**AIM**

This report presents an exploratory data analysis (EDA) on a Pokémon dataset containing multiple numerical and categorical attributes. The objective is to clean the dataset, analyse the distribution of features, and explore relationships between different attributes.

The dataset includes the following key columns:

* **Numerical Variables**: Height (m), Weight (kg), Pokedex Number
* **Categorical Variables**: Type1, Type2, Classification, Generation, Legendary Status

**Data Cleaning**

**Handling Missing Values**

Missing values were detected and handled using forward fill imputation to ensure continuity in categorical attributes.

**Removing Duplicates**

Duplicate records were identified and removed to avoid redundant observations.

**Outlier Detection & Removal**

Outliers in Height, Weight, and Pokedex Number were detected using the Interquartile Range (IQR) method.

Records outside 1.5 times the IQR range were removed to prevent skewed analysis.

**Univariate Analysis**

**Summary Statistics**

* Mean, median, mode, variance, and skewness were computed for numerical variables.
* Key stats being:

Height (m): Mean = 0.94, Median = 0.80, Mode = 0.60, Variance = 0.29, Skewness = 0.78.

Weight (kg): Mean = 32.29, Median = 21.20, Mode = 1.00, Variance = 1018.81, Skewness = 1.27.

**Key findings**

1. Pokedex number is uniformly distributed.
2. Most of the Pokémons are added in generation 1 and 5.
3. There is a right skewness in height and weight graphs indicating most of the pokemons are smaller.
4. Water, normal and grass type most numerous whereas flying type are relatively less in number.
5. Legendary Pokémons are less in number indicating their rarity.

**Visualizations**

* **Histograms** were used to observe the distribution of numerical features.
* **Box plots** highlighted the presence of outliers in weight and height.

**Bivariate Analysis**

**Correlation Matrix**

* A heatmap was generated to examine correlations between numerical variables:

1. Height and Weight showed a moderate positive correlation, indicating taller Pokémon tend to weigh more.
2. Pokedex Number had no strong correlation with other numerical features except for Generation.

**Scatter Plots**

* **Height vs. Weight** showed a non-linear relationship with a few exceptionally large Pokémon outliers.

**Categorical vs. Numerical Analysis**

* **Box plots** illustrated weight differences among Pokémon types, showing significant variation among steel, ground, fighting, ground.
* **Violin plots** showed height distributions across different generations, with significant variation among generation 8.

**Categorical vs. Categorical Analysis**

* **Stacked bar charts** revealed the distribution of Legendary Pokémon across types and generations:
  + Certain types, like **Dragon** and **Psychic**, had a higher proportion of Legendary Pokémon.
  + Generation distribution showed that **later generations introduced more Legendary Pokémon.**

**Multivariate Analysis**

**Pair Plots**

* A pair plot was generated to observe multi-variable relationships, particularly among height, weight, and Pokedex Number.

**Heatmap of Categorical & Numerical Interactions**

* A heatmap was used to analyse the average weight of Pokémon across types and generations:
  + Certain types (e.g., **Steel, Rock**) tend to have heavier Pokémon on average.

**Cluster Map for Hierarchical Clustering**

* A **cluster map** was generated to explore hidden patterns among numerical attributes.
* It provided insights into how certain Pokémon characteristics group together.

**Grouped Comparisons**

* A grouped box plot was used to analyse how Pokémon weight varies across generations for Legendary and Non-Legendary Pokémon.
  + **Legendary Pokémon are generally heavier than non-legendary ones across all generations.**

**Conclusion**

* **Data Cleaning** helped remove inconsistencies, ensuring a reliable dataset.
* **Univariate Analysis** showed that Pokémon attributes have highly skewed distributions.
* **Bivariate Analysis** revealed interesting relationships between categorical and numerical features, particularly Pokémon weight by type.
* **Multivariate Analysis** uncovered patterns in Pokémon types and their generational differences.
* The findings provide a deeper understanding of Pokémon attributes and their interdependencies, demonstrating the importance of EDA in data analysis.