```python
import numpy as np
import pandas as pd

df = pd.read_csv('SuperStore_Sales_Dataset.csv')

df.head()
```

```
   Row ID+O6G3A1:R6  Order ID  Order Date   Ship Date      Ship Mode  \
0              4918         1  01-01-2019  07-01-2019  Standard Class
1              4919         2  02-01-2019  07-01-2019  Standard Class
2              4920         3  02-01-2019  07-01-2019  Standard Class
3              3074         4  03-01-2019  05-01-2019     First Class
4              8604         5  03-01-2019  08-01-2019  Standard Class


  Customer ID      Customer Name     Segment         Country
City  ...  \
0    BM-11575       Brendan Murry   Corporate   United States
Gaithersburg  ...
1    BM-11575       Brendan Murry   Corporate   United States
Gaithersburg  ...
2    BM-11575       Brendan Murry   Corporate   United States
Gaithersburg  ...
3    LR-16915        Lena Radford    Consumer   United States   Los
Angeles  ...
4    CA-12310  Christine Abelman   Corporate   United States   San
Antonio  ...

          Category Sub-Category  \
0         Furniture    Bookcases
1         Furniture    Bookcases
2        Technology       Phones
3  Office Supplies      Storage
4        Technology  Accessories

                                        Product Name    Sales  Quantity
\
0  Bush Westfield Collection Bookcases, Medium Ch...    73.94         1

1  Bush Westfield Collection Bookcases, Medium Ch...   173.94         3

2                                       GE 30522EE2   231.98         2

3  Recycled Steel Personal File for Hanging File ...   114.46         2

4                  Imation Clip USB flash drive - 8 GB    30.08         2
```

```
       Profit  Returns  Payment Mode  ind1  ind2
0  28.2668      NaN        Online   NaN   NaN
1  38.2668      NaN        Online   NaN   NaN
2  67.2742      NaN         Cards   NaN   NaN
3  28.6150      NaN        Online   NaN   NaN
4  -5.2640      NaN        Online   NaN   NaN

[5 rows x 23 columns]
```

```python
#looking the unique values
df['Ship Mode'].unique()
```

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

```python
df.columns
```

```
Index(['Row ID+O6G3A1:R6', 'Order ID', 'Order Date', 'Ship Date',
'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'Country', 'City',
'State',
       'Region', 'Product ID', 'Category', 'Sub-Category', 'Product
Name',
       'Sales', 'Quantity', 'Profit', 'Returns', 'Payment Mode',
'ind1',
       'ind2'],
      dtype='object')
```

```python
df.columns.str.lower()
```

```
Index(['row id+o6g3a1:r6', 'order id', 'order date', 'ship date',
'ship mode',
       'customer id', 'customer name', 'segment', 'country', 'city',
'state',
       'region', 'product id', 'category', 'sub-category', 'product
name',
       'sales', 'quantity', 'profit', 'returns', 'payment mode',
'ind1',
       'ind2'],
      dtype='object')
```

```python
#converting uppercase to lowercase
df.columns = df.columns.str.lower()

df.head()
```

```
   row id+o6g3a1:r6  order id  order date   ship date        ship mode
\
0              4918         1  01-01-2019  07-01-2019  Standard Class
```

```
1                    4919          2  02-01-2019  07-01-2019  Standard Class

2                    4920          3  02-01-2019  07-01-2019  Standard Class

3                    3074          4  03-01-2019  05-01-2019     First Class

4                    8604          5  03-01-2019  08-01-2019  Standard Class


  customer id       customer name     segment          country
city  ...  \
0   BM-11575       Brendan Murry    Corporate  United States
Gaithersburg  ...
1   BM-11575       Brendan Murry    Corporate  United States
Gaithersburg  ...
2   BM-11575       Brendan Murry    Corporate  United States
Gaithersburg  ...
3   LR-16915        Lena Radford     Consumer  United States   Los
Angeles  ...
4   CA-12310  Christine Abelman    Corporate  United States   San
Antonio  ...


          category sub-category  \
0        Furniture    Bookcases
1        Furniture    Bookcases
2       Technology       Phones
3  Office Supplies      Storage
4       Technology  Accessories


                                       product name    sales quantity
\
0  Bush Westfield Collection Bookcases, Medium Ch...   73.94        1

1  Bush Westfield Collection Bookcases, Medium Ch...  173.94        3

2                                       GE 30522EE2  231.98        2

3  Recycled Steel Personal File for Hanging File ...  114.46        2

4                  Imation Clip USB flash drive - 8 GB   30.08        2


    profit  returns  payment mode  ind1  ind2
0  28.2668      NaN        Online   NaN   NaN
1  38.2668      NaN        Online   NaN   NaN
2  67.2742      NaN         Cards   NaN   NaN
3  28.6150      NaN        Online   NaN   NaN
4  -5.2640      NaN        Online   NaN   NaN

[5 rows x 23 columns]
```

