Name: Shubham Verma    SRN: 22099668    Email: *sv23abk@herts.ac.uk*    Course: MSc Data Science
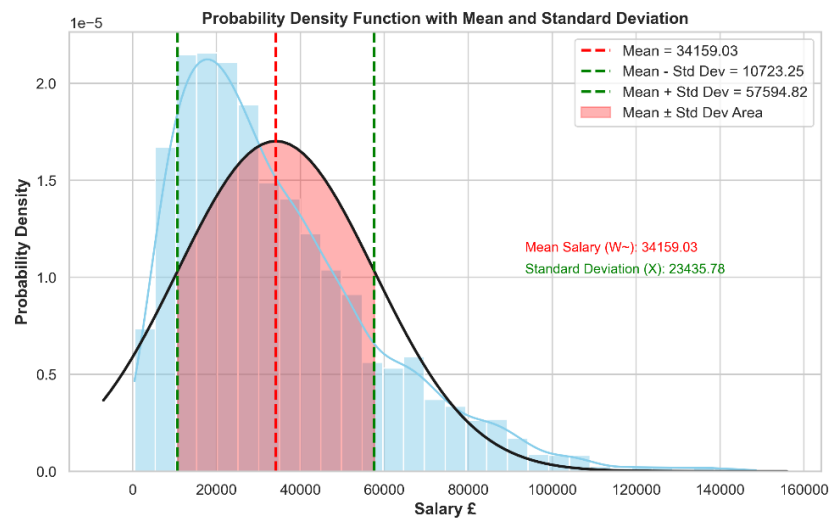
# Fundamentals of Data Science: Coding Project



### Description of data

The given information is the probability density function (PDF) of the annual salary expressed in euros in a European country. The PDF shows the probability that the salary will fall in a certain range. The average salary in the centre of the PDF is 34,159.03. The standard deviation of the salary distribution is 23,435.78.

### Description of the distribution

The distribution of annual salary data is bell-shaped and symmetrical. This means that most salaries are close to the mean salary and fewer salaries are far from the mean salary. The distribution is also known as the normal distribution.

### Calculation of the mean value

The mean of salaries is the average of all salaries in the data set. It is calculated by adding up all the salaries and dividing by the number of salaries.

mean = (sum of all salaries) / (number of salaries)

The mean value we get from our data is, **Mean Salary (W~): 34159.03**

### Calculation of the standard deviation value

X is the standard deviation of salaries, it represents the typical amount by which salaries deviate from the mean salary. In other words, it measures the spread of the salary distribution around the mean value. A higher standard deviation indicates that salaries are more dispersed, while a lower standard deviation suggests that salaries are more concentrated around the mean.

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

#### where:

$x_i$ = Value of the $i^{th}$ point in the data set

$x$ = The mean value of the data set

$n$ = The number of data points in the data set

The standard deviation value we get from our data is, **Standard Deviation (X): 23435.78**

***GitHub Links:*** Repository, Data file location, Source code