```
pip install numpy

Requirement already satisfied: numpy in d:\lib\site-packages
(2.2.3)Note: you may need to restart the kernel to use updated
packages.



[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip

pip install pandas

Requirement already satisfied: pandas in d:\lib\site-packages (2.2.3)
Requirement already satisfied: numpy>=1.26.0 in d:\lib\site-packages
(from pandas) (2.2.3)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\lib\site-
packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\lib\site-packages
(from pandas) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in d:\lib\site-packages
(from pandas) (2025.1)
Requirement already satisfied: six>=1.5 in d:\lib\site-packages (from
python-dateutil>=2.8.2->pandas) (1.17.0)
Note: you may need to restart the kernel to use updated packages.


[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip

pip install matplotlib

Requirement already satisfied: matplotlib in d:\lib\site-packages
(3.10.1)
Requirement already satisfied: contourpy>=1.0.1 in d:\lib\site-
packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in d:\lib\site-packages
(from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in d:\lib\site-
packages (from matplotlib) (4.55.0)
Requirement already satisfied: kiwisolver>=1.3.1 in d:\lib\site-
packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in d:\lib\site-packages
(from matplotlib) (2.2.3)
Requirement already satisfied: packaging>=20.0 in d:\lib\site-packages
(from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in d:\lib\site-packages (from
matplotlib) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\lib\site-
packages (from matplotlib) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in d:\lib\site-
```

```
packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in d:\lib\site-packages (from
python-dateutil>=2.7->matplotlib) (1.17.0)
Note: you may need to restart the kernel to use updated packages.


[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip
```

```
pip install seaborn
```

```
Note: you may need to restart the kernel to use updated
packages.Requirement already satisfied: seaborn in d:\lib\site-
packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in d:\lib\site-
packages (from seaborn) (2.2.3)
Requirement already satisfied: pandas>=1.2 in d:\lib\site-packages
(from seaborn) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in d:\lib\site-
packages (from seaborn) (3.10.1)
Requirement already satisfied: contourpy>=1.0.1 in d:\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.1)
Requirement already satisfied: cycler>=0.10 in d:\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in d:\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.55.0)
Requirement already satisfied: kiwisolver>=1.3.1 in d:\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)
Requirement already satisfied: packaging>=20.0 in d:\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: pillow>=8 in d:\lib\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in d:\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\lib\site-packages
(from pandas>=1.2->seaborn) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in d:\lib\site-packages
(from pandas>=1.2->seaborn) (2025.1)
Requirement already satisfied: six>=1.5 in d:\lib\site-packages (from
python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)



[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip
```

Load datasets

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

transactions_1997 = pd.read_csv('F:\\Data Analyst\\File\\
Maven+Market+CSV+Files (1)\\MavenMarket_Transactions_1997.csv')
transactions_1997
```

```
      transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1997  12/31/1996         869         3449         6

1              1/1/1997  12/31/1996        1472         3449         6

2              1/1/1997  12/28/1996          76         3449         6

3              1/1/1997  12/26/1996         320         3449         6

4              1/1/1997  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

86832       12/30/1997  12/29/1997        1376         7043        14

86833       12/30/1997  12/26/1997        1380         7043        14

86834       12/30/1997  12/23/1997         637         7043        14

86835       12/30/1997  12/29/1997         930         2145        14

86836       12/30/1997  12/28/1997         427         2145        14


       quantity
0             5
1             3
2             4
3             3
4             4
...         ...
86832         2
86833         1
86834         2
86835         1
86836         2

[86837 rows x 6 columns]
```

```
transactions_1998 = pd.read_csv('F:\\Data Analyst\\File\\
Maven+Market+CSV+Files (1)\\MavenMarket_Transactions_1998.csv')
transactions_1998
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1998  12/25/1997           4         2439        10

1              1/1/1998  12/28/1997          11         4284        10

2              1/1/1998  12/28/1997          12          534        10

3              1/1/1998  12/29/1997          14         9743        10

4              1/1/1998  12/27/1997          16         3608        10

...                 ...         ...         ...          ...       ...

