

Obtención de estadísticas descriptivas

Jorge Guerrero Díaz A01411752
José Sebastián Pedrero Jiménez A01703331
Armando Gutiérrez Rojo A01702748

La base de datos elegida cuenta con 13 columnas que especifican diferentes atributos de cada Pokemon. Como son 800 Pokemon, la base de datos cuenta con 800 renglones.

Las columnas son: #, Name, Type1, Type2, Total, HP, Attack, Defense, Sp. Atk, Sp. Def, Speed, Generation, Legendary.

	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	False
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	False
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	False
3	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	False
4	4	Charmander	Fire	NaN	309	39	52	43	60	50	65	1	False
...
795	719	Diancie	Rock	Fairy	600	50	100	150	100	150	50	6	True
796	719	DiancieMega Diancie	Rock	Fairy	700	50	160	110	160	110	110	6	True
797	720	HoopaHoopa Confined	Psychic	Ghost	600	80	110	60	150	130	70	6	True
798	720	HoopaHoopa Unbound	Psychic	Dark	680	80	160	60	170	130	80	6	True
799	721	Volcanion	Fire	Water	600	80	110	120	130	90	70	6	True

Las columnas Name, Type1, Type2 y Legendary cuentan con datos categóricos nominales, ya que son utilizados para etiquetar variables y no tienen ningún valor cuantitativo y su orden no importa. Por otro lado, las columnas HP, Attack, Defense, Sp. Atk, Sp. Def y Speed cuentan con datos numéricos discretos, ya que se pueden calcular proporciones, media, mediana y moda. Finalmente las columnas # y Generation cuentan con datos numéricos continuos, ya que no hay cero absoluto ni relación de proporción.

De acuerdo al análisis realizado los rangos en los que se encuentra cada variable son:

- Total de atributos: 180 - 780
- HP: 1 - 255
- Attack: 5 - 190
- Defense: 5 - 230
- Sp. Atk: 10 - 194
- Sp. Def: 20 - 230
- Speed: 5 - 180

Esto nos puede servir para saber si un Pokémon se encuentra en los puntos altos de cada variable, en los bajos o en los puntos medios.

También se calculó la media, mediana y desviación estándar estándar de cada variable.

Media

Media de Total es: 435.1025

Media de HP es: 69.25875
Media de Attack es: 79.00125
Media de Defense es: 73.8425
Media de Sp. Atk es: 72.82
Media de Sp. Def es: 71.9025
Media de Speed es: 68.2775

La media de las variables nos puede servir para saber si un Pokémon se encuentra arriba del promedio o por debajo de este en cada variable, lo que puede servir para conocer qué Pokémon es mejor en cada rubro. Por los datos obtenidos, sabemos que el promedio de atributos que tiene un Pokémon es de 435.1025, la vida promedio es de 69.25875, el ataque promedio es de 79.00125, el super ataque promedio es de 72.82, la super defensa promedio es de 71.9025 y la velocidad promedio es de 68.2775.

Mediana

Mediana de Total es: 450.0
Mediana de HP es: 65.0
Mediana de Attack es: 75.0
Mediana de Defense es: 70.0
Mediana de Sp. Atk es: 65.0
Mediana de Sp. Def es: 70.0
Mediana de Speed es: 65.0

La mediana de cada variable nos indica cuál es el dato que se encuentra justo a la mitad, por lo tanto si un Pokémon tiene más puntos que la mediana, significa que se encuentra en la mitad superior. De igual forma, si tiene menos puntos que la mediana, se encuentra en la mitad inferior.

Desviación estándar

La desviación estándar de Total es: 119.963039755519 y su varianza es de: 14391.130907384233
La desviación estándar de HP es: 25.534669032332047 y su varianza es de: 652.0193225907373
La desviación estándar de Attack es: 32.45736586949843 y su varianza es de: 1053.4805991864816
La desviación estándar de Defense es: 31.183500559332927 y su varianza es de: 972.410707133917
La desviación estándar de Sp. Atk es: 32.72229416880157 y su varianza es de: 1070.748535669585
La desviación estándar de Sp. Def es: 27.82891579711745 y su varianza es de: 774.4485544430531
La desviación estándar de Speed es: 29.06047371716149 y su varianza es de: 844.5111326658338

La desviación estándar es una medida que se utiliza para cuantificar la dispersión de un conjunto de datos numéricos. Si la desviación estándar es baja significa que la mayor parte de los datos de una muestra tienden a estar agrupados cerca de su media. En los datos

obtenidos se puede observar que la variable que tiene más dispersión es la de total de atributos y que la que tiene menor dispersión es la de HP.

Después de analizar la base de datos utilizando Python y Pandas, nos dimos cuenta de cómo podemos utilizar herramientas computacionales para simplificar el análisis de datos y la importancia de tener datos ordenados y fáciles de leer.

Link del repositorio Github:

https://github.com/MandukoGR/TC1002S_Equipo2