In [3]:

```
#import libraries
#!pip install kaggle
import kaggle
```

In [4]:

```
#Downloading dataset from kaggle
!kaggle datasets download ankitbansal06/retail-orders -f orders.csv
```

```
Warning: Looks like you're using an outdated API Version, please consider updating (serve
r 1.6.17 / client 1.6.14)
Dataset URL: https://www.kaggle.com/datasets/ankitbansal06/retail-orders (https://www.kag
gle.com/datasets/ankitbansal06/retail-orders)
License(s): CC0-1.0
orders.csv.zip: Skipping, found more recently modified local copy (use --force to force d
ownload)
```

In [5]:

```
# Extracting file from zip file
import zipfile
zip_ref=zipfile.ZipFile('orders.csv.zip')
#Extracting file to directory
zip_ref.extractall()
zip_ref.close()#closing the file
```

In [6]:

```
# Importing the Libraries
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set()

import warnings
warnings.filterwarnings('ignore')

```

In [7]:

```
df = pd.read_csv('c://Users//Chitra//orders.csv',na_values=['Not Available','Unknown']
```

In [8]:

```
df.columns=df.columns.str.replace(' ','_')
df.columns=df.columns.str.lower()
df.columns
```

Out[8]:

```
Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country', 'city',
       'state', 'postal_code', 'region', 'category', 'sub_category',
       'product_id', 'cost_price', 'list_price', 'quantity',
       'discount_percent'],
      dtype='object')
```

In [9]:

```
df['ship_mode'].unique()
```

Out[9]:

```
array(['Second Class', 'Standard Class', nan, 'unknown', 'First Class',
       'Same Day'], dtype=object)
```

In [10]:

```
df.shape
```

Out[10]:

```
(9994, 16)
```

In [11]:

```
# Deriving new columns discount, sale_price and profit
#discount= list_price x discount_percent.
df['discount']=df['list_price'] * df['discount_percent']*.01
```

In [12]:

```
df['sale_price'] = df['list_price'] - df['discount']
```

In [13]:

```
df['profit'] = df['sale_price'] - df['cost_price']
```

In [14]:

```
df.dtypes
```

Out[14]:

```
order_id              int64
order_date           object
ship_mode            object
segment              object
country              object
city                 object
state                object
postal_code           int64
region               object
category             object
sub_category         object
product_id           object
cost_price            int64
list_price            int64
quantity              int64
discount_percent      int64
discount            float64
sale_price          float64
profit              float64
dtype: object
```

In [15]:

```
df['order_date']=pd.to_datetime(df['order_date'],format="mixed")
```

In [16]:

```
df.drop(columns=['discount_percent','list_price','cost_price'],inplace=True)
```

In [17]:

```
df.columns
```

Out[17]:

```
Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country', 'city',
       'state', 'postal_code', 'region', 'category', 'sub_category',
       'product_id', 'quantity', 'discount', 'sale_price', 'profit'],
      dtype='object')
```

In [18]:

```
#Loading the dataset in to sql server
import sqlalchemy
import pyodbc, os
```

In [19]:

```
import platform
print(platform.node())
```

```
DESKTOP-84H1QFD
```

In [20]:

```
import socket
socket.gethostname()
```

Out[20]:

```
'DESKTOP-84H1QFD'
```

In [21]:

```
pyodbc.drivers()
```

Out[21]:

```
['SQL Server',
 'SQL Server Native Client 11.0',
 'ODBC Driver 11 for SQL Server',
 'Microsoft Access Driver (*.mdb, *.accdb)',
 'Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb)',
 'Microsoft Access dBASE Driver (*.dbf, *.ndx, *.mdx)',
 'Microsoft Access Text Driver (*.txt, *.csv)']
```

In [22]:

```
engine = sqlalchemy.create_engine('mssql://DESKTOP-84H1QFD\SQLEXPRESS/project?driver=O
conn = engine.connect()
```

In [110]:

```
df.to_sql('stores',con = conn,if_exists = 'append',index = False)
```

Out[110]:

```
-1
```

In [111]:

```
df.to_excel('c://Users//Chitra//Desktop//orders_analysis.xlsx',index = False)
```