**Short Report**

**Name: Hemanth Tak**

**Student Id:22073706**

GitHub Repository: https://github.com/Hemanthtak2000/Codingproject

**Description of the Given Dataset**

Forty thousand rows of a single field make it a very huge and complex dataset because it is difficult to assume the conclusive message or determine specific data values from this given big data. Each data value provides information about an individual salary. Therefore, the big data can be interpreted as total wages of the population residing across the different nations of the European continent. All this information of wages is stored in the CSV files. If the data type is considered for this dataset, there is only one data type called integer justifying its uniformity over adhering to a single data type. Hence, a statistical representation can be expected from this dataset while describing the outcome.

**Description of the Gained Distribution**

Thefrequency of different wages earned by the people residing in the various countries of the European continent gained from the obtained distribution with the support of probability density. The resultant distribution has been found in the form of histogram diagram where the x-axis represents the salary ranges earned by the European population and the values along the y-axis indicates the probability densities with the range of 0 to 50. In addition, the ranges of salaries have been kept from 0 to 20000. It is a known fact that probability density is used for calculating the frequency of the corresponding value. Based on the Histogram visualization, the probability density has reached its maximum level when it ranges from 0 to 20000 Euros and the height of bars gradually declines.

**Procedure to Calculate the Mean Value and its Outcome**

The procedure to calculate the mean value of all the total wages of the population belonging to the European continent starts from a given dataset. The given data facilitated with values that can be substituted in the following formula:

Bar (x) = 1/n ∑ni=1 xi

where Mean Salary is indicated by Bar (x); number of observations is indicated by ‘n’ and individual salary is conveyed by xi.

Hence, the Mean value is 26847.65.

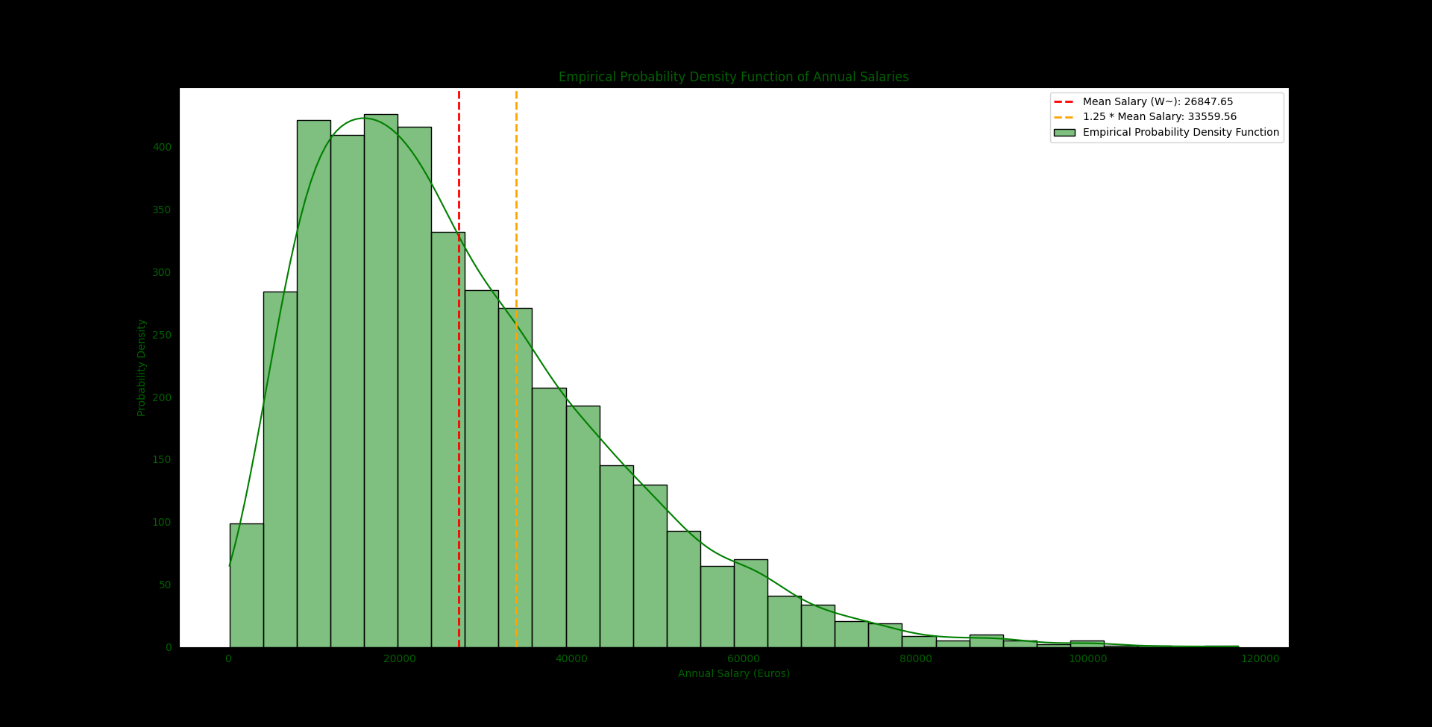
**Procedure to Calculate the Required Value X and its Outcome**

Firstly, the numbers of salaried employees have been determined in the range of W and 1.25W. The calculations of total wages were completed as its second step and the following formula has been implicated below:

X = Salaried employees between W and 1.25W/ Total Wages

We got, X= 33559.56

**Outcome as Histogram Diagram**

****