**MEMORIA ANALISIS EXPLORATORIO DE DATOS**

**OBJETIVO DEL TRABAJO**

El objetivo de este proyecto es aprender a realizar un EDA (Análisis Exploratorio de Datos).

Para desarrollarlo he empleado el IDE Visual Studio Code, usando el lenguaje de programación de Python, con el que he creado un notebook en el cuál está escrito todo el código que se ha ido empleando para dar estructura a los datos y crear las gráficas.

**CÓDIGO Y TRANSFORMACIÓN DE LOS DATOS**

En primer lugar, me he descargado la base de datos en la web de Kaggle, desde el siguiente enlace: <https://www.kaggle.com/drgilermo/nba-players-stats>

En el archivo descargado, he usado las siguientes bases de datos:

* Players.csv
* Seasons\_Stats.csv

Mediante estos 2 archivos unidos, se ha creado el dataframe con el que se trabaja durante todo el EDA.

Para realizar el código se ha hecho uso de las siguientes librerías, de las cuales se ha consultado también la documentación oficial cuando ha sido necesario:

* Pandas: [API reference — pandas 1.2.3 documentation (pydata.org)](https://pandas.pydata.org/docs/reference/index.html)
* Numpy: [NumPy Reference — NumPy v1.20 Manual](https://numpy.org/doc/1.20/reference/index.html)
* Seaborn: [API reference — seaborn 0.11.1 documentation (pydata.org)](https://seaborn.pydata.org/api.html)
* Matplotlib: [Gallery — Matplotlib 3.3.4 documentation](https://matplotlib.org/stable/gallery/index.html)

Como referencia para resolver dudas, además de la documentación de estas librerías, también se han consultado los notebooks vistos en clase.

Una vez cargadas y juntadas las bases de datos, se ha procedido a limpiar y dar forma a los datos. Para ello, se ha usado la librería de Pandas con la que se han ido eliminando columnas que no se fueran a utilizar a lo largo del proyecto, se han creado otras nuevas, se ha cambiado el formato de algunas columnas, comprobado si había valores Nan, se han ordenado las columnas, etc.

A partir de este dataframe, se ha filtrado para que se cogiese a partir del año 1980, año donde se introdujo el triple en la NBA. También se ha comprobado las diferentes posiciones de jugadores existentes y se han agrupado todas en 3 posiciones principales que son las de base, alero y pívot, posiciones sobre las que nos basaremos luego para representar los datos en las gráficas.

Tras modelar todos los datos y ya disponibles para trabajar con ellos de la forma que quería, he comprado los estadísticos de todos los datos (media, desviación típica, máximos y mínimos, etc.), las correlaciones entre todas las variables numéricas y los sumatorios totales de cada variable en base a la década y las posiciones, así como otro por años solo.

**HIPÓTESIS, REPRESENTACIÓN GRÁFICA Y CONCLUSIONES**

Después de ver todas las estadísticas, he pasado a la representación gráfica buscando responder a las hipótesis planteadas a la hora de comenzar el trabajo.

En la primera grafica se ha buscado ver si había una evolución física de los jugadores a lo largo de los años mediante las variables de peso y altura, sin embargo, al no apreciarse casi diferencias, no me servía para comprobar si ha habido una mejora o no de la forma física de los jugadores, por lo que he decidido no mostrarlo al final puesto que se deberían tener seguramente más datos antropomorfológicos y de test físicos de los jugadores para ver dicha evolución y extraer conclusiones precipitadas.

Por otro lado, el resto de las gráficas sí que se han empleado para comprobar la evolución que ha tenido el juego de la NBA a lo largo de los años desde la introducción del triple. El cambio del juego se ha visto afectado debido a que el número de tiros ha ido incrementando poco a poco con el paso de los años, notándose de forma significativa en los lanzamientos de triples y un ligero descenso de los tiros de 2 puntos. Además, el aumento total de tiros de campo ha hecho que el juego sea más rápido y que se produzcan más acciones durante los partidos, viendo esta evolución en el incremento, generalmente, del resto de estadísticas.

Atendiendo a los tiros de campo por las 3 posiciones principales de los jugadores, se puede observar que los tiros de 2 puntos tanto de los bases como de los pívots se mantienen o crecen ligeramente, pero los aleros han reducido considerablemente este tipo de tiros. Mientras que los triples se han visto incrementados en todas las posiciones, sobre todo en los aleros, que han evolucionado su juego claramente hacia el juego exterior. También es destacable que los pívots, los jugadores más altos y menos habilidosos del equipo, han cambiado esa tendencia y han ampliado su rango de tiro, pasando de no lanzar casi ningún triple hasta lanzar casi los mismos que los bases, los jugadores más bajos y encargados de dirigir al equipo en la pista.

La evolución del juego, que ha adoptado una tendencia más ofensiva y centrándose menos en la parcela defensiva, no ha tenido casi impacto en el número de robos y de pérdidas por partidos, cuyas estadísticas se mantienen más o menos similares, no se notan casi diferencias entre diferentes épocas. Esto difiere con mi hipótesis inicial, ya que pensaba que dichas estadísticas habrían aumentado al centrarse más en atacar y despreocupándose un poco más por la defensa.

**DIFICULTADES DURANTE EL PROYECTO**

A la hora de realizar el proyecto, he encontrado dificultades para crear las gráficas y es de lo que más tiempo me ha llevado durante todo el trabajo. Al intentar crear otro tipo de gráficas a las realizadas finalmente, por dar variedad al trabajo con gráficas de distintos tipos, no sabía cómo hacerlas, por lo que todas las gráficas son de barras y de líneas.

Al comienzo de comenzar a graficar, también me resultó complicado en pensar qué es lo que quería representar y cómo exactamente, por lo que tuve que realizar pruebas e incluso volver al código para eliminar columnas y moldear algunos datos más.