Mini Project 2: Semi- structured Data Report (Pokémon Go Analysis)
Submitted by: Kruti Kotadia
Date: 17th November 2023

## The Data and its Source: -

The dataset used for this analysis is on Pokémon Go dataset, which was retrieved from GitHub. This report embarks on the collation and analysis of a dataset focused on Pokémon attributes, leveraging Python for the analytical process. The dataset originates from a JSON file containing comprehensive information about various Pokémon, including details like height, weight, spawn chances, evolutions, weaknesses, and more. The JSON file was stored in collab drive.

## Data Exploration and Cleaning: -

The process of data exploration and cleaning is crucial to ensure the dataset's integrity and prepare it for meaningful analysis. In this analysis of Pokémon attributes, several key steps were taken to enhance the quality of the dataset.

**Conversion of Attributes:**

The attributes 'height' and 'weight' were initially stored as strings with a unit suffix ('m' and 'kg', respectively). To facilitate numeric analysis, these attributes were cleaned by removing the units and converting them into float data types.

**Handling Spawn Time:**

The 'spawn time' attribute, representing the time Pokémon spawn, was initially loaded with potential 'NaN' values. These values were replaced and subsequently converted to datetime format for further analysis.

**Weakness Exploration:**

To analyze weaknesses effectively, all weaknesses for each Pokémon were extracted into a single list. The occurrences of each weakness were then counted to identify the most common weaknesses. The dataset was expanded to explore weaknesses based on Pokémon types, leading to a comprehensive understanding of type-specific weaknesses.

## Comparison Questions and Analysis: -

1. During which hours of the day do we observe the highest and lowest numbers of Pokémon spawns?
    - Unit of Analysis: Hour of the day
    - Comparison Value: Number of Pokémon Spawns
    - Computation: (Count the number of Pokémon spawns for each hour of the day.)
      Analyzing Pokémon spawns, we identify peak and off-peak hours, considering the

unit of analysis as the hour of the day, with a computation based on counting the number of Pokémon spawns for each hour.

2. How does the average height and weight of Pokémon vary across different types of Pokémon?

    - Unit of Analysis: Pokémon Type
    - Comparison Value: Average Height and Weight
    - Computation: (Calculate the average height and weight for each Pokémon type.)
    - Examining Pokémon diversity, we assess how the average height and weight differ across types, focusing on the Pokémon type as the unit of analysis and computing the average height and weight for each type.

3. How does the distribution of weaknesses vary among different Pokémon types, revealing patterns or concentrations within specific types?

    - Unit of Analysis: Pokémon Type
    - Comparison Value: Count of Weaknesses
    - Computation: (Identify the Pokémon type with the most and least weaknesses against it each type.) Investigating Pokémon weaknesses, we explore variations in weakness distribution among types, unveiling patterns and concentrations within specific types by considering Pokémon type as the unit of analysis and counting weaknesses for each type.

## Program Description: -

The Python program leverages popular libraries such as json, pandas, matplotlib, seaborn, and NumPy to conduct a comprehensive analysis and visualization of data sourced from a Pokémon dataset. The dataset is loaded from a JSON file, offering information on various Pokémon species.

## Conclusion: -

In conclusion, the cleaning and exploration steps ensured the data's quality and reliability. By answering three specific comparison questions, we gained valuable insights into the dataset, which can be useful for various purposes.