

# MINI PROJECT 1 - Data Analysis and Visualization



## Group Members Name

1. Dinesh Subramanian

2. S.Aishwarya

3. Vinod A

In [84]:

```
1 #import libraries
2 import numpy as np
3 import pandas as pd
4 import seaborn as sns
5 import matplotlib.pyplot as plt
6 import warnings
7 warnings.filterwarnings("ignore")
```

1. Read the dataset 'pokemon.csv'. [1 mark]

```
In [85]: 1 df = pd.read_csv("Pokemon.csv")
        2 df
```

Out[85]:

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

800 rows × 13 columns



## 2.List the columns in the dataframe. [2 marks]

```
In [86]: 1 df.columns
```

Out[86]: Index(['#', 'Name', 'Type 1', 'Type 2', 'Total', 'HP', 'Attack', 'Defense', 'Sp. Atk', 'Sp. Def', 'Speed', 'Generation', 'Legendary'], dtype='object')

## 3.Drop the column '#'. [1 mark]

```
In [87]: 1 df = df.drop(columns="#")
        2 df
```

Out[87]:

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

800 rows × 12 columns



#### 4. Check if the dataset has null values. [2 marks]

```
In [88]: 1 # to check with respect to each data
        2 df.isnull()
        3
        4 # to check the number of null values present in each column
        5 df.isnull().sum()
        6
        7 # to check the number of null values present in dataset
        8 total_number = df.isnull().sum().sum()
        9
        10 # therefore
        11
        12 if (total_number != 0):
        13     print("The dataset have", total_number, "null values")
        14 else:
        15     print("The dataset doesn't have null values")
```

The dataset have 386 null values

#### 5. Drop the records with null values. [2 marks]

```
In [89]: 1 df = df.dropna()
         2 df
```

Out[89]:

	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Lev
0	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	
1	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	
2	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	
3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	
6	Charizard	Fire	Flying	534	78	84	78	109	85	100	1	
...	...	...	...	...	...	...	...	...	...	...	...	...
795	Diancie	Rock	Fairy	600	50	100	150	100	150	50	6	
796	DiancieMega Diancie	Rock	Fairy	700	50	160	110	160	110	110	6	
797	HoopaHoopa Confined	Psychic	Ghost	600	80	110	60	150	130	70	6	
798	HoopaHoopa Unbound	Psychic	Dark	680	80	160	60	170	130	80	6	
799	Volcanion	Fire	Water	600	80	110	120	130	90	70	6	

414 rows × 12 columns

## 6. Find the most frequent type of pokemon in Type 1 and Type 2. [2 marks]

```
In [90]: 1 df[['Type 1', 'Type 2']].mode()
```

Out[90]:

	Type 1	Type 2
0	Water	Flying

## 7. Find the mean defense value and attack value of the pokemons. [2 marks]

```
In [91]: 1 df[['Attack', 'Defense']].mean()
```

