**ANNUAL SALARY DISTRIBUTION**

A graph of a salary distribution

Description automatically generated

**Describe the data:**

The given data is stored in a CSV file named 'data4-2.csv.' It is loaded into a Pandas Data Frame (df) with a single column labelled 'salary.' The sample data variable is assigned the values from this 'salary' column.

**Describe the distribution you get:**

The distribution of the data is visualized using a histogram. The histogram is created with 30 bins, normalized (density=True), and has a transparency of 0.5. The color of the bars is set to 'dark cyan.' This allows you to observe the frequency distribution of salary values in the dataset.

**Calculate the mean value:**

The mean value of the salary data is calculated using NumPy's np.mean function. The result is stored in the variable mean\_salary. This mean value represents the average annual salary in the dataset.

**Calculate the required value X:**

The value 'X' is calculated using the norm.ppf function from the SciPy library. Specifically, it calculates the value of 'X' such that 25% of the data falls below it in a standard normal distribution. The parameters passed to norm.ppf are the desired quantile (0.25 for 25%), the mean (mu), and the standard deviation (std) calculated from the dataset. The resulting value of 'X' is stored in the variable X. This value represents the threshold below which 25% of the salaries in the dataset lie.

**Result:**

In summary, this code is performing exploratory data analysis on salary data, visualizing its distribution, calculating the mean salary, and determining a threshold value 'X' below which 25% of salaries fall. The results are visualized in a histogram with additional annotations for the mean and 'X' values.