



PROGRAMACION PARA LA EXTRACCION DE DATOS

OMAR FLORES BECERRIL

PROYECTO FINAL

JOSUE MIGUEL FLORES PARRA

Contenido Proyecto

PLANTEAMIENTO DEL PROBLEMA	3
LIGAS A USAR.....	3
OBJETIVOS	3
DASHBOARD 1	4
DASHBOARD 2.....	4
DIAGRAMA-ENTIDAD RELACIONAL.....	5
DICCIONARIOS DE DATOS	5
DASHBOARDS	6
DASHBOARD GENERAL:	6
DASHBOARD POR POSICIÓN:	8
REFERENCIAS	9

PLANTEAMIENTO DEL PROBLEMA

El objetivo principal de este proyecto es analizar el rendimiento de los jugadores de fútbol a partir de datos específicos como goles, asistencias, minutos jugados, distancia cubierta y velocidad máxima. Los datos se obtienen de un conjunto preexistente y se presentan mediante dashboards interactivos, este proyecto se base en contestar preguntas estadísticas hacerca de los jugadores de la champions league por lo cual puede ser actualizado para futuras temporadas o para bien usarlo para hacer un analisis con temporadas pasadas, decidi hacerlo de este tema ya que es algo que a mi me gusta mucho y que me dio la idea de que me seria de gran utilidad para analizar mas a detalle el deporte que me encanta.

LIGAS A USAR

1. Las ligas a usar solo use una para poder de esa manera hacer una mejor extraccion de datos y esta liga me permitio obtener los datos que deseo para poder generar la estadisitica necesaria, el tamaño de los datos que se extraen se puede modificar para poder extraer mas o menos según sea la preferencia.

<https://www.uefa.com/uefachampionsleague/statistics/players/>

OBJETIVOS

Extracción de Datos: Obtener datos precisos y detallados de jugadores de fútbol, mediante la utilizacion de codigo en python apoyandonome de librerias como selenium, beautifulsoup4, pandas y los webdrivers, estas librerias me permitieron realizar la extraccion de manera mas eficaz y simple, para su utilizacion me apoye en el codigo que ya nos habia enseñado el profesor en clase.

Limpieza y Comprobaciones: Asegurar la calidad de los datos mediante la eliminación de duplicados y la gestión de valores faltantes, esta limpieza de datos fue primero en observacion de los datos, ya que mi extraccion de datos fue precisa y sin errores la limpieza es mas facil ya que no tenia tantos datos anormales, de igual manera comprabe duplicados y datos anormales para eliminarlos o en si oara separarlos como fue el caso de equipo y posicion que estaban unidas.

Creación de Estadísticas para Dashboards: Generar estadísticas relevantes que puedan ser visualizadas de manera interactiv.

Creación de Base de Datos MySQL: Organizar los datos en una base de datos estructurada en python mediante la conexión entre el gestor y la base dedatos, realice la creacion de esta base de datos en mi gestor.

Mandar los Datos a las Tablas Correspondientes en MySQL Workbench: Importar los datos limpios a MySQL mediante el uso de código python.

Creación de Dashboards con Python: Desarrollar dashboards interactivos utilizando Dash y Plotly, apoyandome de el código python que ya conozco y que me fue de gran utilidad.

Mandar los Datos Estadísticos a las Gráficas de los Dashboards: Visualizar los datos en gráficos interactivos, se mandaron los datos ya limpios a los dashboards para poder visualizar las estadísticas.

Dashboard 1

Goles por Jugador Grafica 1

- ¿Cuántos goles ha anotado cada jugador?
- ¿Quiénes son los máximos goleadores?
- Tabla: Nombre, equipo, posicion, goles

Asistencias por Jugador Grafica 2

- ¿Cuántas asistencias ha hecho cada jugador?
- ¿Quiénes son los máximos asistentes?
- Tabla: Nombre, equipo, posicion, asistencias

Minutos Jugados vs. Distancia Cubierta Grafica 3

- ¿Cuál es la relación entre los minutos jugados y la distancia cubierta por cada jugador?
- ¿Qué jugadores han jugado más minutos y cubierto más distancia?
- Tabla: Nombre, equipo, posicion, minutos_jugados, distancia_cubierta

Dashboard 2

Distribución de Goles por Posición Grafica 1

- ¿Cuál es la distribución de goles entre las diferentes posiciones?
- Tabla: Nombre, equipo, posicion, goles

