados que fueron eliminados del historial, se creó un filtro global para que no se tomen en cuenta.
- En la página **Inicio** se crea un marcador **ClickHereInicio**. El mismo es un cuadro de texto en donde se explica porque fueron eliminados del historial los resultados mencionados anteriormente. Se accede al mismo presionando sobre un cuadro de texto.
- Se crean filtros por página y por elemento para mostrar los resultados queridos.
- Se crean tooltips para que aparezca más información si se pasa por arriba del gráfico.
- En la tabla **Cities** se creó la siguiente medida:

Count_Cities = DISTINCTCOUNT(Cities[City])

La misma cuenta la cantidad de ciudades por la que paso el TDF.

- En la tabla **Date_Start_Finish** se creó las siguientes medidas:

Ave_Duration = AVERAGE(Date_Start_Finish[Duration])

Promedio de días que duro el TDF en el lapso de tiempo estudiado.

Max_Duration = MAX(Date_Start_Finish[Duration])

Promedio máximo de días que duro el TDF en el lapso de tiempo estudiado.

Min_Duration = MIN(Date_Start_Finish[Duration])

Promedio mínimo de días que duro el TDF en el lapso de tiempo estudiado.

- En la tabla **Race_Data** se crearon las siguientes medidas:

Max_Km = MAX(Race_Data[Total_Km])

Promedio máximo de km recorridos en una edición.

Min_Km = MIN(Race_Data[Total_Km])

Promedio mínimo de km recorridos en una edición.

Sum_Entrants = SUM(Race_Data[Entrants])

Cantidad total de corredores que inicio el TDF en el lapso de tiempo estudiado.

Cantidad total de corredores que finalizaron el TDF en el lapso de tiempo estudiado.

Cantidad total de corredores que finalizaron el TDF en el lapso de tiempo estudiado.

Cantidad total de km recorridos en el lapso de tiempo estudiado.

- Promedio de velocidad en el lapso de tiempo estudiado.

Cantidad de nacionalidades de los corredores del TDF en el lapso de tiempo estudiado.

Cantidad de equipos en el lapso de tiempo estudiado.

Máximo promedio de velocidad de un ganador en el lapso de tiempo estudiado.

Mínimo promedio de velocidad de un ganador en el lapso de tiempo estudiado.

- $$\text{Ave_Withdrawals} = ([\text{Withdrawals}] * 100) / \text{Entrants}$$

- ```
Duration = Duration.Days([EndDate]-[Start_Days])
```

- En la tabla **Cities** se creó la función **Fx\_Country** y la columna **Country**. Lo que la función hace es conectarse a la API de BING y averiguar el país teniendo como dato la longitud y la latitud. La columna son los resultados de la función.

```
Xml.Tables(Web.Contents("http://dev.virtualearth.net/REST/v1/Locations/"&lat&","&long&"?&format=xml"))
```



```

o=xml&key=yprxQ0r9fx0EIzqNjn4m~bOwNFoUgWB2FolmEWx5w9w~AidsLHcSsfkFp-
DbTBR0fyyQC0NjIw_Cqwndw5FLgb48UBfXG5ZUbVP0Q7tOzg04")),
 #"Tipo cambiado" = Table.TransformColumnTypes(Origen,{{"Copyright", type text},
{"BrandLogoUri", type text}, {"StatusCode", Int64.Type}, {"StatusDescription", type text},
{"AuthenticationResultCode", type text}, {"TraceId", type text}}),
 ResourceSets = #"Tipo cambiado"{0}[ResourceSets],
 ResourceSet = ResourceSets{0}[ResourceSet],
 #"Tipo cambiado1" = Table.TransformColumnTypes(ResourceSet,{{"EstimatedTotal",
Int64.Type}}),
 Resources = #"Tipo cambiado1"{0}[Resources],
 Location = Resources{0}[Location],
 #"Tipo cambiado2" = Table.TransformColumnTypes(Location,{{"Name", type text},
{"Confidence", type text}, {"MatchCode", type text}}),
 Address = #"Tipo cambiado2"{0}[Address],
 #"Tipo cambiado3" = Table.TransformColumnTypes(Address, {"CountryRegion", type
text})
in
 #"Tipo cambiado3"
in
 Country

```

## 12. Visualización de los datos

La visualización del tablero se realizó en Power BI.

