

Memoria EDA Tenis

EDA

Primero, realizar la búsqueda de datos en internet. Tras barajar diferentes APIs y páginas web que ofrecían servicios relacionados con el mundo del tenis, me decanté por utilizar Ultimate Tennis Statistics y la página oficial ATP tour.

Segundo, con los datos ya descargados en formato csv procedemos a su limpieza y filtrado para obtener aquellos valores que nos interesan. Una vez que los datos están limpios se juntan en un mismo dataframe.

Tercero, con los dataframes correspondientes ya montados realizamos representaciones gráficas de los mismos para entender mejor los datos.

Cuarto, los datos que necesitamos, que ya están limpios, ordenados, y representativos, los utilizamos para realizar una presentación que será explicada a la audiencia.

Organización del EDA

¿Por qué? A raíz de que Djokovic ganara Wimbledon y sumase su vigésimo grand slam se reanimó el debate de quien era el mejor jugador de todos los tiempos. Este se debe a que el Big 3 (Federer, Nadal, Djokovic) tengan el mismo número de grand slams.

Por lo tanto, mi idea principal fue debatir de una manera objetiva (basándome en las estadísticas de cada jugador durante sus respectivas carreras) quién realmente el mejor jugador de la historia actualmente, 25 de Julio de 2021.

En este punto empezamos a analizar la parte objetiva del EDA.

- Se analizan estadísticas durante su carrera (servicio, saque, duración de los partidos, victorias/derrotas como favorito/no favorito)

- Se analiza el rendimiento que han tenido durante sus carreras (superficies, dureza mental)

- Se analizan líneas temporales (grand slams y dominancia en el circuito)

- Se analizan sus respectivos enfrentamientos (head to head)

- Se enseña los récords personales de cada uno

Analizamos quien es el mejor de los tres por superficie de una manera más detallada.

Segundo lugar, análisis de la parte subjetiva del EDA.

Se analizan una serie de encuestas acerca de quién es el mejor jugador de la historia con sus respectivos resultados.

Para finalizar, se ofrece una conclusión que resume un poco lo analizado anteriormente.

Librerías utilizadas en el EDA

Para la lectura de datos y la limpieza de estos se ha utilizado Numpy y Pandas. Con estas librerías se han filtrado los datos hasta llegar a lo que se necesitaba.

Para la parte de visualización del proyecto, se ha utilizado la librería plotly para la realización de las gráficas.

Por último se ha utilizado la librería Streamlit, su funcionalidad es presentar el EDA en formato de página web.

Sources empleadas para la búsqueda de datos

Los datos utilizados en el EDA han sido recolectados de:

Ultimate Tennis Statistics: página web muy completa en la que se puede encontrar todo tipo de dato estadístico relacionado con el tenis.

Link: <https://www.ultimatetennisstatistics.com/>

ATP tour: página web oficial de la ATP que ofrece información actualizada.

Link: <https://www.atptour.com/es/>

Las encuestas utilizadas en el EDA se han obtenido de:

The Delite: <https://www.thedelite.com/the-greatest-mens-tennis-players-of-all-time/25/>

Give me sport: <https://www.givemesport.com/1480600-roger-federer-wins-public-vote-for-greatest-tennis-player-of-all-time>

SportBible: <https://www.sportbible.com/tennis/reactions-news-legends-roger-federer-voted-the-greatest-tennis-player-of-all-time-20190714>

Ceoworld: <https://ceoworld.biz/2020/06/15/revealed-worlds-greatest-tennis-players-of-all-time/>

How they play: <https://howtheyplay.com/individual-sports/Top-10-Greatest-Male-Tennis-Players-of-All-Time>