Lab 4: Data Visualization and EDA

CPE232 Data Models

In [1]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
```

1. Load all pokemon data (from previous Homework)

In [2]:

```python
# write your code here
df = pd.read_csv("pokemon.csv")
df
```

Out[2]:

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

800 rows × 13 columns

2. Are there any missing values? If so, in which column?

   Ans: there are missing values in "Type 2" column

In [13]:

```
df.isnull().sum()
```

Out[13]:

```
#               0
Name            0
Type 1          0
Type 2        386
Total           0
HP              0
Attack          0
Defense         0
Sp. Atk         0
Sp. Def         0
Speed           0
Generation      0
Legendary       0
dtype: int64
```

3. Calculate the average 'Attack' of each pokemon 'Type 1'. Which type has the most Attack stats?

   Ans: Dragon type

In [16]:

```
df.groupby("Type 1")["Attack"].mean()
```

Out[16]:

```
Type 1
Bug          70.971014
Dark         88.387097
Dragon      112.125000
Electric     69.090909
Fairy        61.529412
Fighting     96.777778
Fire         84.769231
Flying       78.750000
Ghost        73.781250
Grass        73.214286
Ground       95.750000
Ice          72.750000
Normal       73.469388
Poison       74.678571
Psychic      71.456140
Rock         92.863636
Steel        92.703704
Water        74.151786
Name: Attack, dtype: float64
```

4. Aggregate count of each Pokemon Type 1

In [21]:

```
type_count =df.groupby("Type 1")["Name"].count()
type_count
```

Out[21]:

```
Type 1
Bug          69
Dark         31
Dragon       32
Electric     44
Fairy        17
Fighting     27
Fire         52
Flying        4
Ghost        32
Grass        70
Ground       32
Ice          24
Normal       98
Poison       28
Psychic      57
Rock         44
Steel        27
Water       112
Name: Name, dtype: int64
```

5. Create a visualization that show proportion of each pokemon Type 1 in the dataset
   Hint: plotly

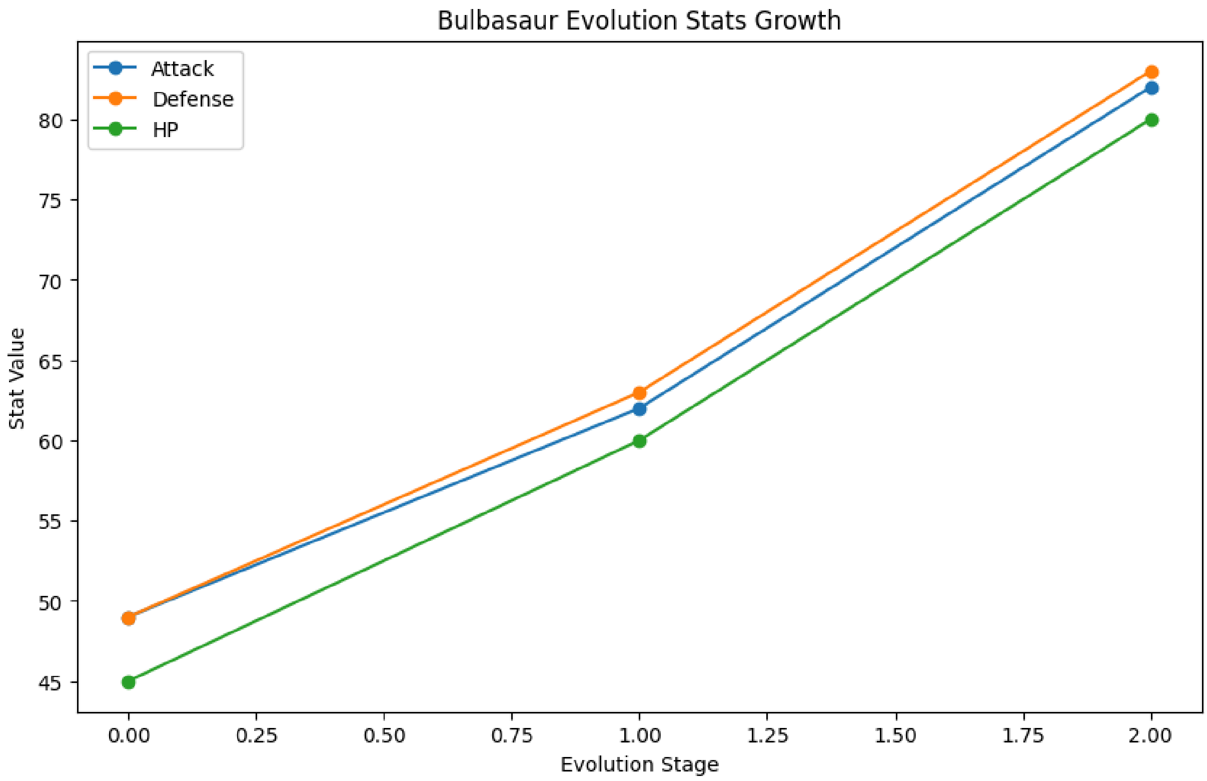
In [38]:

```
# write your code here
fig = px.pie(df, names='Type 1', title='Pokemon Type 1 Proportion', hole=0.3)
fig.show()
import plotly.io as pio
pio.renderers.default = "png+notebook_connected+vscode"
```

6. Create a line plot of 'Attack', 'Defense', and 'HP' of Bulbasaur evolution.
Bulbasaur -> Ivysaur -> Venusaur

In [31]:

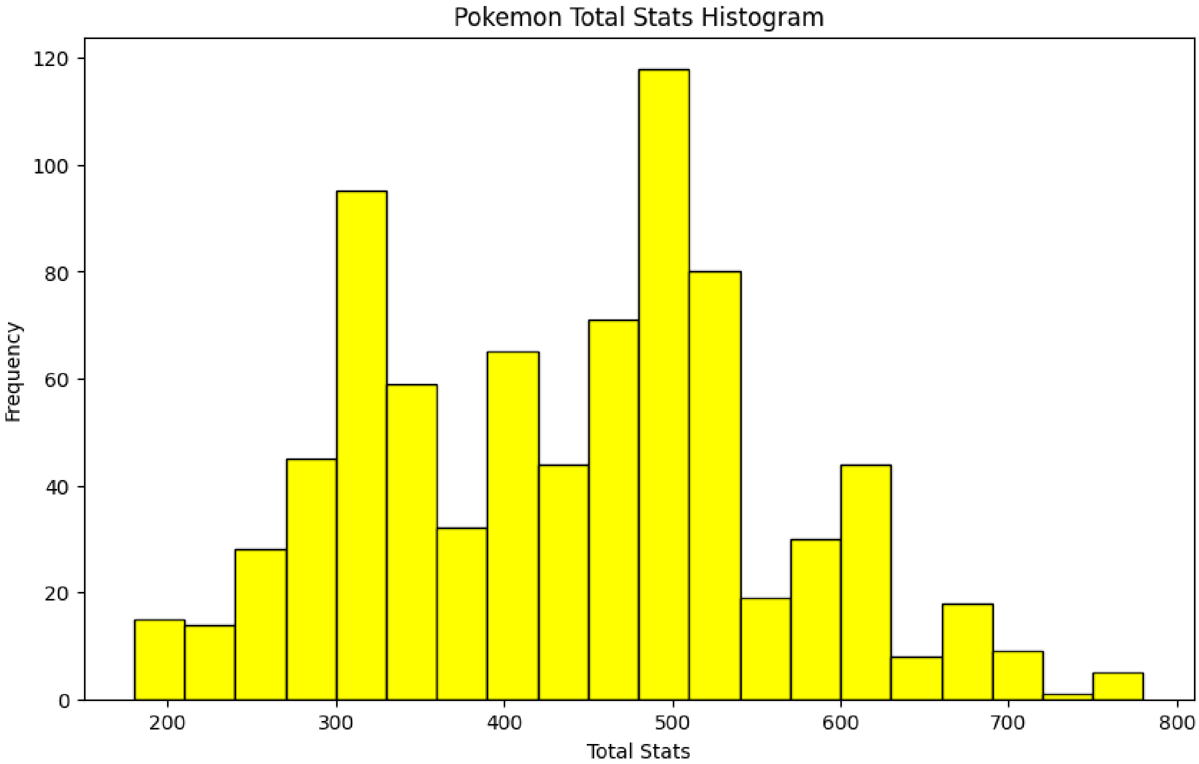
```
evo = df.iloc[:3]
plt.figure(figsize=(10, 6))
plt.plot(evo.index, evo['Attack'], label='Attack', marker='o')
plt.plot(evo.index, evo['Defense'], label='Defense', marker='o')
plt.plot(evo.index, evo['HP'], label='HP', marker='o')
plt.title('Bulbasaur Evolution Stats Growth')
plt.xlabel('Evolution Stage')
plt.ylabel('Stat Value')
plt.legend()
plt.show()
```



7. Create a histogram of Pokemon total stats

In [34]:

```python
plt.figure(figsize=(10, 6))
plt.hist(df['Total'], bins=20, color='yellow', edgecolor='black')
plt.title('Pokemon Total Stats Histogram')
plt.xlabel('Total Stats')
plt.ylabel('Frequency')
plt.show()
```


Pokemon Total Stats Histogram
Frequency
0
20
40
60
80
100
120
200
300
400
500
600
700
800
Total Stats