**7PAM2020-0105-2022- Fundamentals of Data Science**

**Coding Project**

**Student ID:** 21066861

**Student Name:** Sanshiya Rameshkumar

**Link to dataset**: https://github.com/Sanshiya-23/Fundamentals-of-Data-Science-Coding-Project-.git

**Fundamentals of Data Science – Coding Project**

**About Data:**

* The purpose of the code is to visualize the distribution of newborn weights and to highlight the sample mean and a specific x value. According to my dataset (data1), the value of X should be such that 33% of new- borns from the distribution are born with a weight above X. It shows a histogram of the data with 20 bins, normalizes the distribution. It plots the normalized distribution as a bar chart and adds vertical lines to indicate the sample mean and a specified x value. The code also includes text annotations to display the mean and x value. The plot title is "Distribution of Newborn Weights" and the x-axis label is "Weight (in pounds)".

**About distribution**

* The distribution of newborn weights appears to be approximately bell-shaped, with a peak around 7 pounds and a spread of weights ranging from around 4 to 10 pounds. The distribution is roughly symmetric, with a slightly longer tail on the right-hand side of the distribution. The histogram shows that the majority of newborns weigh between 6 and 8 pounds. The discrete probability density function (PDF) is shown by the purple bars and the cumulative distribution function (CDF) is shown by the green line. The mean value of the weight distribution is calculated to be 7.026 pounds. The X value represents the weight at which 33% of the data is below this value and 67% of the data is above this value. The X value is calculated to be 6.533 pounds, which means that 33% of newborns weigh less than 6.533 pounds and 67% of newborns weigh more than 6.533 pounds.

**Mean Weight Calculation**

* The mean weight is calculated using the formula:

**xmean = np.sum(xdst\*ydst)**

where xdst is the bin center locations and ydst is the normalized distribution. The value of the mean weight is approximately 7.288 pounds, rounded to three decimal places. This value is shown in the plot with the label "Sample Mean" and a yellow dashed line.

**X Value Calculation**

* The value of X is the weight value on the x-axis that corresponds to the cumulative distribution of 0.67. In the code, we find the index of the element in the cdst array that is closest to 0.67 using np.argmin(np.abs(cdst-0.67)). Then we use this index to find the corresponding weight value in the oedge array, which is assigned to xhigh.

In this case, the value of X is approximately 7.053 pounds (rounded to 3 decimal places).