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

df.columns = df.columns.str.replace('-','_')

df.columns

Index(['row_id+o6g3a1:r6', 'order_id', 'order_date', 'ship_date',
'ship_mode',
       'customer_id', 'customer_name', 'segment', 'country', 'city',
'state',
       'region', 'product_id', 'category', 'sub_category',
'product_name',
       'sales', 'quantity', 'profit', 'returns', 'payment_mode',
'ind1',
       'ind2'],
      dtype='object')

df.head()

   row_id+o6g3a1:r6  order_id  order_date   ship_date        ship_mode
\
0                4918         1  01-01-2019  07-01-2019  Standard Class

1                4919         2  02-01-2019  07-01-2019  Standard Class

2                4920         3  02-01-2019  07-01-2019  Standard Class

3                3074         4  03-01-2019  05-01-2019     First Class

4                8604         5  03-01-2019  08-01-2019  Standard Class


  customer_id      customer_name     segment          country
city  ...  \
0    BM-11575      Brendan Murry  Corporate  United States
Gaithersburg  ...
1    BM-11575      Brendan Murry  Corporate  United States
Gaithersburg  ...
2    BM-11575      Brendan Murry  Corporate  United States
Gaithersburg  ...
3    LR-16915       Lena Radford   Consumer  United States   Los
Angeles  ...
4    CA-12310  Christine Abelman  Corporate  United States   San
Antonio  ...


          category sub_category  \
0         Furniture    Bookcases
1         Furniture    Bookcases
2        Technology       Phones
3   Office Supplies      Storage
4        Technology  Accessories
```

```
                                          product_name   sales quantity
\
0  Bush Westfield Collection Bookcases, Medium Ch...   73.94        1

1  Bush Westfield Collection Bookcases, Medium Ch...  173.94        3

2                                     GE 30522EE2  231.98        2

3  Recycled Steel Personal File for Hanging File ...  114.46        2

4                 Imation Clip USB flash drive - 8 GB   30.08        2


     profit  returns  payment_mode  ind1  ind2
0   28.2668      NaN        Online   NaN   NaN
1   38.2668      NaN        Online   NaN   NaN
2   67.2742      NaN         Cards   NaN   NaN
3   28.6150      NaN        Online   NaN   NaN
4   -5.2640      NaN        Online   NaN   NaN

[5 rows x 23 columns]
```

```python
df.drop(columns = ['row_id+o6g3a1:r6'], inplace = True)

df.head()
```

```
   order_id  order_date  ship_date       ship_mode     customer_name
\
0         1  01-01-2019  07-01-2019  Standard Class      Brendan Murry

1         2  02-01-2019  07-01-2019  Standard Class      Brendan Murry

2         3  02-01-2019  07-01-2019  Standard Class      Brendan Murry

3         4  03-01-2019  05-01-2019     First Class       Lena Radford

4         5  03-01-2019  08-01-2019  Standard Class  Christine Abelman


      segment        country         city        state   region  \
0  Corporate  United States  Gaithersburg     Maryland     East
1  Corporate  United States  Gaithersburg     Maryland     East
2  Corporate  United States  Gaithersburg     Maryland     East
3   Consumer  United States   Los Angeles   California     West
4  Corporate  United States   San Antonio        Texas  Central

       product_id      category sub_category  \
0  FUR-BO-10004709     Furniture    Bookcases
1  FUR-BO-10004709     Furniture    Bookcases
2  TEC-PH-10000455    Technology       Phones
```

```
3  OFF-ST-10003692  Office Supplies       Storage
4  TEC-AC-10002217       Technology  Accessories

                                         product_name    sales  quantity
\
0  Bush Westfield Collection Bookcases, Medium Ch...    73.94         1