A continuación, se detalla un breve resumen de cada una de las páginas creadas:

### - Página de inicio.

En la misma se da una breve descripción de lo que es el TDF.

Se agregan botones hacia las demás hojas de la presentación.

Se agrega la fecha de la última actualización del archivo de Power BI.





Se agrega un marcador para indicar porque hay datos que no fueron tomados en cuenta para el análisis.

INFORMACIÓN IMPORTANTE A TENER EN CUENTA. PRESIONE AQUÍ.

**El TDF es una carrera ciclista anual de etapas múltiples que se lleva a cabo principalmente en Francia. Aunque ocasionalmente pasa por países cercanos.**

**Es de las competiciones deportivas más vistas del mundo. Emitida en 180 países y vista por 3.500 millones de personas cada año.**

**El Tour se creó en 1903 y apenas se detuvo durante las 2 Guerras Mundiales.**

**Para el presente trabajo se tiene información desde 1903 a 2016 inclusive.**

INFO GENERAL

CARRERA

PARTICIPANTES

Fecha de última actualización  
11/9/2022 21:38:18

El 22 de octubre de 2012, Lance Armstrong es despojado formalmente de los siete títulos del Tour de Francia que ganó entre 1999 y 2005.

Se le prohibió de por vida competir en ciclismo después de comprobarse que había usado drogas ilegales.

Para este trabajo no se tuvo en cuenta los datos de esos años.

#### - Página Info General.

En esa página se trata de explicar aún más lo que significa el TDF. Además de dar información sobre la competición.

Se agregan los botones necesarios para volver al Inicio o pasar a las demás Hojas.

Ciudades

49

Cantidad total etapas

1928

Cantidad total de ciclista que comenzaron

13284

Cantidad total de ciclistas que terminaron

8031

Ciclismo y Tour de Francia van de la mano; no se entiende el uno sin el otro.

La inauguración de la vuelta al país galo fue el 1 de julio de 1903, concretamente en Montgeron, muy próximo a París. Tan solo 60 ciclistas salieron de la línea de meta en aquel primer año.

Fue la primera carrera por etapas en un país. Tiene un recorrido ideal para la competición: consta de montañas y llanuras.

CARRERA

PARTICIPANTES

INICIO

Nacionalidades de los participantes

13

Distancia total recorrida (Km)

411891

Velocidad promedio (Km/h)

34

Cantidad de equipos participantes

42

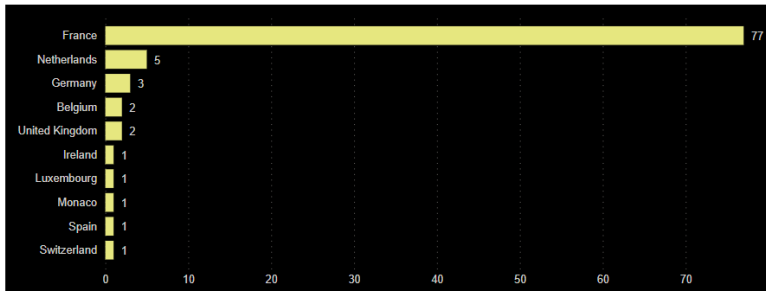
## - Página Carrera.

Se agrega un slicer para que el usuario pueda seleccionar los años que le interesa.

