**Prerequisites:** A VM instance with Hadoop cluster with Hive installed on it

**Step 1: *Load data into hive***

**After logging in, clone the code and required files**

>> cd ~

>> git clone https://github.com/Mahi230701/visualizationapplication.git

**Connect to hive**

>> hive

**Create table in hive with name sales**

hive>> CREATE TABLE sales (

invoice\_id STRING,

branch STRING,

city STRING,

customer\_type STRING,

gender STRING,

product\_line STRING,

unit\_price FLOAT,

quantity INT,

tax FLOAT,

total FLOAT,

purchase\_date STRING,

purchase\_time STRING,

payment STRING,

cogs FLOAT,

gross\_margin\_percentage FLOAT,

gross\_income FLOAT,

rating FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**View tables in the default database**

hive>> show tables;

**Load csv data into hive table**

hive>> LOAD DATA LOCAL INPATH '/root/visualizationapplication.git/dataset/sales.csv' OVERWRITE INTO TABLE sales;

**Query the data and verify**

hive>> select \* from sales;

**Come out of hive shell**

hive>> exit;

**Step 2: *Run the interactive Visualizer application***

**Setup libraries**

>> pip3 install flask, plotly

>> sudo apt install unixodbc-dev

>> sudo apt-get install libsasl2-dev

>> pip3 install sasl

**Run the application**

>> cd /root/visualizationapplication.git/code

>> python app.py

**Access the application on browser with ip:5000**