182878       12/30/1998  12/29/1998        1521         7197        11

182879       12/30/1998  12/23/1998        1536         5223        10

182880       12/30/1998  12/23/1998        1542         8077        10

182881       12/30/1998  12/28/1998        1544         4485        10

182882       12/30/1998  12/29/1998        1549         5223        10


        quantity
0              3
1              3
2              3
3              2
4              3
...          ...
182878         3
182879         2
182880         4
182881         2
182882         3

[182883 rows x 6 columns]
```

```
returns = pd.read_csv('F:\\Data Analyst\\File\\Maven+Market+CSV+Files
(1)\\MavenMarket_Returns_1997-1998.csv')
returns
```

```
     return_date  product_id  store_id  quantity
0       1/1/1997         250         6         1
1       1/1/1997         628         6         1
```

```
2       1/1/1997          869          6          1
3       1/2/1997          469         11          1
4       1/2/1997          532         23          2
...          ...          ...        ...        ...
7082  12/30/1998         1037         11          2
7083  12/30/1998         1048         10          1
7084  12/30/1998         1065         10          1
7085  12/30/1998         1154         11          1
7086  12/30/1998         1291         11          1

[7087 rows x 4 columns]
```

```
customers = pd.read_csv('F:\\Data Analyst\\File\\
Maven+Market+CSV+Files (1)\\MavenMarket_Customers.csv')
customers
```

```
       customer_id  customer_acct_num first_name  last_name  \
0                1        87462024688      Sheri     Nowmer
1                2        87470586299    Derrick    Whelply
2                3        87475757600     Jeanne      Derry
3                4        87500482201    Michael     Spence
4                5        87514054179       Maya  Gutierrez
...            ...                ...        ...        ...
10276        10277        87439274191       Fran       Ross
10277        10278        87448420500     Myreda    Calahoo
10278        10279        87453135848       Mary      Ayers
10279        10280        87458639740     Ernest     Aiello
10280        10281        87460163235     Samuel    Cartney

            customer_address  customer_city customer_state_province  \
0           2433 Bailey Road       Tlaxiaco                   Oaxaca
1         2219 Dewing Avenue          Sooke                       BC
2            7640 First Ave.       Issaquah                       WA
3              337 Tosca Way        Burnaby                       BC
4            8668 Via Neruda         Novato                       CA
...                      ...            ...                      ...
10276  5603 Blackridge Drive    Lake Oswego                       OR
10277      263 La Orinda Pl.   N. Vancouver                       BC
10278            6885 Auburn  Lincoln Acres                       CA
10279      5077 Bannock Ct.       Puyallup                       WA
10280     4609 Parkway Drive      Vancouver                       BC

       customer_postal_code customer_country  birthdate marital_status
\
0                     15057           Mexico  8/26/1961              M

1                     17172           Canada   7/3/1915              S

2                     73980              USA  6/21/1910              M
```

```
3                    74674          Canada  6/20/1969              M

4                    57355             USA  5/10/1951              S

...                    ...             ...        ...            ...

10276                52724             USA   2/9/1974              M

10277                71758          Canada  12/8/1926              M

10278                42550             USA  5/18/1913              S

10279                27746             USA   9/6/1968              M

10280                63699          Canada   7/6/1914              S


       yearly_income gender  total_children  num_children_at_home  \
0         $30K - $50K      F               4                     2
1         $70K - $90K      M               1                     0
2         $50K - $70K      F               1                     1
3         $10K - $30K      M               4                     4
4         $30K - $50K      F               3                     0
...               ...    ...             ...                   ...
10276    $90K - $110K      M               4                     3
10277     $30K - $50K      F               0                     0
10278   $130K - $150K      M               3                     0
10279         $150K +      F               5                     2
10280     $50K - $70K      F               5                     0

                 education acct_open_date member_card      occupation
\
0      Partial High School      9/10/1991      Bronze  Skilled Manual

1      Partial High School      3/11/1993      Bronze    Professional

2         Bachelors Degree      6/11/1991      Bronze    Professional

3      Partial High School      5/21/1994      Normal  Skilled Manual

4          Partial College      8/21/1992      Silver          Manual

...                    ...            ...         ...             ...

