REPORT ON ANNUAL SALARIES DISTRIBUTION

Name: ASWATHI BIBEESH

Student Id: 22075385

1. Data Description:

The dataset, which is yearly salary, was taken from 'data5-1.csv'. We used NumPy, Matplotlib, and SciPy, three Python libraries, to try to figure out this dataset's distribution and important statistical values. A histogram with 30 bins has been made for visualization to comprehend the distribution and features of yearly salaries. To model the data's underlying probability density function, a normal distribution has also been fitted.

2. Distribution Overview:

A histogram with 30 bins is made to obtain insight into the annual salary dispersion. A visual depiction of the frequency of wage values within various ranges is given by the histogram. To evaluate how well the data fits into a typical normal distribution, a fitted normal distribution is also superimposed on the histogram. The central tendency and dispersion of the data are described by the computed mean (mu) and standard deviation (std) of the fitted normal distribution.

3. Calculation of Mean (W):

Utilizing the np. mean() function from the numpy package, the mean annual salary (W) is determined. The following is the formula:

$$W = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Where 'n' is the number of data points (salaries) and 'xi' represents individual salary values.

4. Calculation of Required Value X (10th Percentile):

Using np. Percentile (data, 10), the value at the 10th percentile (X) is computed. The wage that 10% of the data falls below is represented by this figure. The following is the percentile calculation formula:

X=Percentile(p)= value of the data point at the p-th percentile, In this case, p=10.

5. Graphical Representation:

Probability Density Function of Annual Salaries
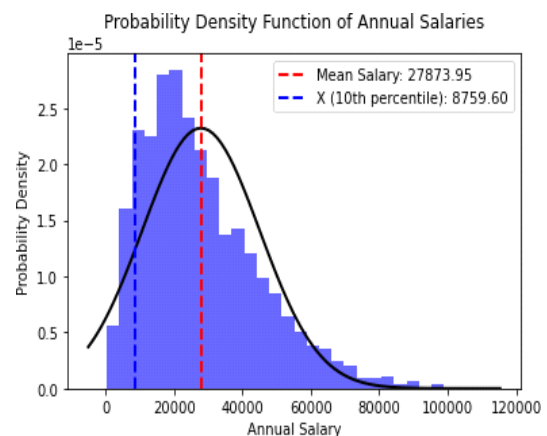
Results:

For the given dataset, the calculated mean salary (W) is 27873.95 and standard deviation. (X) is 8759.60 are displayed on the graph generated by the python code.

Conclusion:

Through the display of important statistical metrics like the mean and the value at a certain percentile, this study sheds light on the distribution of yearly incomes. Our comprehension of the dataset's properties is improved by the normal distribution fit and graphical depiction, which supports well-informed salary decision-making.