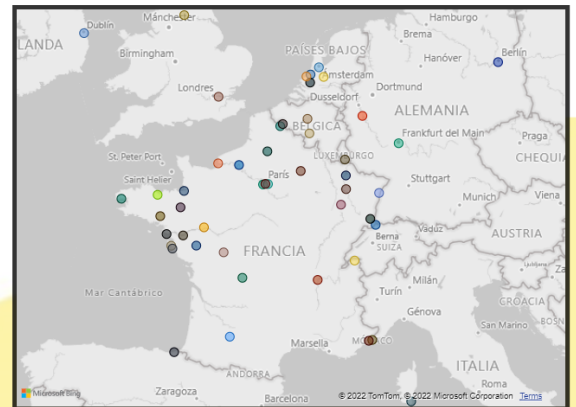
En esta Hoja se abordan los datos generales de la carrera, sus competidores y los países por los que se desarrolló.

Se agrega información sobre los temas antes indicados.

Países en donde comenzó el Tour.

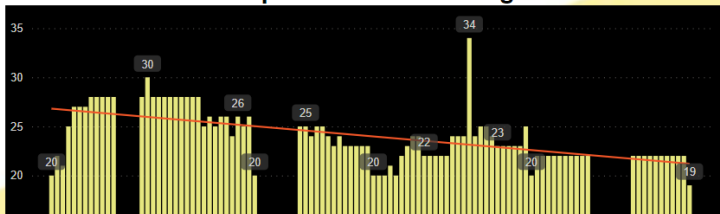


El primer país donde inició el TDF, fuera de Francia, fue Holanda en 1954.



Con el pasar de los años, la duración del TDF ha sido mas corta.

Cantidad de días que duro TDF a lo largo de los años.



Promedio de duración (Días) **24** TDF mas corto (Días) **19** TDF mas largo (Días) **34**

1903 1929 se corrió de manera individual

## - Página Participantes.

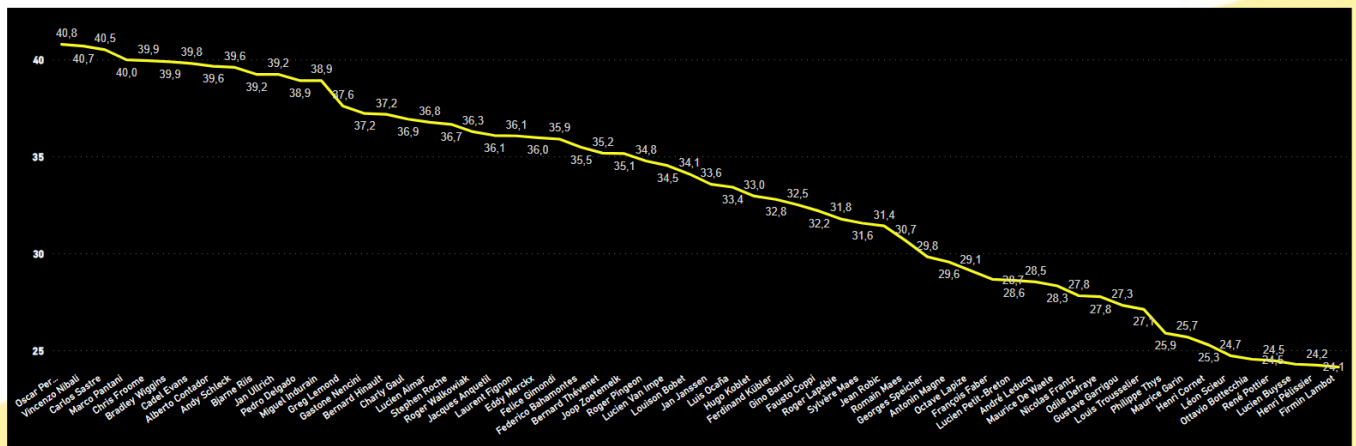
Es la última hoja de la presentación.

En la misma se aborda todo lo relacionado sobre los corredores.

El español Oscar Pereiro cuenta con la mayor velocidad promedio en un TDF. Mientras que el belga Firmin Lambot cuenta con las mas lenta.