10276  Partial High School      3/14/1991      Bronze      Management

10277      Partial College      3/20/1992      Bronze    Professional

10278  Partial High School     11/22/1991      Bronze      Management

10279   High School Degree      5/26/1991      Golden    Professional
```

```
10280     Bachelors Degree      7/22/1993      Bronze      Management


      homeowner
0             Y
1             N
2             Y
3             N
4             N
...         ...
10276         N
10277         N
10278         Y
10279         Y
10280         N

[10281 rows x 20 columns]
```

```
products = pd.read_csv('F:\\Data Analyst\\File\\Maven+Market+CSV+Files
(1)\\MavenMarket_Products.csv')
products
```

```
      product_id product_brand                    product_name
product_sku  \
0              1    Washington          Washington Berry Juice
90748583674
1              2    Washington          Washington Mango Drink
96516502499
2              3    Washington     Washington Strawberry Drink
58427771925
3              4    Washington           Washington Cream Soda
64412155747
4              5    Washington            Washington Diet Soda
85561191439
...          ...           ...                             ...
...
1555        1556           CDR        CDR Creamy Peanut Butter
29538288712
1556        1557           CDR        CDR Strawberry Preserves
50687324404
1557        1558           CDR  CDR Extra Chunky Peanut Butter
84930775761
1558        1559           CDR              CDR Apple Preserves
75317577719
1559        1560           CDR                  CDR Grape Jelly
54896665215

      product_retail_price  product_cost  product_weight  recyclable
low_fat
```

```
0                    2.85          0.94           8.39          NaN
NaN
1                    0.74          0.26           7.42          NaN
1.0
2                    0.83          0.40          13.10          1.0
1.0
3                    3.64          1.64          10.60          1.0
NaN
4                    2.19          0.77           6.66          1.0
NaN
...                   ...           ...            ...          ...
...
1555                 2.65          1.14           6.94          1.0
1.0
1556                 1.20          0.48          15.40          1.0
NaN
1557                 2.16          0.82          11.50          NaN
1.0
1558                 1.62          0.62          21.00          NaN
NaN
1559                 1.60          0.74          12.50          NaN
1.0

[1560 rows x 9 columns]
```

```
stores = pd.read_csv('F:\\Data Analyst\\File\\Maven+Market+CSV+Files
(1)\\MavenMarket_Stores.csv')
stores
```

```
    store_id  region_id           store_type store_name  \
0          1         28          Supermarket    Store 1
1          2         78        Small Grocery    Store 2
2          3         76          Supermarket    Store 3
3          4         27  Gourmet Supermarket    Store 4
4          5          4        Small Grocery    Store 5
5          6         47  Gourmet Supermarket    Store 6
6          7          3          Supermarket    Store 7
7          8         26   Deluxe Supermarket    Store 8
8          9          2     Mid-Size Grocery    Store 9
9         10         24          Supermarket   Store 10
10        11         22          Supermarket   Store 11
11        12         25   Deluxe Supermarket   Store 12
12        13         23   Deluxe Supermarket   Store 13
13        14          1        Small Grocery   Store 14
14        15         18          Supermarket   Store 15
15        16         87          Supermarket   Store 16
16        17         84   Deluxe Supermarket   Store 17
17        18         25     Mid-Size Grocery   Store 18
18        19          5   Deluxe Supermarket   Store 19
19        20          6     Mid-Size Grocery   Store 20
```

```
20         21         106   Deluxe Supermarket   Store 21
21         22          88        Small Grocery   Store 22
22         23          89     Mid-Size Grocery   Store 23
23         24           7          Supermarket   Store 24

          store_street_address     store_city store_state store_country
\
0                2853 Bailey Rd       Acapulco    Guerrero        Mexico

1            5203 Catanzaro Way     Bellingham          WA           USA

2            1501 Ramsey Circle      Bremerton          WA           USA

3              433 St George Dr        Camacho   Zacatecas        Mexico

4            1250 Coggins Drive    Guadalajara     Jalisco        Mexico

5   5495 Mitchell Canyon Road  Beverly Hills          CA           USA

6              1077 Wharf Drive    Los Angeles          CA           USA

7          3173 Buena Vista Ave         Merida     Yucatan        Mexico

8          1872 El Pintado Road    Mexico City          DF        Mexico

9             7894 Rotherham Dr        Orizaba    Veracruz        Mexico

10          5371 Holland Circle       Portland          OR           USA

11          1120 Westchester Pl        Hidalgo   Zacatecas        Mexico

12              5179 Valley Ave          Salem          OR           USA

13               4365 Indigo Ct  San Francisco          CA           USA

14          5006 Highland Drive        Seattle          WA           USA

15             5922 La Salle Ct        Spokane          WA           USA

16              490 Risdon Road         Tacoma          WA           USA

17               6764 Glen Road        Hidalgo   Zacatecas        Mexico

18           6644 Sudance Drive      Vancouver          BC        Canada

19             3706 Marvelle Ln       Victoria          BC        Canada

20           4093 Steven Circle     San Andres          DF        Mexico

21             9606 Julpum Loop    Walla Walla          WA           USA
```

```
22            3920 Noah Court         Yakima           WA            USA

23           2342 Waltham St.      San Diego           CA            USA


     store_phone first_opened_date last_remodel_date  total_sqft
grocery_sqft
0   262-555-5124          1/9/1982         12/5/1990       23593
17475
1   605-555-8203          4/2/1970          6/4/1973       28206
22271
2   509-555-1596         6/14/1959        11/19/1967       39696
24390
3   304-555-1474         9/27/1994         12/1/1995       23759
16844
4   801-555-4324         9/18/1978         6/29/1991       24597
15012
5   958-555-5002          1/3/1981         3/13/1991       23688
15337
6   477-555-7967         5/21/1971        10/20/1981       23598
14210
7   797-555-3417         9/23/1958        11/18/1967       30797
20141
8   439-555-3524         3/18/1955          6/7/1959       36509
22450
9   212-555-4774         4/13/1979         1/30/1982       34791
26354
10  685-555-8995         9/17/1976         5/15/1982       20319
16232
11  151-555-1702         3/25/1968        12/18/1993       30584
21938
12  977-555-2724         4/13/1957        11/10/1997       27694
18670
13  135-555-4888        11/24/1957          1/7/1958       22478
15321
14  893-555-1024         7/24/1969        10/19/1973       21215
13305
15  643-555-3645         8/23/1974         7/13/1977       30268
22063
16  855-555-5581         5/30/1970         6/23/1976       33858
22123
17  528-555-8317         6/28/1969         8/30/1975       38382
30351
18  862-555-7395         3/27/1977        10/25/1990       23112
16418
19  897-555-1931          2/6/1980          4/9/1987       34452
27463
20  493-555-4781          2/7/1986         4/16/1990       32717
25453
21  881-555-5117         1/24/1951        10/17/1969       35918
```

```
24837
22  170-555-8424         7/16/1977         7/24/1987       29182
19283
23  111-555-0303         5/22/1979         4/20/1986       27372
18293
```

```
regions = pd.read_csv('F:\\Data Analyst\\File\\Maven+Market+CSV+Files
(1)\\MavenMarket_Regions.csv')
regions
```

```
     region_id sales_district    sales_region
0            1  San Francisco    Central West
1            2    Mexico City  Mexico Central
2            3    Los Angeles      South West
3            4    Guadalajara     Mexico West
4            5      Vancouver     Canada West
..         ...             ...             ...
104        105       Victoria     Canada West
105        106    Mexico City  Mexico Central
106        107    Mexico City  Mexico Central
107        108    Mexico City  Mexico Central
108        109    Mexico City  Mexico Central

[109 rows x 3 columns]
```

Combine transaction data

```
transactions = pd.concat([transactions_1997, transactions_1998],
ignore_index=True)
transactions
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1997  12/31/1996         869         3449         6

1              1/1/1997  12/31/1996        1472         3449         6

2              1/1/1997  12/28/1996          76         3449         6

3              1/1/1997  12/26/1996         320         3449         6

4              1/1/1997  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       12/30/1998  12/29/1998        1521         7197        11

269716       12/30/1998  12/23/1998        1536         5223        10
```

```
269717       12/30/1998  12/23/1998        1542          8077         10

269718       12/30/1998  12/28/1998        1544          4485         10

269719       12/30/1998  12/29/1998        1549          5223         10


        quantity
0              5
1              3
2              4
3              3
4              4
...          ...
269715         3
269716         2
269717         4
269718         2
269719         3

