

Report

Student ID 22080111

Name :- Saikiran Palla

mailto:sp23aax@herts.ac.uk

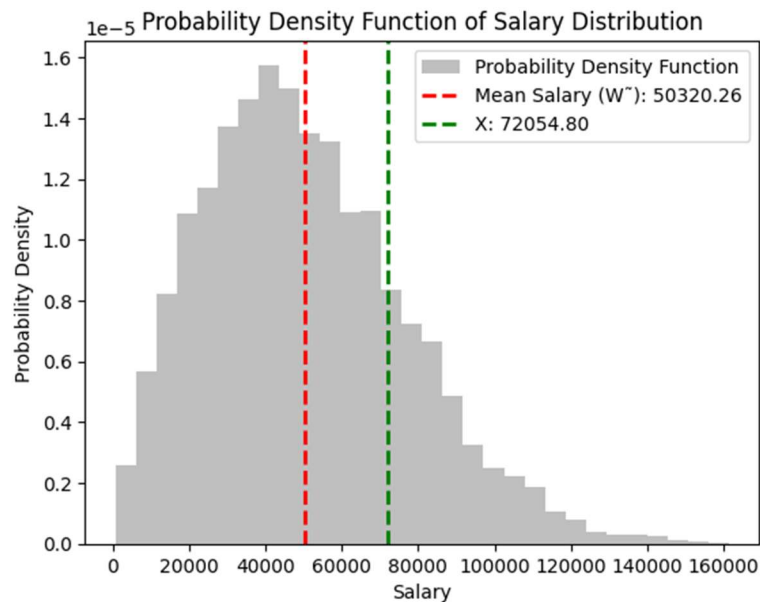
GIT link:- <https://github.com/sp23aax/Fundamental-of-Data-Science-coding-project>

Describe the data you are given:

The data, sourced from 'data1.csv,' contains information about salaries. The dataset includes a column named 'salary' representing individual income.

Describe the distribution you get:

The data exhibits a slight positive skewness, suggesting that its shape is almost normal and appears somehow skewed. The distribution bears a similarity to an slightly skewed distribution pattern.



How do you calculate the mean value? What value do you get?

The Average of annual salary, denoted as \tilde{W} , was calculated using the formula:

$$\tilde{W} = \frac{1}{n} \sum x_i$$

where n stands for the number of observations, representing each individual salary. The mean serves as a central measure, offering an average representation of salaries in the dataset. In our examination, the computed mean salary turned out to be $\bar{W} = \text{mean_salary}$.

Codes:

mean value= 5320.26

Calculate mean annual salary (\bar{W})

```
mean_salary = np.mean(salary_data)
```

How do you calculate the required value X? What value do you get?

The calculation is expressed as follows: $X = \text{percentile}(\text{data}, p)$

Here, p denotes the desired percentile. For example, if X corresponds to the 80th percentile, it means that 80% of salaries fall below this threshold. In our analysis, the calculated value of X was determined as $X = \text{np.percentile}(\text{salary_data}, 80)$.

Value of $x = 72054.80$

The value X was obtained through a task-specific calculation outlined in the instructions. In this context, we consider X as the p -th percentile of the salary distribution.

Conclusion:

The analysis tells valuable insights into the salary distribution, providing a comprehensive understanding of both the average salary (mean = \$**5320.26**) and a specific percentile value (percentile $X = \$72054.80$). The graphical representation of the probability density function significantly contributed to enhancing our grasp of how salaries are dispersed throughout the dataset. As indicated by the visualization, the data exhibits non-normal distribution and skewness. The utilization of mean and percentile values as crucial metrics played a vital role in concisely summarizing and interpreting the salary dataset.