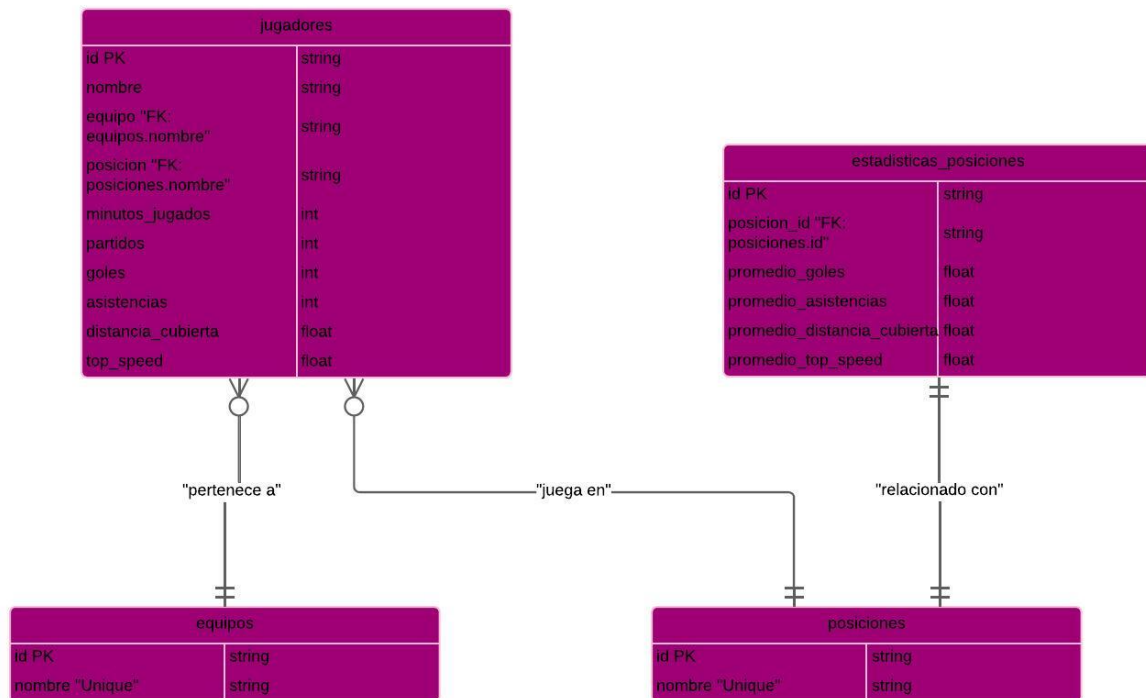
Distribución de Asistencias por Posición Grafica 2

- ¿Cuál es la distribución de asistencias entre las diferentes posiciones?
- ¿Existen posiciones que tienden a tener más asistencias?
- Tabla: Nombre, equipo, posicion, asistencias

Promedio de Distancia Cubierta por Posición Grafica 3

- ¿Cuál es el promedio de distancia cubierta por cada posición?
- ¿Qué posiciones tienden a cubrir más distancia en el campo?
- Tabla: Nombre, equipo, posicion, distancia_cubierta

DIAGRAMA-ENTIDAD RELACIONAL



DICCIONARIOS DE DATOS

Tabla jugadores:

id: Identificador único del jugador.

nombre: Nombre del jugador.

equipo: Equipo en el que juega.

posicion: Posición en el campo.

minutos_jugados: Minutos jugados por el jugador.

partidos: Número de partidos jugados.

goles: Goles anotados.

asistencias: Asistencias realizadas.

distancia_cubierta: Distancia cubierta en kilómetros.

top_speed: Velocidad máxima alcanzada.

Tabla equipos:

id: Identificador único del equipo.

nombre: Nombre del equipo.

Tabla posiciones:

id: Identificador único de la posición.

nombre: Nombre de la posición.

Tabla estadisticas_posiciones:

id: Identificador único de la estadística.

posicion_id: Identificador de la posición.

promedio_goles: Promedio de goles por posición.

promedio_asistencias: Promedio de asistencias por posición.

promedio_distancia_cubierta: Promedio de distancia cubierta por posición.

promedio_top_speed: Promedio de velocidad máxima por posición.

DASHBOARDS

Dashboard General:

Goles por Jugador Grafica 1

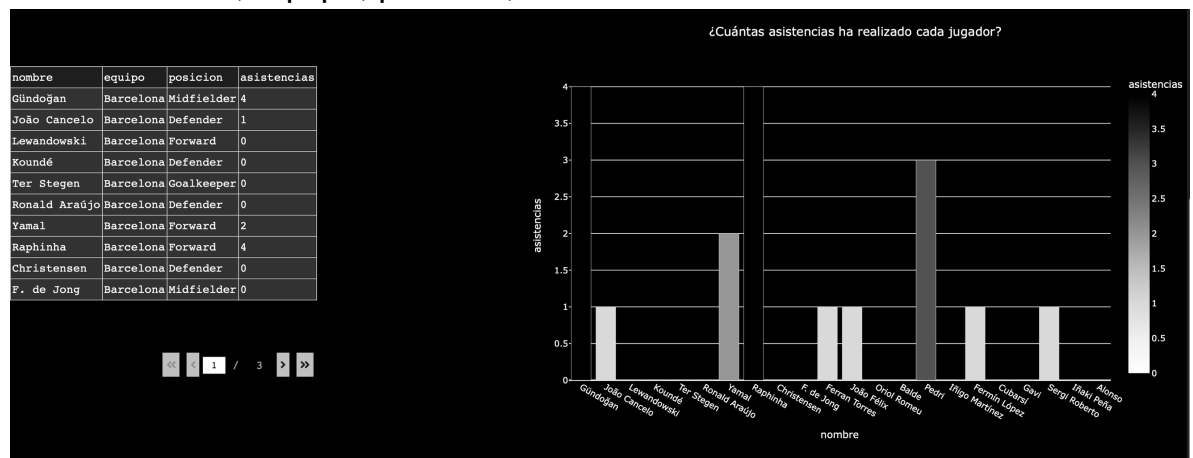
- ¿Cuántos goles ha anotado cada jugador?
- ¿Quiénes son los máximos goleadores?

- Tabla: Nombre, equipo, posicion, goles



Asistencias por Jugador Grafica 2

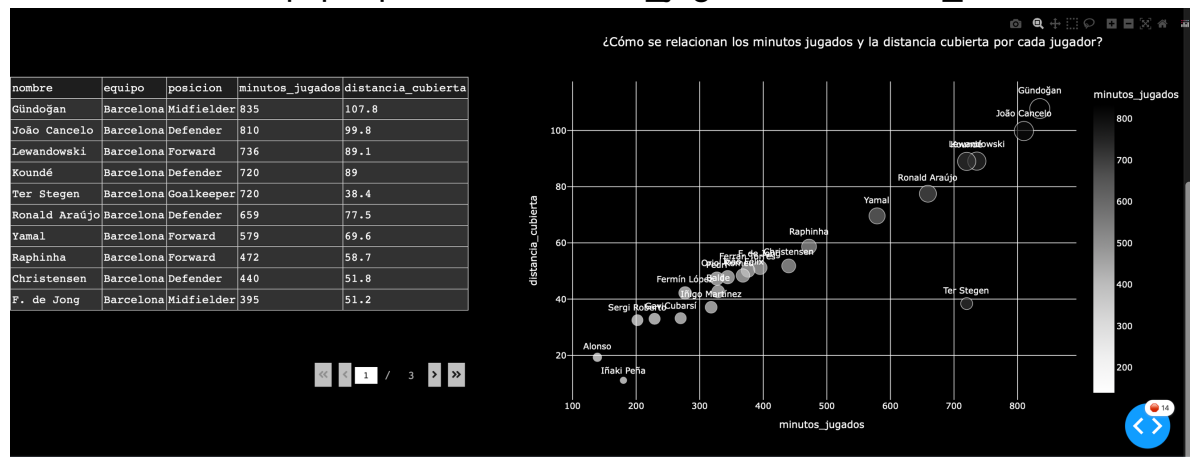
- ¿Cuántas asistencias ha hecho cada jugador?
- ¿Quiénes son los máximos asistentes?
- Tabla: Nombre, equipo, posicion, asistencias



Minutos Jugados vs. Distancia Cubierta Grafica 3

- ¿Cuál es la relación entre los minutos jugados y la distancia cubierta por cada jugador?
- ¿Qué jugadores han jugado más minutos y cubierto más distancia?

- Tabla: Nombre, equipo, posicion, minutos_jugados, distancia_cubierta



Dashboard por Posición:

