

Informe del trabajo final. Análisis Exploratorio de Datos y Visualización.

La selección del tema de nuestro trabajo final fue el mundo Pokémon, que está compuesto de un conjunto de videojuegos en los que se entrenan (suben de nivel) y combaten entre ellas criaturas ficticias que poseen distintas características: un identificador propio, nombre, un tipo o dos (planta, agua...), nivel de vida, de ataque, ataque especial, defensa, defensa especial, velocidad, y la suma de todas ellas. Además, si son legendarios o no, y cada uno creado en una determinada generación (cuanto menor generación antes se creó). El análisis se ha realizado con las 6 primeras generaciones de pokémon.

El objetivo principal es determinar qué características de los pokémon son más relevantes a la hora de ganar un combate 1 contra 1, sin olvidarnos de otros puntos interesantes acerca de las 800 muestras que analizaremos. Este análisis puede ayudar tanto a principiantes como a gente avanzada para saber qué cosas se ha de tener en cuenta a la hora de capturar un pokémon que le acompañe a lo largo de su aventura. Hemos decidido hacer el análisis sobre este mundo porque es un juego muy popular y porque nos parece un tema interesante.

Por lo tanto, las cuestiones que resolveremos serán:

- Averiguar cuáles son los mejores y peores pokémon
- Conocer los tipos más comunes en cómputo global y por generaciones
- Conocer qué relevancia tiene cada una de las características con el win rate (tasa de victorias) de los pokémon y posibles correlaciones entre ellas
- Conocer la relación de las estadísticas respecto a los tipos de los pokémon
- Estudiar si la generación influye en la proporción de tipos y en el win rate de los pokémon
- Establecer las diferencias entre pokémon no legendarios y legendarios observando sus características y tipos

Los datos con los que se realiza el análisis y visualización los hemos obtenido de:

<https://www.kaggle.com/datasets/terminus7/pokemon-challenge?select=combats.csv>
<https://www.kaggle.com/datasets/tuannguyenvanh/pokemon-dataset-with-team-combat>

Utilizaremos los datasets: 'combats.csv' y 'pokemon.csv'. En este último se representan 800 filas, cada una es un pokémon y 12 columnas en las que se encuentran las características mencionadas con anterioridad: tipo 1 (p.e fuego), tipo 2 (p.e volador), suma de todas las estadísticas, hp (vida), ataque, defensa, ataque especial, defensa especial, velocidad, nº de generación y si es legendario o no. Esta información está sacada directamente de los datos del videojuego. En 'combats.csv' se presentan 50000 filas y tres columnas que representan combates 1vs1 entre los diferentes pokémon. La primera columna es el id del pokémon que empieza atacando, la segunda es el id del otro pokémon y la tercera el id del ganador. El ganador se ha generado por un algoritmo personalizado que simula combates 1 contra 1 entre los 800 pokémon, y tras la simulación se obtiene el id que representa al pokémon con más posibilidades de ganar.

Respecto a la preparación de los datos, primero hay que ejecutar el código de python que es en el que hemos realizado ingeniería de atributos, creando una nueva columna llamada Win.Percentage que es en la que calculamos la probabilidad de ganar un combate de cada uno de los pokemon (lo que los hace mejor que los demás con menor porcentaje). Esto ha sido realizado con el dataframe 'combats.csv', hallando la proporción de victorias frente al número de combates totales de cada pokémon. Además, también hemos creado la columna "Total" que es la suma de las estadísticas numéricas ya mencionadas con anterioridad. Estas dos nuevas variables se han añadido al dataframe 'pokemon.csv' originando ahora un nuevo dataframe con las modificaciones llamado: 'pokemon_data.csv'.

Para generar las gráficas hemos utilizado íntegramente R y librerías como ggplot2, gganimate, tidyverse, dbplyr, etc. Los tipos de gráficas utilizados han sido diagramas de barras (barplots), siendo animados algunos de estos (formato gif), diagramas de puntos (scatterplots) con líneas de regresión, mapas de calor (heatmaps), matriz de correlación y diagramas de tartas (pie charts).

La historia visual la hemos realizado en Genially ya que nos ha permitido crear una presentación con diapositivas de manera sencilla, bonita, y con un resultado muy visual y fácil de seguir para el espectador, ya que además hemos optado por un formato en el que no haya mucho texto en la presentación y las imágenes y gráficas sean protagonistas.

El enlace a la presentación de Genially se encuentra en el documento 'historia visual' donde, además, debajo de este se encuentra otro enlace que conduce a un documento de drive donde se encuentran las interpretaciones de cada una de las gráficas que se comunicarán de manera oral en la presentación.

Los resultados y conclusiones que hemos obtenido se podrían resumir en:

- Los mejores pokémon son legendarios y mega-evoluciones. Y los peores las primeras etapas evolutivas.
- El tipo base más común es el agua, y como segundo tipo el volador.
- Los pokémon más rápidos son los que más win rate tienen, por lo que la velocidad es muy importante a la hora de los combates.
- Los pokémon de tipo acero son los que más defensa tienen seguido del roca, el volador el más rápido y el dragón el de más ataque. Los pokémon con más win rate son el volador (probablemente debido a su gran velocidad), dragón (por su ataque) y siniestro (buen ataque como segundo tipo). Es decir, se reafirma que lo más importante es la velocidad con diferencia y seguidamente el ataque.
- La segunda generación es la que menos winrate tiene porque se incorporaron muchas etapas preevolutivas.
- Los pokémon legendarios tienen mejores estadísticas que los no legendarios, y por tanto mucho mayor win rate, además sus tipos más comunes coinciden con los mejores tipos (volador, dragón...)

Los resultados obtenidos son bastante claros y reveladores, ya que desde un primer momento quedó claro que los pokémon más rápidos son los que tienen mayor posibilidad de ganar un combate 1vs1 (aunque sin olvidarnos tampoco de sus capacidades ofensivas como el ataque). Básicamente, quien golpea primero tiene mayor posibilidad de ganar. El tipo volador es el que más winrate tiene entre todos los tipos ya que estos pokémon suelen

ser generalmente rápidos, como se puede intuir simplemente escuchando el nombre y como hemos mostrado gráficamente mediante dos mapas de calor.

En definitiva, no hemos tenido grandes complicaciones a la hora de generar o interpretar los gráficos ya que los resultados son bastante esclarecedores y consideramos que las gráficas son de buena calidad y bastante ilustrativas, ya que, por ejemplo, en las gráficas acerca de los tipos de pokemon hemos aplicado colores relacionados a esos tipos (por ejemplo, naranja con el tipo fuego o amarillo con el tipo eléctrico).