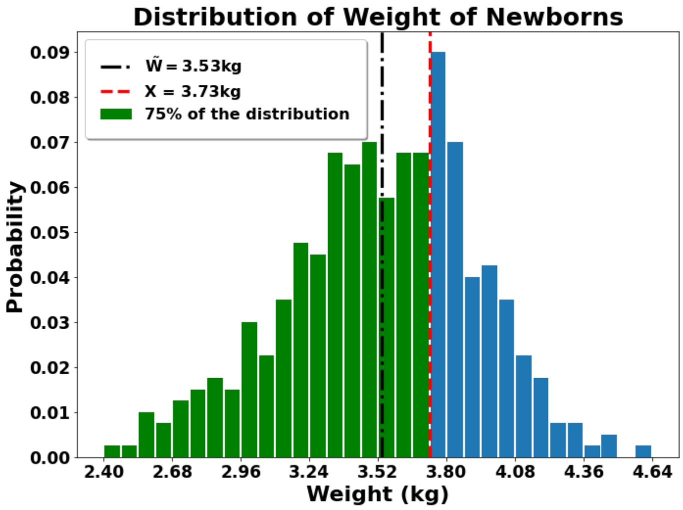
CODING PROJECT REPORT

Name: Aaron Modiyil Joseph

Student ID: 22018497

The dataset given contains 400 entries of weights of new-borns in certain locations of Europe over a given time period. This collection of data provides information regarding the weight distribution of new-borns in these regions and allows for potential analysis and comparison between them.



When the distribution is visualized as a histogram, we can observe that the distribution has a bell-shaped curve, with a majority of the data being clustered towards the right and less on the left. Therefore, we can conclude that the distribution follows a left skewed normal distribution or a negatively skewed normal distribution. Hence, for this distribution, the mean is lower than both the median and the mode.

We can see that this is a discrete probability density function. Hence, to calculate the mean, we can use the formula: Mean, $\mu = \sum_{k=-\infty}^{+\infty} x_k f(x_k)$. Here, for this project, the x_k are the X values we considered for the histogram and $f(x_k)$ are the corresponding Y values of the histogram. Therefore, by using numpy' s sum function to the array formed by multiplying the X values with the corresponding Y values of the histogram we can determine the average weight of the new-born babies, W̃.

The calculated average weight, W̃ = 3.53kg

To calculate the value X that corresponds to the point where 75% of the distribution is below X, we first calculate the cumulative distribution of the normalized distribution we obtained from the histogram. We then find the index of the value in the cumulative distribution that is closest to 0.75. This is done by taking the absolute difference between the cumulative distribution and 0.75 and finding the index of the minimum value of this difference. Finally, we use the index value to obtain the corresponding edge value of the histogram which gives us the value of X such that 0.75 of the distribution are below X. This value of X is the weight at which 75% of new-borns from the distribution are born with a weight below X.

The calculated value such that 75% of the given data are below it, X = 3.73kg