# Coding Project : Probability Distribution Function Analysis

**Name: Saman Qayyum**
**Student ID: 23010599**

**Data Description:** The provided data is extracted from a CSV file containing the information about salaries (Euros) in a European country. Each data point represents an individual's salary. The dataset is read and processed to analyze the probability distribution of salaries.

**Distribution Analysis:** The probability distribution is visualized by a probability distribution function (PDF). The distribution reveals the insights into salary structure, providing a clear overview of how salaries are distributed across the population. In order to make the PDF, bin width is calculated using Freedman-Diaconis rule:

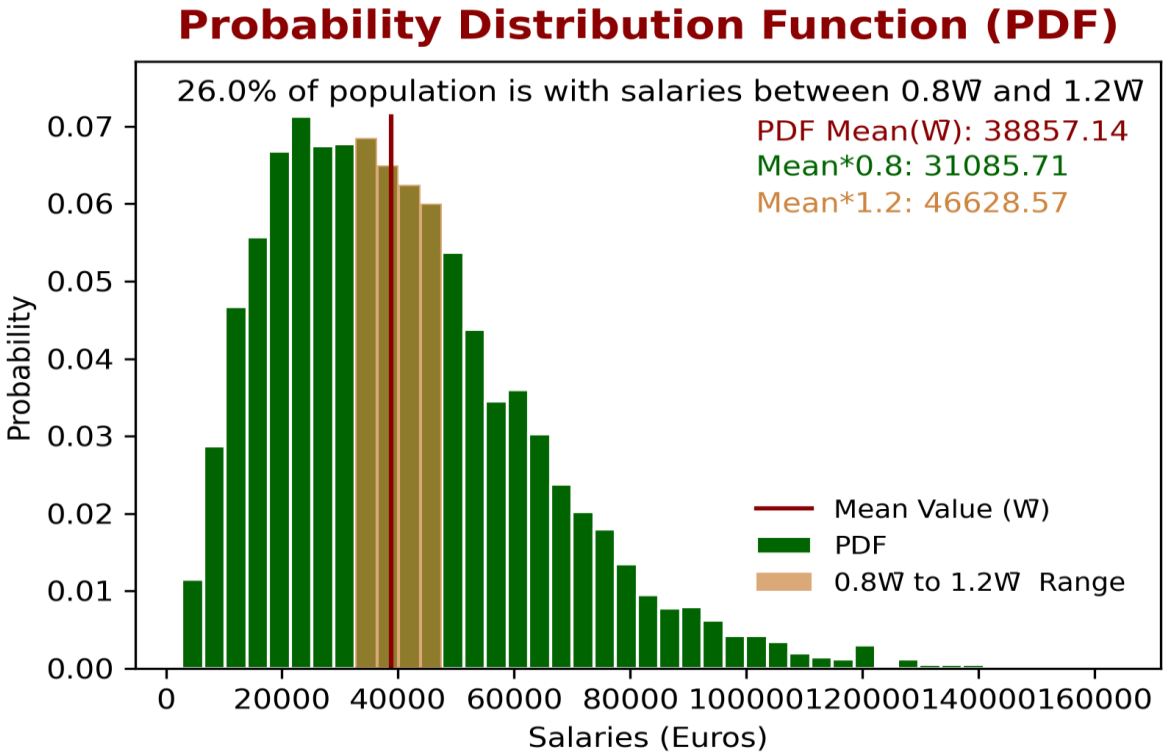
$$\text{bin width (binw)} = \frac{2(iqr)}{n^{1/3}}$$

Interquartile range (iqr) is the difference between the 75th and 25th percentiles of the dataset where n is the number of points in the dataset.

$$\text{bin size (bins)} = \frac{max.value - min.value\ (from\ dataset)}{bin\ width} = 43$$

$$\text{bin center location (xdst)} = \frac{lower\ boundry + upper\ boundry\ (of\ bin\ edge)}{2}$$

$$\text{PDF (ydist)} = \frac{no.of\ entries\ in\ each\ bin}{sum\ of\ all\ no.of\ entries\ in\ each\ bin}$$



**Mean Value Calculation:** The PDF mean annual salary of the population is calculated as the weighted average of salary values. This is achieved by summing the product of each salary and its corresponding probability in the distribution. The resulting mean value (W̄) is a representative measure of central tendency of the salary distribution.

$$\bar{W} = \sum x_{dist} \, . \, y_{dist}$$

W̄ is the mean annual salary for PDF, $x_{dist}$ is bin center location and $y_{dist}$ is discrete PDF.

**PDF Mean Annual Salary = 38857.14 €**

**Required Value X Calculation**: A specific value X, which is the fraction of population with salaries between 0.8 W̄ and 1.2 W̄ is calculated. This range provide insights into the distribution of salaries whose values are:

$$X_{0.8} = \bar{W} * 0.8 = 31085.71$$

$$X_{1.2} = \bar{W} * 1.2 = 46628.57$$

The fraction of population with salaries between 0.8 W̄ and 1.2 W̄ is calculated by,

$$X = (\sum h_{dist}) * 100$$

$$h_{dist} = y_{dist} * (x_{dist} \geq \bar{W} * 0.8) \& (x_{dist} \leq \bar{W} * 1.2)$$

Where, $h_{dist}$ is a modified probability distribution in which the probabilities outside the range (0.8 W̄ and 1.2 W̄) are set as zero.

**The fraction of population with salaries between 0.8W̄ and 1.2W̄ = 26.0 %.** This provides a qualitative measure expressed as a percentage, highlighting the portion of the population within the specified range, also it implies that 26% of the population have the salaries between $X_{0.8}$ and $X_{1.2}$ values.

**Comment:**

This analysis and information is useful for making decisions and understanding how salaries vary in the country. It gives us insights that can help us for planning and making policies that affect people's incomes.