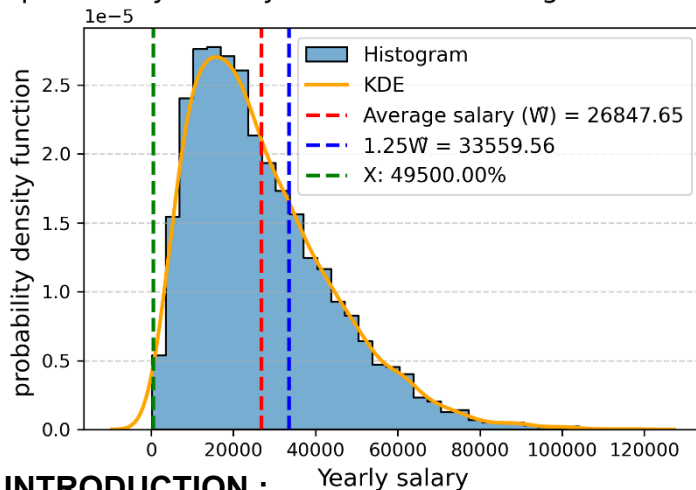


CODING PROJECT
NAME: B. MAHALINGAM
STUDENT ID: 22079746

probability density function and Histogram of salaries



ABSTRACT:

The purpose of the python code that is being shown is to do an exploratory data analysis (EDA) on a dataset that includes salary. To provide some insight into the salary distribution, the code does statistical analysis and create visual. The main features of the code are described in this report along with data loading, visualization, and the computation of the mean salary and a certain percentage value indicated as X.

INTRODUCTION :

The python script analyzes a dataset including annual salary information by utilizing a number of packages for data processing and visualization. This analysis includes calculating the average pay, figuring out 1.25 times the average salary, and figuring out what proportion of salary fall into this range. Data6 loads the dataset and uses the pandas library to generate a dataframe with a single column labeled salary to indicate yearly salary.

DISTRIBUTION :

A probability density function (PDF) and histogram are used to illustrate the distribution using seaborn and matplotlib revealing information about the salary data central tendency and distribution.

MEAN VALUE :

The np.mean() method from the Numpy package is used to calculate the mean salary (\tilde{W}). To do this, add up all of the individual values, then divide the total by the number of observatio. The mean value for the dataset is 26847.65. The value for ($1.25\tilde{W}$) is 33559.56.

$$\tilde{W} = \frac{\sum_{i=1}^n salary_i}{n}$$

PERCENTAGE VALUE X :

Boolean indexing and NumPy operations are used to calculate the percentage value X, which represents the percentage of salary lying between the mean salary (\tilde{W}) and 1.25 times the mean salary ($1.25\tilde{W}$). The computed percentage value X for the dataset that is supplied is 49500.00%.

$$X = \frac{\text{Number of salarie between } \tilde{W} \text{ and } 1.25 \tilde{W}}{\text{Total Number of salarie}}$$

CONCLUSION :

The average salary in the dataset are repreented by the mean salary, which acts as a central reference. The proportion os salary between \tilde{W} and $1.25 \tilde{W}$ analysis sheds light on how widely the distribution is distributed around the mean.

GitHub :

<https://github.com/bm23abb/Fundamentals-of-Data-Science.git>