1  Bush Westfield Collection Bookcases, Medium Ch...   173.94         3

2                                    GE 30522EE2   231.98         2

3  Recycled Steel Personal File for Hanging File ...   114.46         2

4                 Imation Clip USB flash drive - 8 GB    30.08         2


    profit payment_mode
0  28.2668         Online
1  38.2668         Online
2  67.2742          Cards
3  28.6150         Online
4  -5.2640         Online
```

#for droping the columns permanently
```python
df.drop(columns = ['customer_id', 'returns', 'ind1', 'ind2'], inplace
=True)

df.head()
```

```
   order_id  order_date   ship_date      ship_mode      customer_name
\
0         1  01-01-2019  07-01-2019  Standard Class      Brendan Murry

1         2  02-01-2019  07-01-2019  Standard Class      Brendan Murry

2         3  02-01-2019  07-01-2019  Standard Class      Brendan Murry

3         4  03-01-2019  05-01-2019     First Class       Lena Radford

4         5  03-01-2019  08-01-2019  Standard Class  Christine Abelman


     segment         country         city       state   region  \
0  Corporate  United States  Gaithersburg    Maryland     East
1  Corporate  United States  Gaithersburg    Maryland     East
2  Corporate  United States  Gaithersburg    Maryland     East
3   Consumer  United States   Los Angeles  California     West
4  Corporate  United States   San Antonio       Texas  Central

        product_id        category sub_category  \
0  FUR-BO-10004709       Furniture    Bookcases
```

```
1  FUR-BO-10004709        Furniture    Bookcases
2  TEC-PH-10000455        Technology      Phones
3  OFF-ST-10003692  Office Supplies      Storage
4  TEC-AC-10002217        Technology  Accessories

                                    product_name   sales  quantity
\
0  Bush Westfield Collection Bookcases, Medium Ch...   73.94         1

1  Bush Westfield Collection Bookcases, Medium Ch...  173.94         3

2                                      GE 30522EE2  231.98         2

3  Recycled Steel Personal File for Hanging File ...  114.46         2

4                  Imation Clip USB flash drive - 8 GB   30.08         2


    profit payment_mode
0  28.2668       Online
1  38.2668       Online
2  67.2742        Cards
3  28.6150       Online
4  -5.2640       Online
```

`!pip install sqlalchemy`

```
Requirement already satisfied: sqlalchemy in c:\users\kokje\appdata\
local\programs\python\python311\lib\site-packages (2.0.37)


[notice] A new release of pip available: 22.3 -> 25.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip


Requirement already satisfied: greenlet!=0.4.17 in c:\users\kokje\
appdata\local\programs\python\python311\lib\site-packages (from
sqlalchemy) (3.1.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\
kokje\appdata\local\programs\python\python311\lib\site-packages (from
sqlalchemy) (4.12.2)
```

```python
# for connecting jupyter notebook to poatgreSQL

import sqlalchemy as sal
from urllib.parse import quote_plus


USERNAME = "postgres"
PASSWORD = quote_plus("Nevergiveup@2024")
HOST = "localhost"
```

```python
PORT = "5432"
DATABASE = "superstore_sql_project"

# Create engine
engine = sal.create_engine(f'postgresql://{USERNAME}:
{PASSWORD}@{HOST}:{PORT}/{DATABASE}')

# Test connection
try:
    conn = engine.connect()
    print("Connected successfully!")
except Exception as e:
    print("Connection failed:", e)
```

Connected successfully!

```python
df.to_sql('store_record', con=conn, index=False, if_exists =
'replace')
```

901

```python
df.columns
```

```
Index(['order_id', 'order_date', 'ship_date', 'ship_mode',
'customer_name',
       'segment', 'country', 'city', 'state', 'region', 'product_id',
       'category', 'sub_category', 'product_name', 'sales',
'quantity',
       'profit', 'payment_mode'],
      dtype='object')
```

```python
df.drop(columns = ['customer_name', 'ship_date'], inplace =True)
```

```python
df.dtypes
```

```
order_id          int64
order_date       object
ship_mode        object
segment          object
country          object
city             object
state            object
region           object
product_id       object
category         object
sub_category     object
product_name     object
sales           float64
quantity          int64
profit          float64
```

```
payment_mode       object
dtype: object

df.columns

Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country',
'city',
       'state', 'region', 'product_id', 'category', 'sub_category',
       'product_name', 'sales', 'quantity', 'profit', 'payment_mode'],
      dtype='object')

# for appending the tabel in the database
df.to_sql('store_record', con=conn, index=False, if_exists = 'append')

901
```