Exploring Salary Distribution

There are 4,000 observations in the dataset, with a mean salary of $32,023.36 and a standard deviation of $17,369.71.
The starting wage is $574, and the maximum is $115,330. Salary levels are separated into 25%, 50%, and 75%, with the 25th percentile at $19,238.75, the median at $29,496, and the 75th percentile at $42,378.25.

The salary distribution in the plot is skewed towards the lower end with more salaries falling in that range than the higher end. The median wage is likely lower than the mean salary. The distribution appears to be approximately uniform, with the mean wage being the central point. However, there are a few high salaries that are outliers, and they create a long tail in the distribution. As a result, the salary range is quite broad.

The mean salary is calculated by dividing the total salary values in the 'Salaries' column of the DataFrame by the total number of observations, assigning the result to the variable mean salary.

The mean value ($\widetilde{W}$) is calculated using the formula:

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

In this formula:

- X_i represents individual salary values.
- $\sum_{i=1}^{n}$ denotes the sum across all observations.
- n is the total number of observations.

Mean salary = 32023.35

The calculation of the salary value (X) below which 33% have a salary is done using the Cumulative Distribution Function (CDF) of the normal distribution. The formula is given by the inverse of the CDF, the Quantile Function.

$$X = \text{norm.ppf}(0.33, \mu, \sigma)$$

Here:

- $\text{norm.ppf}(0.33, \mu, \sigma)$ represents the quantile function of the normal distribution.
- 0.33 is the probability below the threshold.
- μ is the mean of the distribution.
- σ is the standard deviation of the distribution.

Percentile_33_salary = 24382.19