Short Report on Evaluation of Dataset

Description of Data

The dataset over which analysis needs to be conducted comprises information of annual salary belonging to the European nations and it is represented in a statistical manner. These data values have been stored within a CSV file where each entry indicates the annual salary of an individual in Euros. In addition, the dataset has one column and 4000 rows conveying a large volume of information depicting the European salaries.

Description of Distribution

Based on the analysis of the distribution for the given dataset, the empirical probability density functions were visually represented in the form of histogram. It exhibited the frequency of various ranges of wages facilitating insights as a shape as well as attributes of income distribution.

Procedure to Calculate Mean value and its Result

The following formula has been used to calculate the mean value is

Bar (x) = 1/n $\sum^{n}_{i=1} x_i$

The bar ‘x’ mentioned above conveys the mean salary, \(n\) is considered as number of observations and x_i indicates every person’s wages. With the calculated mean value, the central tendency of the income distribution can be recognized.

The result obtained as mean salary value as 26847.65

Procedure to Calculate required value X and it’s Result

The needed value X indicates the division of population by salaries between W and 1.25W using the following formula:

X= Number of Salaries between W and 1.25W / Total number of Salaries

It provides us with perception to understand the proportion of individuals that salaries come under the range with respect to the mean.

This result of the calculation of value X is 33559.56.

Summary

The analysis of dataset with exploratory method exploited a vital insight toward the income distribution of the residence of European countries. The visualization we have obtained in the form of histogram as well as computed statistics along with mean salary and value of x, develops a strong understanding over the dataset.