[269720 rows x 6 columns]
```

Display basic info

```
print("Transactions Data Info:\n")
transactions.info()
```

```
Transactions Data Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 269720 entries, 0 to 269719
Data columns (total 6 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   transaction_date  269720 non-null  object
 1   stock_date        269720 non-null  object
 2   product_id        269720 non-null  int64
 3   customer_id       269720 non-null  int64
 4   store_id          269720 non-null  int64
 5   quantity          269720 non-null  int64
dtypes: int64(4), object(2)
memory usage: 12.3+ MB
```

```
print("\nReturns Data Info:\n")
print(returns.info())
```

```
Returns Data Info:
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7087 entries, 0 to 7086
Data columns (total 4 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   return_date  7087 non-null   object
 1   product_id   7087 non-null   int64 
 2   store_id     7087 non-null   int64 
 3   quantity     7087 non-null   int64 
dtypes: int64(3), object(1)
memory usage: 221.6+ KB
None
```

Merge transactions with products and stores

```
df = transactions.merge(products, on='product_id', how='left')
df


df = df.merge(stores, on='store_id', how='left')
df
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1997  12/31/1996         869         3449         6

1              1/1/1997  12/31/1996        1472         3449         6

2              1/1/1997  12/28/1996          76         3449         6

3              1/1/1997  12/26/1996         320         3449         6

4              1/1/1997  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       12/30/1998  12/29/1998        1521         7197        11

269716       12/30/1998  12/23/1998        1536         5223        10

269717       12/30/1998  12/23/1998        1542         8077        10

269718       12/30/1998  12/28/1998        1544         4485        10

269719       12/30/1998  12/29/1998        1549         5223        10


        quantity product_brand                      product_name
```

```
product_sku  \
0              5     Nationeel        Nationeel Grape Fruit Roll
52382137179
1              3     Fort West          Fort West Fudge Cookies
37276054024
2              4     Red Spade         Red Spade Sliced Chicken
62054644227
3              3     Excellent        Excellent Cranberry Juice
36570182442
4              4    Washington            Washington Cream Soda
64412155747
...          ...           ...                               ...
...
269715         3   Top Measure  Top Measure White Zinfandel Wine
19118239283
269716         2           CDR                  CDR White Sugar
92759070992
269717         4           CDR                 CDR Tomato Sauce
95931351780
269718         2           CDR               CDR Grape Preserves
65897241234
269719         3           CDR                         CDR Salt
44236547350

        product_retail_price  ...  store_name
store_street_address  \
0                       2.12  ...     Store 6  5495 Mitchell Canyon
Road
1                       2.20  ...     Store 6  5495 Mitchell Canyon
Road
2                       1.69  ...     Store 6  5495 Mitchell Canyon
Road
3                       3.26  ...     Store 6  5495 Mitchell Canyon
Road
4                       3.64  ...     Store 6  5495 Mitchell Canyon
Road
...                      ...  ...         ...                        .
..
269715                  0.70  ...    Store 11        5371 Holland
Circle
269716                  1.33  ...    Store 10          7894 Rotherham
Dr
269717                  0.95  ...    Store 10          7894 Rotherham
Dr
269718                  1.46  ...    Store 10          7894 Rotherham
Dr
269719                  1.52  ...    Store 10          7894 Rotherham
Dr
```

```
           store_city  store_state  store_country   store_phone  \
0       Beverly Hills           CA            USA  958-555-5002
1       Beverly Hills           CA            USA  958-555-5002
2       Beverly Hills           CA            USA  958-555-5002
3       Beverly Hills           CA            USA  958-555-5002
4       Beverly Hills           CA            USA  958-555-5002
...               ...          ...            ...           ...
269715       Portland           OR            USA  685-555-8995
269716        Orizaba     Veracruz         Mexico  212-555-4774
269717        Orizaba     Veracruz         Mexico  212-555-4774
269718        Orizaba     Veracruz         Mexico  212-555-4774
269719        Orizaba     Veracruz         Mexico  212-555-4774

       first_opened_date last_remodel_date total_sqft grocery_sqft
0               1/3/1981         3/13/1991      23688        15337
1               1/3/1981         3/13/1991      23688        15337
2               1/3/1981         3/13/1991      23688        15337
3               1/3/1981         3/13/1991      23688        15337
4               1/3/1981         3/13/1991      23688        15337
...                  ...               ...        ...          ...
269715         9/17/1976         5/15/1982      20319        16232
269716         4/13/1979         1/30/1982      34791        26354
269717         4/13/1979         1/30/1982      34791        26354
269718         4/13/1979         1/30/1982      34791        26354
269719         4/13/1979         1/30/1982      34791        26354

