**Name : Mohit Manish Bhavsar**

**Roll No : 20U437**

**Div : 4**

import pandas as pd

df=pd.read\_csv('Salary Dataset.csv') df

# Company Name Job Title Salaries

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | **Reported** |  | |
| **0** | Mu Sigma | Data Scientist | 105.0 | Bangalore | ₹6,48,573/yr |
| **1** | IBM | Data Scientist | 95.0 | Bangalore | ₹11,91,950/yr |

**Location Salary**

1. Tata Consultancy

Services

Data Scientist 66.0 Bangalore ₹8,36,874/yr

1. Impact Analytics Data Scientist 40.0 Bangalore ₹6,69,578/yr
2. Accenture Data Scientist 32.0 Bangalore ₹9,44,110/yr

**...** ... ... ... ... ...

**4339** TaiyōAI Machine Learning Scientist

1.0 Mumbai ₹5,180/mo

**4340** Decimal Point Analytics

Machine Learning

Developer

1.0 Mumbai ₹7,51,286/yr

**4341** MyWays Machine Learning Developer

1.0 Mumbai ₹4,10,952/yr

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| df.head()  **Company Name** | **Job Title** | **Salaries** | **Reported** | **Location** | **Salary** |
| **0** Mu Sigma | Data Scientist |  | 105.0 | Bangalore | ₹6,48,573/yr |
| **1** IBM | Data Scientist |  | 95.0 | Bangalore | ₹11,91,950/yr |
| **2** Tata Consultancy Services | Data Scientist |  | 66.0 | Bangalore | ₹8,36,874/yr |
| **3** Impact Analytics | Data Scientist |  | 40.0 | Bangalore | ₹6,69,578/yr |
| **4** Accenture | Data Scientist |  | 32.0 | Bangalore | ₹9,44,110/yr |

df.tail()

# Company Name Job Title Salaries Reported

**Location Salary**

**4339** TaiyōAI

Machine Learning 1.0 Mumbai ₹5,180/mo Scientist

# 4340

Decimal Point

Analytics

Machine Learning 1.0 Mumbai ₹7,51,286/yr Developer

**4341** MyWays

Machine Learning 1.0 Mumbai ₹4,10,952/yr Developer

df.columns

Index(['Company Name', 'Job Title', 'Salaries Reported', 'Location', 'Salary'], dtyp

df[['Company Name','Job Title']]

# Company Name Job Title

1. Mu Sigma Data Scientist
2. IBM Data Scientist
3. Tata Consultancy Services Data Scientist
4. Impact Analytics Data Scientist
5. Accenture Data Scientist

**...** ... ...

**4339** TaiyōAI Machine Learning Scientist **4340** Decimal Point Analytics Machine Learning Developer **4341** MyWays Machine Learning Developer **4342** Market Pulse Technologies Software Engineer - Machine Learning **4343** vPhrase Machine Learning Engineer 4344 rows × 2 columns

df.shape

(4344, 5)

df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 4344 entries, 0 to 4343

Data columns (total 5 columns):

# Column Non-Null Count Dtype

1. Company Name 4341 non-null object
2. Job Title 4344 non-null object
3. Salaries Reported 4342 non-null float64
4. Location 4344 non-null object
5. Salary 4344 non-null object dtypes: float64(1), object(4)

memory usage: 169.8+ KB

df.isna()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Company Name** | **Job** | **Title** | **Salaries** | **Reported** | **Location** | **Salary** |
| **0** False |  | False |  | False | False | False |
| **1** False |  | False |  | False | False | False |
| **2** False |  | False |  | False | False | False |
| **3** False |  | False |  | False | False | False |
| **4** False |  | False |  | False | False | False |
| **...** ... |  | ... |  | ... | ... | ... |
| **4339** False |  | False |  | False | False | False |
| **4340** False |  | False |  | False | False | False |
| **4341** False |  | False |  | False | False | False |
| **4342** False |  | False |  | False | False | False |
| **4343** False |  | False |  | False | False | False |
| df.isna().sum() |  |  | | | | |
| Company Name | 3 |
| Job Title | 0 |
| Salaries Reported | 2 |
| Location | 0 |
| Salary | 0 |

dtype: int64

4344 rows × 5 columns

df['Company Name'][3]

'Impact Analytics'

df['Location'][4]

'Bangalore'

df['Location'][0]=None

/usr/local/lib/python3.7/dist-packages/ipykernel\_launcher.py:1: SettingWithCopyWarni A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/u](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) """Entry point for launching an IPython kernel.

df.isna().sum()

|  |  |
| --- | --- |
| Company Name | 3 |
| Job Title | 0 |
| Salaries Reported | 2 |
| Location | 1 |

Salary 0

dtype: int64

df['Location'][0]='Pune'