Out[91]: Attack 83.173913  
Defense 79.676329  
dtype: float64

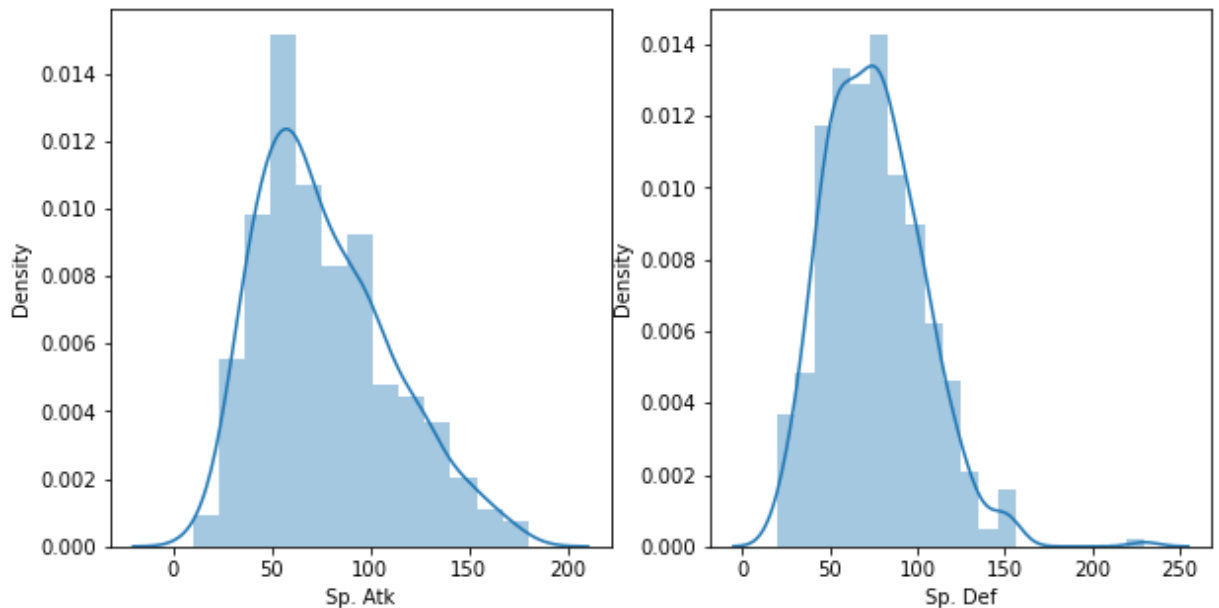
## 8. What is the maximum 'total' of the pokemons? List the pokemons with top 5 total. [2 marks]

```
In [92]: 1 df.sort_values(by=['Total'],ascending=False, inplace=True)
2         df["Name"].head(5)
```

```
Out[92]: 426      RayquazaMega Rayquaza
163      MewtwoMega Mewtwo X
424      GroudonPrimal Groudon
494      GarchompMega Garchomp
413      MetagrossMega Metagross
Name: Name, dtype: object
```

**9.How skewed are the features 'Sp. Def' and 'Sp. Atk'? Write down your inference of the same. [4 marks]**

```
In [93]: 1 fig = plt.subplots(figsize=(10,5))
2         plt.subplot(1, 2, 1)
3
4         sns.distplot(df['Sp. Atk'])
5         plt.subplot(1, 2, 2)
6         sns.distplot(df['Sp. Def'])
7         plt.show()
8         print ("Shew Values for \nSp.Atck :",df['Sp. Atk'].skew(),"\nSp.Def :",df['S
9
10        print("Inference : As part of plots we can figure out that the 'Sp. Atk' and
```



Shew Values for

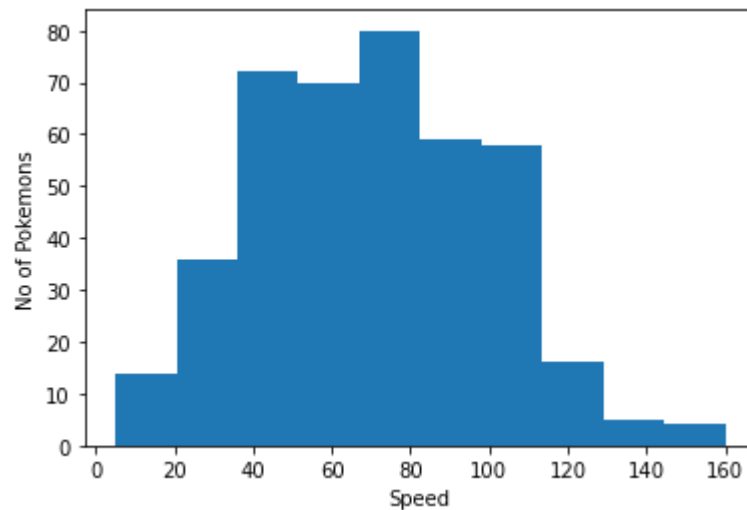
Sp.Atck : 0.6531678282717396

Sp.Def : 0.7653450163443561

Inference : As part of plots we can figure out that the 'Sp. Atk' and 'Sp. Def' are been postive skewed. Therefore, mean > median > mode.

**10.Plot a histogram for the feature 'Speed'. What are your inferences from the same? [2 marks]**

```
In [94]: 1 plt.hist(x = df['Speed'])
2 plt.xlabel('Speed')
3 plt.ylabel('No of Pokemons')
4 plt.show()
5 print("Inference : ")
6 print("Least number of Pokemons have the top speed between 140-160")
7 print("Max Number of Pokemons have the speed comes under the top speed betwe
```



Inference :

Least number of Pokemons have the top speed between 140-160

Max Number of Pokemons have the speed comes under the top speed between 75-80

```
In [ ]: 1
```