1. Let us consider a random experiment of rolling a die variable number of times (each die roll can be denoted as X_1, X_2, …, X_{50}) ranging from 1 to 50 and collecting their sum (No need to collect the sum when you roll the die once). Repeat the experiments 100, 500, 1000, 5000, 10000, 50000, and 100000 times. Plot and save the histograms of each (For example: you should have 6 histograms for one die roll, 6 for 2 dice roll, 6 for 3 dice roll, and so on). Find the mean and variance for each case.
2. Consider a binomial distribution. Assume that a product is either rated positive or negative in an ecommerce website. You got 97 positive reviews out of 100 for a product. Construct a binomial distribution with different success rates to find out an estimated success rate for the product. Use histograms to assert your answer.
3. For each of the datasets you have used so far (3 datasets), find out the entropy for each column. [Hint: If it is a categorical variable, then it is straight forward. If it is a continuous value, try binning it and then find entropy.]