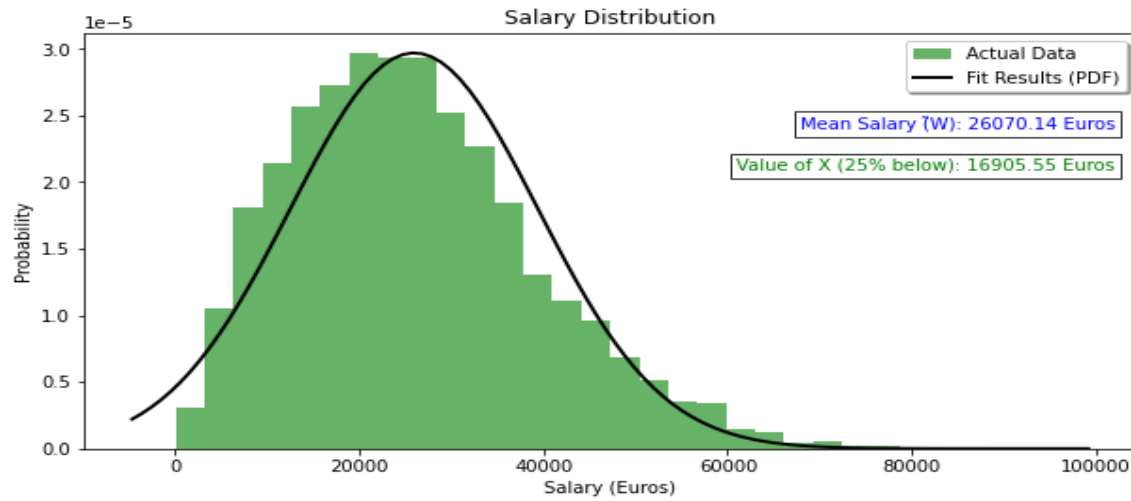# Fundamentals of Data Science

Probability Density Function

Student ID: 22066704

About the Data:

The dataset we're working with holds salary figures in a European country, from a file named 'data4-2_1575043522.csv.'

2. Picture of the Numbers:

The graph above provides a visual representation of how salaries are dispersed and structured. The resulting histogram, obtained from the probability density function (PDF), takes on the classic shape of a bell curve. The distinctive bell curve shape observed in the histogram signifies a symmetric and balanced distribution of salaries. It indicates that a significant portion of individuals earns salaries close to the average, with fewer extremes at both ends.

3. Figuring Out the Average:

The mean value "mu" is calculated using the `norm.fit()` function, which estimates the mean of the normal distribution fitted to the provided salary data. In the context of the code, it's achieved with the following snippet:

```
mu, std = norm.fit(data['salary'], loc=0)
```

Here, `mu` represents the estimated mean of the normal distribution. In simpler terms, it's the average salary value around which the data is centered.

The initial mean value obtained from the data is (25965.34) Euros, rounded to two decimal places.

```
Mean Value Intially
25965.3385
Standard Deviation
13432.069174606635

In [30]:
```

4. Finding the Right Number:

The required value (X) is calculated using the `norm.ppf()` function, which is particularly useful for finding the value associated with a specific percentile point in a normal distribution. In the context of your code, \(X\) is determined with the following line:

Calculate the value of X for 25% of people above X

```
X = norm.ppf(0.25, mu, std)
```

This code finds the salary value ((X)) below which 25% of individuals fall. Mathematically, it's represented as (**X = ppf(0.25, mu, sigma)**).

The calculated value for (X) is (**16905.55**) Euros, rounded to two decimal places.

Assignment GitHub Link:
https://github.com/sanaullahgithub/Fundamental-of-data-science.git