[269720 rows x 26 columns]
```

Convert date columns

```
df['transaction_date'] = pd.to_datetime(df['transaction_date'])
df
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0            1997-01-01  12/31/1996         869         3449         6

1            1997-01-01  12/31/1996        1472         3449         6

2            1997-01-01  12/28/1996          76         3449         6

3            1997-01-01  12/26/1996         320         3449         6

4            1997-01-01  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       1998-12-30  12/29/1998        1521         7197        11
```

```
269716       1998-12-30  12/23/1998        1536         5223        10

269717       1998-12-30  12/23/1998        1542         8077        10

269718       1998-12-30  12/28/1998        1544         4485        10

269719       1998-12-30  12/29/1998        1549         5223        10


        quantity product_brand                      product_name
product_sku  \
0              5     Nationeel        Nationeel Grape Fruit Roll
52382137179
1              3     Fort West           Fort West Fudge Cookies
37276054024
2              4     Red Spade          Red Spade Sliced Chicken
62054644227
3              3     Excellent         Excellent Cranberry Juice
36570182442
4              4    Washington             Washington Cream Soda
64412155747
...          ...           ...                              ...
...
269715         3   Top Measure  Top Measure White Zinfandel Wine
19118239283
269716         2           CDR                  CDR White Sugar
92759070992
269717         4           CDR                 CDR Tomato Sauce
95931351780
269718         2           CDR               CDR Grape Preserves
65897241234
269719         3           CDR                         CDR Salt
44236547350

        product_retail_price  ...  store_name
store_street_address  \
0                       2.12  ...     Store 6  5495 Mitchell Canyon
Road
1                       2.20  ...     Store 6  5495 Mitchell Canyon
Road
2                       1.69  ...     Store 6  5495 Mitchell Canyon
Road
3                       3.26  ...     Store 6  5495 Mitchell Canyon
Road
4                       3.64  ...     Store 6  5495 Mitchell Canyon
Road
...                      ...  ...         ...                        .
..
269715                  0.70  ...    Store 11        5371 Holland
Circle
```

```
269716                  1.33  ...    Store 10          7894 Rotherham
Dr
269717                  0.95  ...    Store 10          7894 Rotherham
Dr
269718                  1.46  ...    Store 10          7894 Rotherham
Dr
269719                  1.52  ...    Store 10          7894 Rotherham
Dr

           store_city  store_state  store_country   store_phone  \
0       Beverly Hills           CA            USA  958-555-5002
1       Beverly Hills           CA            USA  958-555-5002
2       Beverly Hills           CA            USA  958-555-5002
3       Beverly Hills           CA            USA  958-555-5002
4       Beverly Hills           CA            USA  958-555-5002
...               ...          ...            ...           ...
269715       Portland           OR            USA  685-555-8995
269716        Orizaba     Veracruz         Mexico  212-555-4774
269717        Orizaba     Veracruz         Mexico  212-555-4774
269718        Orizaba     Veracruz         Mexico  212-555-4774
269719        Orizaba     Veracruz         Mexico  212-555-4774

       first_opened_date last_remodel_date total_sqft grocery_sqft
0               1/3/1981         3/13/1991      23688        15337
1               1/3/1981         3/13/1991      23688        15337
2               1/3/1981         3/13/1991      23688        15337
3               1/3/1981         3/13/1991      23688        15337
4               1/3/1981         3/13/1991      23688        15337
...                  ...               ...        ...          ...
269715         9/17/1976         5/15/1982      20319        16232
269716         4/13/1979         1/30/1982      34791        26354
269717         4/13/1979         1/30/1982      34791        26354
269718         4/13/1979         1/30/1982      34791        26354
269719         4/13/1979         1/30/1982      34791        26354

[269720 rows x 26 columns]
```

Sales analysis

```
df['Total_Sales'] = df['quantity'] * df['product_cost']
df

       transaction_date  stock_date  product_id  customer_id  store_id
\
0            1997-01-01  12/31/1996         869         3449         6