Distribución de Goles por Posición Grafica 1

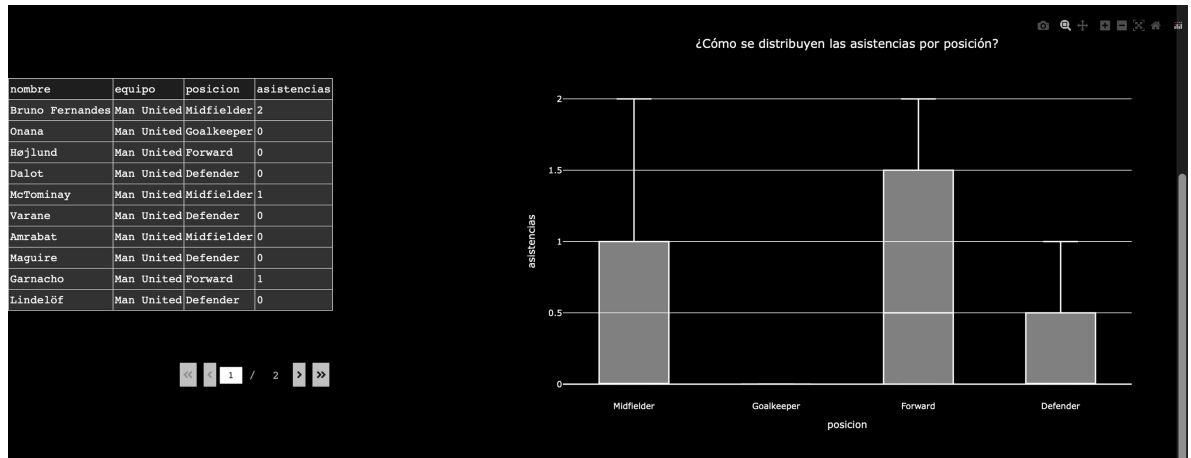
- ¿Cuál es la distribución de goles entre las diferentes posiciones?
- Tabla: Nombre, equipo, posicion, goles



Distribución de Asistencias por Posición Grafica 2

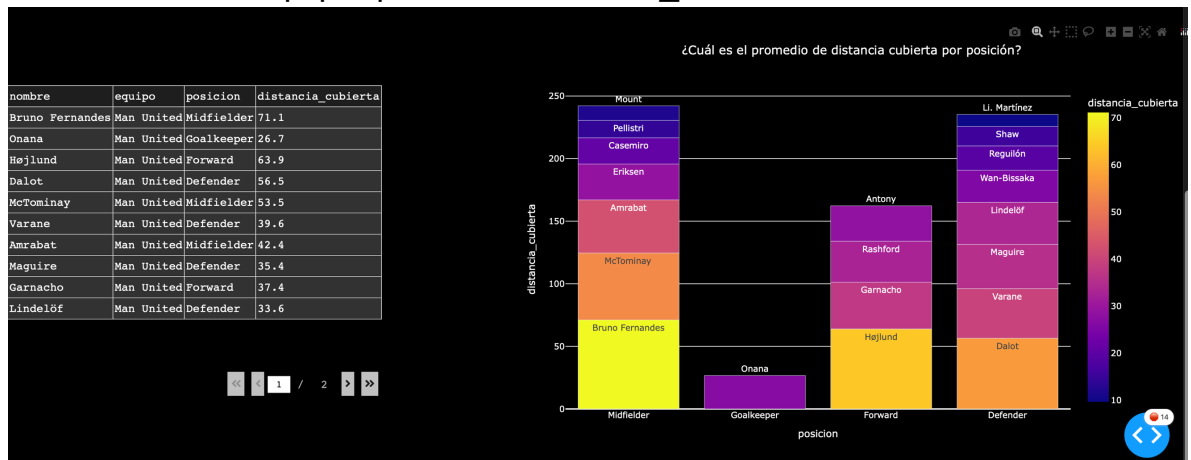
- ¿Cuál es la distribución de asistencias entre las diferentes posiciones?
- ¿Existen posiciones que tienden a tener más asistencias?

- Tabla: Nombre, equipo, posicion, asistencias



Promedio de Distancia Cubierta por Posición Grafica 3

- ¿Cuál es el promedio de distancia cubierta por cada posición?
- ¿Qué posiciones tienden a cubrir más distancia en el campo?
- Tabla: Nombre, equipo, posicion, distancia_cubierta



REFERENCIAS

GitHub: <https://github.com/OmarFloresQ/ProyectoExtraccion.git>

Ligas: <https://www.uefa.com/uefachampionsleague/statistics/players/>

Se utilizo la documentacion de esta liga <https://dash.plotly.com/interactive-graphing> para poder realizar las graficas con un mejor diseño.

<https://dash.plotly.com/workspaces/visualizing-plotly-graphs-in-jupyter-and-workspaces>

<https://dash.plotly.com/dash-core-components/graph>

<https://plotly.com/python/>

Tambien algunos videos que me ayudaron a realizar el codigo para la extraccion de datos:

Curso de Web Scraping en Python | Web Scraping MUY FACIL con BeautifulSoup [Nivel Básico]

<https://youtu.be/kPNHKrOgedI?si=08XUO1cz7yLA2v5a>

Python: Web-Scraping para principiantes usando XPATH y BeautifulSoup dos ejemplos

<https://youtu.be/SuwCyiCLe8Y?si=LF1cllIDZFfFBcAO>