

1. Describe the dataset.

This dataset contains information on a total of 802 Pokémon from all seven generations. The data was scraped from the Serebii site and includes basic statistics of Pokémon, their performance against other species, height and weight, classification, egg steps, experience points, abilities, etc. The dataset is available on Kaggle.

**Characteristics of the Data:**

The dataset contains a total of 41 columns. These columns and their contents are explained in detail below:

abilities: List of abilities that Pokémon can have (Text).

against\_bug: Pokémon's damage rate against bug-type attacks (Decimal).

against\_dark: Pokémon's damage rating against dark-type attacks (Decimal).

against\_dragon: Pokémon's damage rating against dragon-type attacks (Decimal).

against\_electric: Pokémon's damage rating against electric-type attacks (Decimal).

against\_fairy: Pokémon's damage rating against fairy-type attacks (Decimal).

against\_fight: Pokémon's damage rating against fight-type attacks (Decimal).

against\_fire: Pokémon's damage rating against fire-type attacks (Decimal).

against\_flying: Pokémon's damage rating against flying-type attacks (Decimal).

against\_ghost: Pokémon's damage rating against ghost-type attacks (Decimal).

against\_grass: Pokémon's damage rating against grass-type attacks (Decimal).

against\_ground: Pokémon's damage rating against earth-type attacks (Decimal).

against\_ice: Pokémon's damage rating against ice-type attacks (Decimal).

against\_normal: Pokémon's damage rating against normal-type attacks (Decimal).

against\_poison: Pokémon's damage rating against poison-type attacks (Decimal).

against\_psychic: Pokémon's damage rating against psychic-type attacks (Decimal).

against\_rock: Pokémon's damage rating against rock-type attacks (Decimal).

against\_steel: Pokémon's damage rating against steel-type attacks (Decimal).

against\_water: Pokémon's damage rating against water-type attacks (Decimal).

attack: Pokémon's base attack power (Integer).

base\_egg\_steps: The number of steps required for a Pokémon's egg to hatch (Integer).

base\_happiness: Pokémon's base happiness value (Integer).

base\_total: Pokémon's total base statistic value (Integer).

capture\_rate: Pokémon's capture rate (Text).

classification: The classification with which the Pokémon is identified in the Sun and Moon Pokedex (Text).

defense: Pokémon's basic defense (Integer).

experience\_growth: Pokémon's experience point growth rate (Integer).

height\_m: Pokémon's height (in meters) (Decimal).

hp: Pokémon's base health (Integer).

japanese\_name: Pokémon's Japanese name (Text).

name: Pokémon's English name (Text).

percentage\_male: The percentage of the Pokémon type that is male (If left blank, the Pokémon is genderless) (Decimal).

pokedex\_number: Pokémon's entry number in the National Pokedex (Integer).

sp\_attack: Pokémon's basic special attack power (Integer).

sp\_defense: Pokémon's basic special defense (Integer).

speed: Pokémon's base speed (Integer).

type1: Pokémon's primary type (Text).

type2: Pokémon's secondary type (if any) (Text).

weight\_kg: Pokémon's weight (in kilograms) (Decimal number).

generation: The generation in which the Pokémon was first introduced (Integer).

is\_legendary: Value indicating whether the Pokémon is legendary (1: Legendary, 0: Not legendary) (Integer).



You can find the details of the data types of in. json format in the attachment.

1. Analyze the data in terms of descriptive statistics.

You can find summary of descriptive statistics:



General Information:

The dataset contains 801 observations (Pokémon) and 41 variables.

There are no missing values in the dataset.

Numeric Variables:

attack:

Mean: 77.5

Median: 75.0

Standard Deviation: 32.5

Minimum: 5

Maximum: 190

Skewness: 0.53 (Positive skewness, i.e., values are skewed to the right)

Kurtosis: 0.07 (distribution close to normal)

base\_egg\_steps (Egg Steps):

Average 5841.9

Median: 5120.0

Standard Deviation: 2907.5

Minimum: 1200

Maximum: 30720

Skewness: 2.96 (High positive skewness, values spread to the right)

Kurtosis: 7.58 (Very pointed distribution)

base\_total (Total Base Value):

Average 421.7

Median: 435.0

Standard Deviation: 119.0

Minimum: 175

Maximum: 780

Skewness: 0.17 (Slight positive skewness)

Kurtosis: -0.53 (Flat distribution)

height\_m (Height):

Average 1.18 meters

Median: 1.0 meter

Standard Deviation 1.49 meters

Minimum: 0.1 meter

Maximum: 14.5 meters

Skewness: 3.53 (High positive skewness)

Kurtosis: 17.12 (Very pointed distribution)

weight\_kg (Weight):

Average 51.14 kg

Median: 24.9 kg

Standard Deviation 110.0 kg

Minimum: 0.1 kg

Maximum: 999.9 kg

Skewness: 4.87 (High positive skewness)

Kurtosis: 31.74 (Very pointed distribution)

Categorical Variables:

abilities:

Total unique abilities: 482

Most common ability: 'Overgrow', 'Chlorophyll'

type1 (Primary Type):

Total unique type: 18

The most common species: Water

type2 (Secondary Type):

Total unique type: 19

The most common species: Flying

Key Findings and Comments:

Skewness and Kurtosis: Many numerical variables have high skewness and kurtosis values. This indicates that most of the data are clustered around a particular value and that there are some outliers. In particular, height\_m and weight\_kg variables have high skewness and kurtosis values and are variables where outliers are effective.

Attack Value: The average attack value of Pokémon is 77.5. Since the skewness of the values is low, it can be said that the distribution is generally symmetrical.

Egg Steps (base\_egg\_steps): The mean number of steps required for Pokémon eggs to hatch is 5841.9, with a fairly high standard deviation (2907.5). This shows that the number of steps is distributed over a wide range.

Total Base Value (base\_total): The mean of the Pokémon's total base value is 421.7, indicating that the distribution is almost symmetrical (skewness of 0.17).

Height (height\_m) and Weight (weight\_kg): The high skewness and kurtosis values of these two variables indicate that much of the data is concentrated at low values and that some Pokémon are extremely large or heavy.

Plot pairwise variability of parameters:

Bivariate analysis is used to analyze the relationship between two variables. This analysis is important for understanding how variables interact with each other. In your Pokémon dataset, we can use scatter plots and heatmaps to visualize the relationships between numeric variables.

1. Scatter Plot

A scatter plot is used to visualize the relationship between two numerical variables. In this graph, one axis represents one variable, and the other axis represents the other variable. The scatter plot is useful for seeing the distribution of data points and possible correlations.

2. Heatmap

Heatmap is used to visualize correlations between numerical variables. This graph shows the correlation coefficients between pairs of variables with a color scale. Heatmap is especially useful for quickly visualizing relationships between a large number of variables.