**MODULE 6 PACKAGES**

**Implement in Python**

1. For the dataset “Indian\_cities”,
2. Find out top 10 states in female-male sex ratio

***import pandas as pd***

***a = pd.read\_csv("C:/Users/me/Downloads/Python Problem Statements (1)/Python Problem Statements/Indian\_cities.csv")***

***sr=a.nlargest(10,['sex\_ratio'])***

***sr=sr.iloc[:,2:]***

|  |  |  |
| --- | --- | --- |
| **index** | **state\_name** | **sex\_ratio** |
| 278 | KERALA | 1093 |
| 275 | KERALA | 1077 |
| 14 | KERALA | 1076 |
| 461 | KERALA | 1076 |
| 459 | KERALA | 1064 |
| 220 | MANIPUR | 1055 |
| 358 | KERALA | 1053 |
| 245 | ANDHRA PRADESH | 1046 |
| 378 | PUDUCHERRY | 1045 |
| 430 | MEGHALAYA | 1042 |

1. Find out top 10 cities in total number of graduates

***tg=a.nlargest(10,['total\_graduates'])***

|  |  |  |
| --- | --- | --- |
| index | name\_of\_city | total\_graduates |
| 141 | Delhi | 2221137 |
| 185 | Greater Mumbai | 1802371 |
| 72 | Bengaluru | 1591163 |
| 184 | Greater Hyderabad | 1164149 |
| 119 | Chennai | 879695 |
| 274 | Kolkata | 818476 |
| 7 | Ahmadabad | 769858 |
| 380 | Pune | 656508 |
| 288 | Lucknow | 596990 |
| 225 | Jaipur | 533148 |

1. Find out top 10 cities and their locations in respect of total effective\_literacy\_rate.

***elr=a.nlargest(10,[*** “***effective\_literacy\_rate\_total”])***

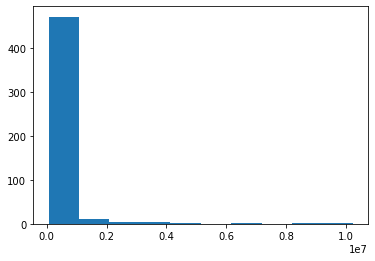
|  |  |  |  |
| --- | --- | --- | --- |
| **index** | **name\_of\_city** | **effective\_literacy\_rate\_total** | **location** |
| 9 | Aizawl | 98.8 | 23.727107,92.7176389 |
| 271 | Kochi | 97.49 | 9.9312328,76.2673041 |
| 461 | Thrissur | 97.24 | 10.5276416,76.2144349 |
| 278 | Kozhikode | 96.8 | 11.2587531,75.78041 |
| 14 | Alappuzha | 96.56 | 9.4980667,76.3388484 |
| 264 | Khardaha | 95.5 | 22.7002943,88.3753455 |
| 333 | Nagercoil | 95.35 | 8.1832857,77.4118996 |
| 13 | Alandur | 95.15 | 12.9974873,80.2006371 |
| 352 | North Barrackpur | 94.78 | 22.7902358,88.367179 |
| 431 | Shimla | 94.67 | 31.1048145,77.1734033 |

1. For the data set “Indian\_cities”
2. Construct histogram on literates\_total and comment about the inferences

***import matplotlib.pyplot as plt***

***import numpy as np***

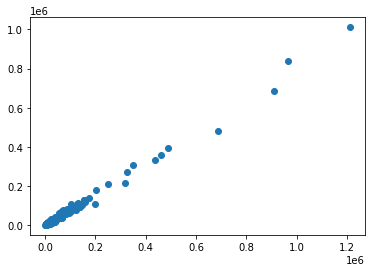
***plt.hist(a.literates\_total )***



* Data is Right/Positively skwed

1. Construct scatter plot between male graduates and female graduates

***plt.scatter(a.male\_graduates,a.female\_graduates)***



1. For the data set “Indian\_cities”

a)Construct Boxplot on total effective literacy rate and draw inferences

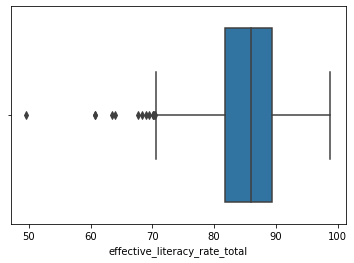
***import matplotlib.pyplot as plt***

***import numpy as np***

***import seaborn as sns***

***sns.boxplot(a.effective\_literacy\_rate\_total)***

***plt.title('Boxplot')***



* Outliers are present In effective\_literacy\_rate\_total
* Data is left/negatively skewed

b)Find out the number of null values in each column of the dataset and delete them.

***a.isna().sum()***

* There is no null values in the dataset.

|  |  |
| --- | --- |
| column names | No of null values |
| name\_of\_city | 0 |
| state\_code | 0 |
| state\_name | 0 |
| dist\_code | 0 |
| population\_total | 0 |
| population\_male | 0 |
| population\_female | 0 |
| 0-6\_population\_total | 0 |
| 0-6\_population\_male | 0 |
| 0-6\_population\_female | 0 |
| literates\_total | 0 |
| literates\_male | 0 |
| literates\_female | 0 |
| sex\_ratio | 0 |
| child\_sex\_ratio | 0 |
| effective\_literacy\_rate\_total | 0 |
| effective\_literacy\_rate\_male | 0 |
| effective\_literacy\_rate\_female | 0 |
| Location | 0 |
| total\_graduates | 0 |
| male\_graduates | 0 |
| female\_graduates | 0 |