Analysis of Annual Salary Data

NAME: ASHIKA MOHAN MUNGATH ID:22063061

Data Description:

The dataset comprises of annual salary information, where each entry in the file shows one annual salary in Euros. The analysis seeks to examine the salary distribution, compute the mean (represented as $\tilde{W}$), and identify a particular value, X. (ie, 33% of people have a salary above X)
A histogram with 30 bins is employed to visually represent the distribution, accompanied by the addition of a fitted normal distribution curve. The observed distribution exhibits a noticeable positive skewness.

Distribution Description:

Median:
The median represents the middle value of the dataset. The calculated middle value of the dataset is €47,241.5.

Mode:
The mode is the most frequently occurring salary value. In this dataset the most frequently occurring salary is €35,049.

Standard Deviation (Std):
The standard deviation of approximately €26,314.45 indicates the extent of variability in the salary data. It is calculated by the equation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Variance:
The square of the standard deviation, the variance, is approximately €692,277,224.05.

Kurtosis:
With a positive kurtosis of 0.1925, the distribution has slightly heavier tails than a normal distribution, indicating a moderate degree of peakedness.

Skewness:
The positive skewness of 0.6017 suggests that the distribution is skewed to the right, with a tail extending towards higher salary values.

Interquartile Range (IQR):
The IQR, representing the range between the 25th and 75th percentiles, is €36,144.0. It provides insights into the spread of the central portion of the data.

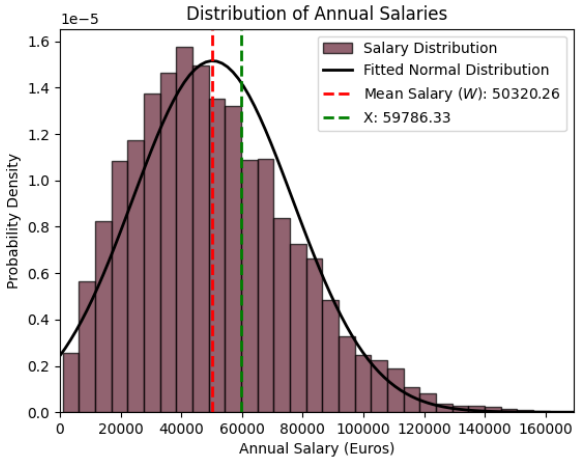
Range of Data:
The overall range of salary values, from the minimum to the maximum, is €160,394.

Mean Calculation:
The mean annual salary is calculated as the average of all salary values. The mean annual salary for the dataset is calculated as €50,320.26.
The mean (μ) of the salary data is calculated using the NumPy library's np. mean function.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where N is the number of data points and x_i is the i-th salary value.

Value X Calculation:
To find the required value X, representing a specific percentile (67th percentile in this case), we use the formula for the percentile:
X=Percentile(X)
X is set to the 67th percentile, ensuring that 33% of people have a salary above X (since 100−67=33). Here the observed x value is 59786.33



Conclusion:
The mean annual salary of €50,320.26 provides a central reference point, signifying the dataset's average. Concurrently, the calculated X value at the 67th percentile, €59,786.33, delineates the salary threshold above which 33% of individuals lie. This analysis furnishes valuable insights into both the central tendency and a specific point of significance within the dataset, contributing to a nuanced comprehension of the salary distribution.The distribution is positively skewed, suggesting a concentration of lower salary values and a tail extending towards higher salaries.
The mean, median, and mode provide a central tendency overview, while the standard deviation and variance offer insights into the spread of the data.
The analysis offers a thorough comprehension of the annual salary distribution, contributing valuable insights for decision-making and future exploration.