1            1997-01-01  12/31/1996        1472         3449         6
```

```
2            1997-01-01  12/28/1996          76         3449          6
3            1997-01-01  12/26/1996         320         3449          6
4            1997-01-01  12/25/1996           4         3449          6
...                 ...         ...         ...          ...        ...
269715       1998-12-30  12/29/1998        1521         7197         11
269716       1998-12-30  12/23/1998        1536         5223         10
269717       1998-12-30  12/23/1998        1542         8077         10
269718       1998-12-30  12/28/1998        1544         4485         10
269719       1998-12-30  12/29/1998        1549         5223         10


        quantity product_brand                       product_name
product_sku  \
0              5     Nationeel        Nationeel Grape Fruit Roll
52382137179
1              3     Fort West           Fort West Fudge Cookies
37276054024
2              4     Red Spade          Red Spade Sliced Chicken
62054644227
3              3     Excellent         Excellent Cranberry Juice
36570182442
4              4    Washington             Washington Cream Soda
64412155747
...          ...           ...                              ...
...
269715         3   Top Measure  Top Measure White Zinfandel Wine
19118239283
269716         2           CDR                  CDR White Sugar
92759070992
269717         4           CDR                CDR Tomato Sauce
95931351780
269718         2           CDR               CDR Grape Preserves
65897241234
269719         3           CDR                         CDR Salt
44236547350

        product_retail_price  ...       store_street_address
store_city  \
0                       2.12  ...  5495 Mitchell Canyon Road  Beverly
Hills
1                       2.20  ...  5495 Mitchell Canyon Road  Beverly
Hills
```

```
2                       1.69  ...  5495 Mitchell Canyon Road  Beverly
Hills
3                       3.26  ...  5495 Mitchell Canyon Road  Beverly
Hills
4                       3.64  ...  5495 Mitchell Canyon Road  Beverly
Hills
...                      ...  ...                        ...
...
269715                  0.70  ...        5371 Holland Circle
Portland
269716                  1.33  ...          7894 Rotherham Dr
Orizaba
269717                  0.95  ...          7894 Rotherham Dr
Orizaba
269718                  1.46  ...          7894 Rotherham Dr
Orizaba
269719                  1.52  ...          7894 Rotherham Dr
Orizaba

        store_state  store_country   store_phone first_opened_date  \
0                CA            USA  958-555-5002          1/3/1981
1                CA            USA  958-555-5002          1/3/1981
2                CA            USA  958-555-5002          1/3/1981
3                CA            USA  958-555-5002          1/3/1981
4                CA            USA  958-555-5002          1/3/1981
...             ...            ...           ...               ...
269715           OR            USA  685-555-8995         9/17/1976
269716     Veracruz         Mexico  212-555-4774          4/13/1979
269717     Veracruz         Mexico  212-555-4774          4/13/1979
269718     Veracruz         Mexico  212-555-4774          4/13/1979
269719     Veracruz         Mexico  212-555-4774          4/13/1979

       last_remodel_date total_sqft grocery_sqft Total_Sales
0              3/13/1991      23688        15337        4.55
1              3/13/1991      23688        15337        2.70
2              3/13/1991      23688        15337        2.76
3              3/13/1991      23688        15337        3.24
4              3/13/1991      23688        15337        6.56
...                  ...        ...          ...         ...
269715         5/15/1982      20319        16232        0.78
269716         1/30/1982      34791        26354        0.96
269717         1/30/1982      34791        26354        1.64
269718         1/30/1982      34791        26354        1.14
269719         1/30/1982      34791        26354        2.22

[269720 rows x 27 columns]
```

Yearly sales comparison

```
df['Year'] = df['transaction_date'].dt.year
sales_by_year = df.groupby('Year')['Total_Sales'].sum()
df
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0            1997-01-01  12/31/1996         869         3449         6

1            1997-01-01  12/31/1996        1472         3449         6

2            1997-01-01  12/28/1996          76         3449         6

3            1997-01-01  12/26/1996         320         3449         6

4            1997-01-01  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       1998-12-30  12/29/1998        1521         7197        11