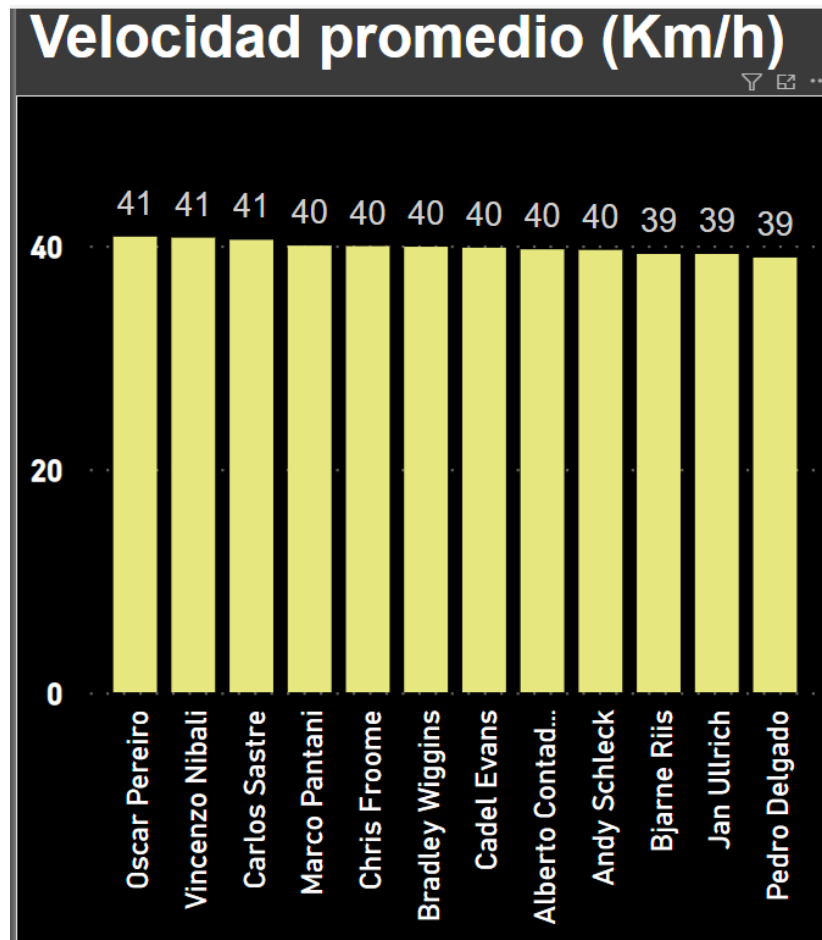
Velocidad promedio (Km/h) más altas y más bajas de ganadores del TDF.



Corredores que iniciaron / abandonaron el TDF y porcentaje de abandonos.

Se agregan los siguientes tooltip:

- **VelProm**



- **Ciudades**

| Año  | Ciudad Largada  | Ciudad Llegada | Duración (Días) |
|------|-----------------|----------------|-----------------|
| 2016 | Mont St. Michel | Paris          | 19              |

- **NacionalidadGanadores**

| Ganador          | Nacionalidad | Victorias |
|------------------|--------------|-----------|
| Bernard Hinault  | France       | 5         |
| Eddy Merckx      | Belgium      | 5         |
| Jacques Anquetil | France       | 5         |

- EquiposAños

| Team           | Year |
|----------------|------|
| La Française   | 1903 |
| Cycles JC      | 1904 |
| Cycles Peugeot | 1905 |

- CiudadesLargada

| City  | Year |
|-------|------|
| Paris | 1903 |
| Paris | 1904 |
| Paris | 1905 |
| Paris | 1906 |

Los tooltips se muestran cuando se pasan por encima de los gráficos.

### **13.Futuras actualizaciones.**

Se considera oportuno actualizar la BDD incluyendo información desde 2017 al 2022 para continuar con el análisis.

Además, se podría conseguir más información sobre las etapas en particular de cada edición para profundizar en el análisis.