BDA2321: Machine Learning Activity Report

Reg. no: 222BDA43

Dataset:https://www.kaggle.com/datasets/PromptCloudHQ/flipkart-products?resource=download

The flipkart dataset contains information of its product name, product details and the cost of the product. The dataset contains 20000 rows × 15 columns .

Steps followed in pre-processing of the dataset as follows:

- ❖ **Downloaded the dataset from Kaggle.**
- ❖ **Imported the flipkart dataset into the workspace.**
- ❖ **Identified the missing values in each column.**
- ❖ **Dropped the unwanted columns.**
- ❖ **Checked for the total numbers of unique brands.**
- ❖ **Identified the missing values and replaced them with NaN values.**
- ❖ **Dropped all the NaN values.**
- ❖ **Plotted the histogram for retail_price.**
- ❖ **Plotted the scatter plot between the retail_price and discounted_price.**

In [3]: