In [1]:
```
# importing all the necessary liberaries
import pymongo
import sys
import pandas as pd
import json
import csv
import os
import seaborn as sns
from scipy import stats
```

In [2]:
```
# all the variables
csv_file_path = 'Walmart.csv'
json_file_path ='Walmart.json'
password='satish321'
connection_string=f'mongodb+srv://satishkandel198:{password}@cluster0.odylp4l.mongodb.net/?retryWrites=true&w=ma
db_name='WALMART'
my_collection='sales_data'
```

In [3]:
```
# create function to load csv file and create a json file
def create_json(csv_file,json_file):
    '''
    this function will take two parameter:
    1. csv_file location from where you want to import  the file
    2. Json file path along with the name where json file will be created
    '''

    # Read CSV file and convert to list of dictionaries
    csv_data = []
    with open(csv_file_path, 'r') as csv_file:
        csv_reader = csv.DictReader(csv_file)
        for row in csv_reader:
            csv_data.append(row)

    # Write JSON file
    with open(json_file_path, 'w') as json_file:
        json.dump(csv_data, json_file, indent=4)

    print(f'Conversion complete. JSON file saved at {json_file_path}')
```

In [4]:
```
# create a function that returns the database connection
def get_connection(connection_string):
    try:
        client=pymongo.MongoClient(connection_string)
        return client
    except Exception as e:
        return  e
```

In [5]:
```
db_name='WALMART'
'''
this function will check if json file is already created or not, if the json file is created it first
creates the json file then it checks the databse if there is a data, if there is already data in the
database then, it does not insert data into the database, else it will insert data into the database.
'''
def insert_data(csv_file,json_file):
    '''
    json_file: the document which will be loaded into the mongodb database
    '''
    csv_file_path=csv_file
    json_file_path=json_file
    # check if the json file is already created or not, if created directly load and insert into the database e
    if os.path.exists(json_file_path):
        js_file=json.load(open(json_file_path))
    else:
        create_json(csv_file_path,json_file_path)
        js_file=json.load(open(json_file_path))

    # get the connection
    client=get_connection(connection_string)
    print(client)
    if client:
        try:
            mydb = client[db_name]
            collection = mydb[my_collection]

            if collection.count_documents({}) > 0:
                print('Data is already loaded.')
            else:
                collection.insert_many(js_file)
                print("Data inserted successfully")

        except Exception as e:
```

```
            print(f"Error accessing database or collection: {e}")

    else:
        print("Connection to MongoDB failed.")
```

In [6]:
```
# function_call
insert_data(csv_file_path,json_file_path)
```

```
MongoClient(host=['ac-ny1yb7u-shard-00-00.odylp4l.mongodb.net:27017', 'ac-ny1yb7u-shard-00-01.odylp4l.mongodb.ne
t:27017', 'ac-ny1yb7u-shard-00-02.odylp4l.mongodb.net:27017'], document_class=dict, tz_aware=False, connect=True
, retrywrites=True, w='majority', authsource='admin', replicaset='atlas-vd77c7-shard-0', tls=True)
Data inserted successfully
```

We have successfully inserted data into the mongodata server following the following steps:

1. Download the csv file from the kaggle.
2. import the csv file and create a new json file.
3. connect with the mongodb database(WALMART).
4. Insert the json document into the sales_data collection

Next step: Data cleaning

In [ ]: