Set 1

1. Analyze dataset (forestfires.csv) predict the burned area of forest fires and build linear regression model using python

   a. Identify missing values and fill it with the mean.
   b. Find out how the data is skewed and identify the kurtosis value
   c. Identify the outliers and plot it.
   d. Find out categorical and numerical data and visualize it.
   e. Prepare the data for modelling.
   f. Build a linear regression model using statistical and machine learning approach.
   g. Train the model
   h. Predict the burned area using the model
   i. Analyze the performance of the model and visualize it.

2. Create product table with the following schema using MongoDB and implement the following:
Product(pid,pname,ptype,price,manufacturer,stockqty,tax{CGST,SCGST})
Note: Insert 10 records.

   a. Display all the products which has price greater than Rs.500/-
   b. Update the tax field based on the tax structure issued by Govenrment
   c. Display the total count of the product based on the type of product.
   d. Delete the product if the manufacturer of the product stops its production.
   e. Update the stock count of the product if more products are delivered
   f. Perform data replication on “Retail management DB”. If needed along with product table, you can have create other supporting tables.

3. Write the steps and install single node hadoop .
   Mention the components of Openstack.