/usr/local/lib/python3.7/dist-packages/ipykernel\_launcher.py:1: SettingWithCopyWarni A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/u](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) """Entry point for launching an IPython kernel.

df.isna().sum()

|  |  |
| --- | --- |
| Company Name | 3 |
| Job Title | 0 |
| Salaries Reported | 2 |
| Location | 0 |
| Salary | 0 |

dtype: int64

df=df.rename(columns={'Company Name':'Company\_Name','Salaries Reported':'Salaries\_Reported

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| df |  | **Company\_Name** | **Job Title** | **Salaries\_Reported** | **Location** | **Salary** |
|  | **0** | Mu Sigma | Data Scientist | 105.0 | Pune | ₹6,48,573/yr |
|  | **1** | IBM | Data Scientist | 95.0 | Bangalore | ₹11,91,950/yr |

* 1. Tata Consultancy

Services

Data Scientist 66.0 Bangalore ₹8,36,874/yr

* 1. Impact Analytics Data Scientist 40.0 Bangalore ₹6,69,578/yr
  2. Accenture Data Scientist 32.0 Bangalore ₹9,44,110/yr

**...** ... ... ... ... ...

**4339** TaiyōAI Machine Learning

|  |  |  |
| --- | --- | --- |
| 1.0 | Mumbai | ₹5,180/mo |
| 1.0 | Mumbai | ₹7,51,286/yr |
| 1.0 | Mumbai | ₹4,10,952/yr |
| 1.0 | Mumbai | ₹16,12,324/yr |

Scientist

**4340** Decimal Point Analytics

Machine Learning

Developer

**4341** MyWays Machine Learning Developer

**4342** Market Pulse Technologies

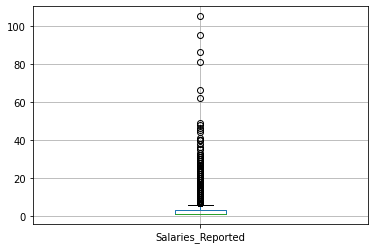
Software Engineer - Machine Learning

df.columns

Index(['Company\_Name', 'Job Title', 'Salaries\_Reported', 'Location', 'Salary'], dtyp

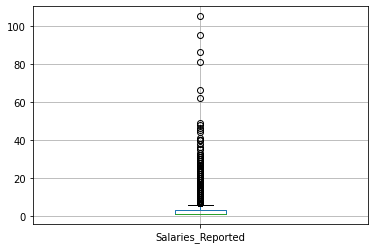
df.boxplot(column=['Salaries\_Reported'],grid=True)

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fdc248eb710>



df.boxplot()

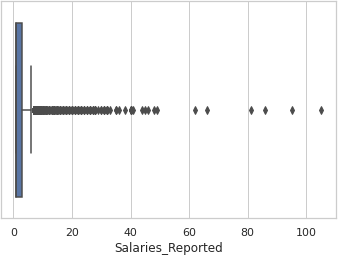
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fdc2443e950>



import seaborn as sns

sns.set\_theme(style="whitegrid")

ax =sns.boxplot(x=df['Salaries\_Reported'])



df['Salaries\_Reported'].mean() 2.7759097190234914

print("Highest allowed",df['Salaries\_Reported'].mean()+3\*df['Salaries\_Reported'].std()) print("Lowest allowed",df['Salaries\_Reported'].mean()-3\*df['Salaries\_Reported'].std())

Highest allowed 18.215492136673088

Lowest allowed -12.663672698626105

df[(df['Salaries\_Reported']>18)]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Company\_Name** | **Job Title** | **Salaries\_Reported** | **Location** | **Salary** |
| **0** | Mu Sigma | Data Scientist | 105.0 | Pune | ₹6,48,573/yr |
| **1** | IBM | Data Scientist | 95.0 | Bangalore | ₹11,91,950/yr |

1. Tata Consultancy

Services

Data Scientist 66.0 Bangalore ₹8,36,874/yr

1. Impact Analytics Data Scientist 40.0 Bangalore ₹6,69,578/yr
2. Accenture Data Scientist 32.0 Bangalore ₹9,44,110/yr

**...** ... ... ... ... ...

**3622** LTI Data Engineer 33.0 Mumbai ₹5,00,000/yr

**3623** Tata Consultancy

Services

Data Engineer 19.0 Mumbai ₹5,00,000/yr

**3748** Quantiphi Machine Learning Engineer

**4115** Amazon Machine Learning Data Associate

28.0 Bangalore ₹8,75,066/yr

38.0 Hyderabad ₹2,75,841/yr

df=df[(df['Salaries\_Reported']<=18)]

import seaborn as sns

sns.set\_theme(style="whitegrid")

ax =sns.boxplot(x=df['Salaries\_Reported'])

