Fundamentals of Data Science

Coding Assignment

Name: Nikhil Dinesh Soni

Student Id: 22045846

Github repository link: https://github.com/Nikhilsxni/FDS-Coding-Assignment

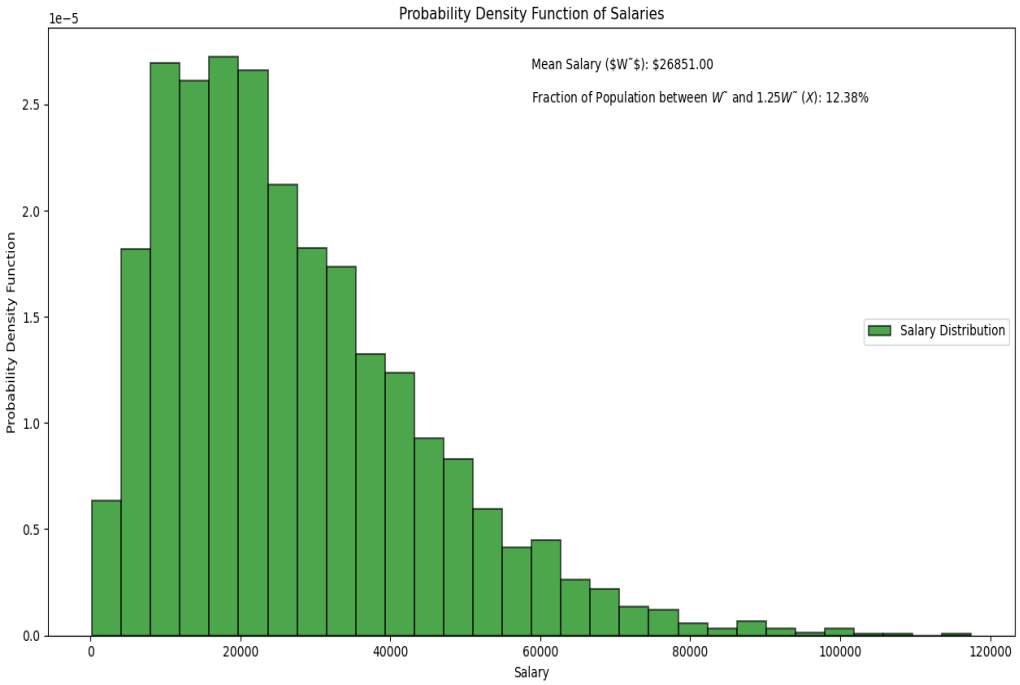

Figure:1

The given data is the salary of some European countries ranges from 0 to 120000 and the **figure 1** represents the probability density function of salaries in a histogram plot. the x-axis represents the range of salaries and the y-axis projects the probability density function (PDF). the PDF is used to specify the probability of the random variable falling within a particular range of values, as opposed to taking on any one value.

There's a sudden incline in the histogram which states that most fraction of population earns around 20000 at 2.5 and the slow decline after 0.5 shows that only small scale of population earns more than 60000. Only some of the population earns 100000, the pdf is less than 0.5.

The mean of the probability density function can be given as $\int_{\infty-\infty} xf(x)dx$. the mean value for this PDF is 26851.

The following part of the code calculates the required value x which is the fraction of population with salaries between W̃ and 1.25W̃ (where W̃ is the mean value in your distribution) salary data is the given salary distribution on which the graph is based on.

```
# Calculate the fraction of population with salaries between W̃ and 1.25W̃ (X)

lower_bound = mean_salary

upper_bound = 1.25 * mean_salary

fraction_population = np.sum((salary_data >= lower_bound) & (

    salary_data <= upper_bound)) / len(salary_data)
```

the fraction of population with salaries between W̃ and 1.25W̃ (X) is 12.38%.