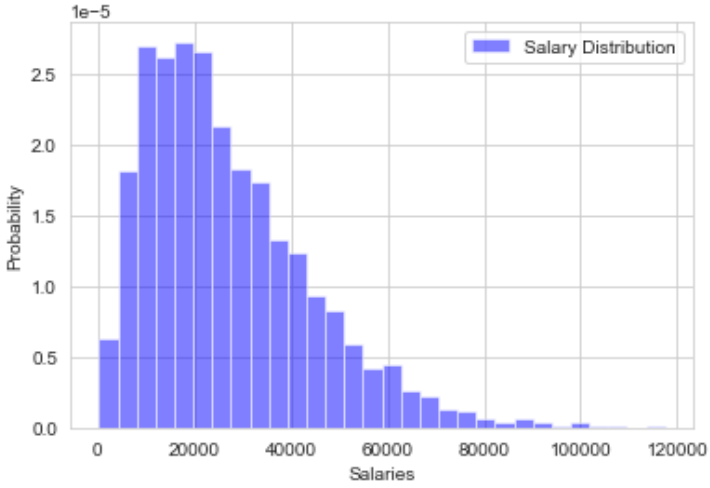
Salary Distribution Analysis

Describe the data :

The data is provided in the form of a CSV file . It is made up of a single column that represents salary. You imported this information into a Pandas DataFrame and retrieved the ‘salary’ column.

Describe the distribution :

The salary distribution is represented graphically via a histogram with 30 bins and probability density ('density=True'). The shape and dispersion of the salary distribution can be seen in



Calculation of Mean Value :

The mean salary (~W) is calculated using the formula for the arithmetic mean:

Mean salary(~W) = 1/N ∑ salary, i = 1 to N

Here, N represents the total number of salaries, and salary denotes each individual salary.

Calculation of Required Value (X):

A population's percentage with salaries between ~W and 1.25 * ~W is represented by the needed value (X). Using the probability density function integrated within the given bounds, this is computed. The probability density function used here is based on the Gaussian (normal)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Where:

- μ is the mean of the salary distribution.
- σ is the standard deviation of the salary distribution.
- x is the variable at which the PDF is evaluated.

The integral to find the fraction of the population with salaries between ~W and 1.25 * ~W is:
Fraction = ∫f(x)dx from limits ~W to 1.25(~W). This integral is solved numerically using the integrate.quad() function from SciPy.

Results :

The code computes the mean annual salary (~W) and calculates the required value (X), representing the fraction of the population with salaries between ~W and 1.25 * ~W. The output values for mean salary (~W) and the fraction of the population with salaries between these limits are printed.