269716       1998-12-30  12/23/1998        1536         5223        10

269717       1998-12-30  12/23/1998        1542         8077        10

269718       1998-12-30  12/28/1998        1544         4485        10

269719       1998-12-30  12/29/1998        1549         5223        10


        quantity product_brand                      product_name
product_sku  \
0              5     Nationeel        Nationeel Grape Fruit Roll
52382137179
1              3     Fort West           Fort West Fudge Cookies
37276054024
2              4     Red Spade          Red Spade Sliced Chicken
62054644227
3              3     Excellent         Excellent Cranberry Juice
36570182442
4              4    Washington             Washington Cream Soda
64412155747
...          ...           ...                               ...
...
269715         3   Top Measure  Top Measure White Zinfandel Wine
19118239283
269716         2           CDR                   CDR White Sugar
```

```
92759070992
269717         4            CDR                   CDR Tomato Sauce
95931351780
269718         2            CDR                CDR Grape Preserves
65897241234
269719         3            CDR                          CDR Salt
44236547350

        product_retail_price  ...     store_city  store_state
store_country  \
0                       2.12  ...  Beverly Hills           CA
USA
1                       2.20  ...  Beverly Hills           CA
USA
2                       1.69  ...  Beverly Hills           CA
USA
3                       3.26  ...  Beverly Hills           CA
USA
4                       3.64  ...  Beverly Hills           CA
USA
...                      ...  ...            ...          ...
...
269715                  0.70  ...       Portland           OR
USA
269716                  1.33  ...        Orizaba     Veracruz
Mexico
269717                  0.95  ...        Orizaba     Veracruz
Mexico
269718                  1.46  ...        Orizaba     Veracruz
Mexico
269719                  1.52  ...        Orizaba     Veracruz
Mexico

         store_phone  first_opened_date last_remodel_date
total_sqft  \
0       958-555-5002           1/3/1981         3/13/1991      23688

1       958-555-5002           1/3/1981         3/13/1991      23688

2       958-555-5002           1/3/1981         3/13/1991      23688

3       958-555-5002           1/3/1981         3/13/1991      23688

4       958-555-5002           1/3/1981         3/13/1991      23688

...              ...                ...               ...        ...

269715  685-555-8995          9/17/1976         5/15/1982      20319

269716  212-555-4774          4/13/1979         1/30/1982      34791
```

```
269717  212-555-4774          4/13/1979          1/30/1982      34791

269718  212-555-4774          4/13/1979          1/30/1982      34791

269719  212-555-4774          4/13/1979          1/30/1982      34791


       grocery_sqft Total_Sales  Year
0             15337        4.55  1997
1             15337        2.70  1997
2             15337        2.76  1997
3             15337        3.24  1997
4             15337        6.56  1997
...             ...         ...   ...
269715        16232        0.78  1998
269716        26354        0.96  1998
269717        26354        1.64  1998
269718        26354        1.14  1998
269719        26354        2.22  1998

[269720 rows x 28 columns]
```

Identify most returned products

```
import squarify
import matplotlib.pyplot as plt
import seaborn as sns

returns_summary = returns.groupby('product_id')
['quantity'].sum().reset_index()
returns_summary = returns_summary.merge(products[['product_id',
'product_name', 'product_brand']], on='product_id', how='left')
returns_category = returns_summary.groupby('product_brand')
['quantity'].sum().reset_index()
returns_category = returns_category.sort_values(by='quantity',
ascending=False).head(20)  # Top 20 brands
returns_category
```

```
    product_brand  quantity
54       Hermanos       274
58        Horatio       240
107     Tri-State       239
100     Tell Tale       238
33          Ebony       223
```

56	High Top	222
77	Nationeel	221
39	Fast	208
88	Red Wing	194
96	Sunset	186
11	Big Time	174
42	Fort West	174
55	High Quality	172
59	Imagine	160
80	Plato	159
8	Best Choice	156
79	PigTail	153
20	Carrington	152
69	Landslide	150
26	Cormorant	149

Calculate percentage contribution of each brand

```
# Step 1: Calculate the percentage
returns_category['percentage'] = (returns_category['quantity'] /
returns_category['quantity'].sum()) * 100
# Step 2: Create label column
returns_category['label'] = returns_category.apply(lambda x:
f"{x['product_brand']}\n{x['percentage']:.1f}%", axis=1)
returns_category
```

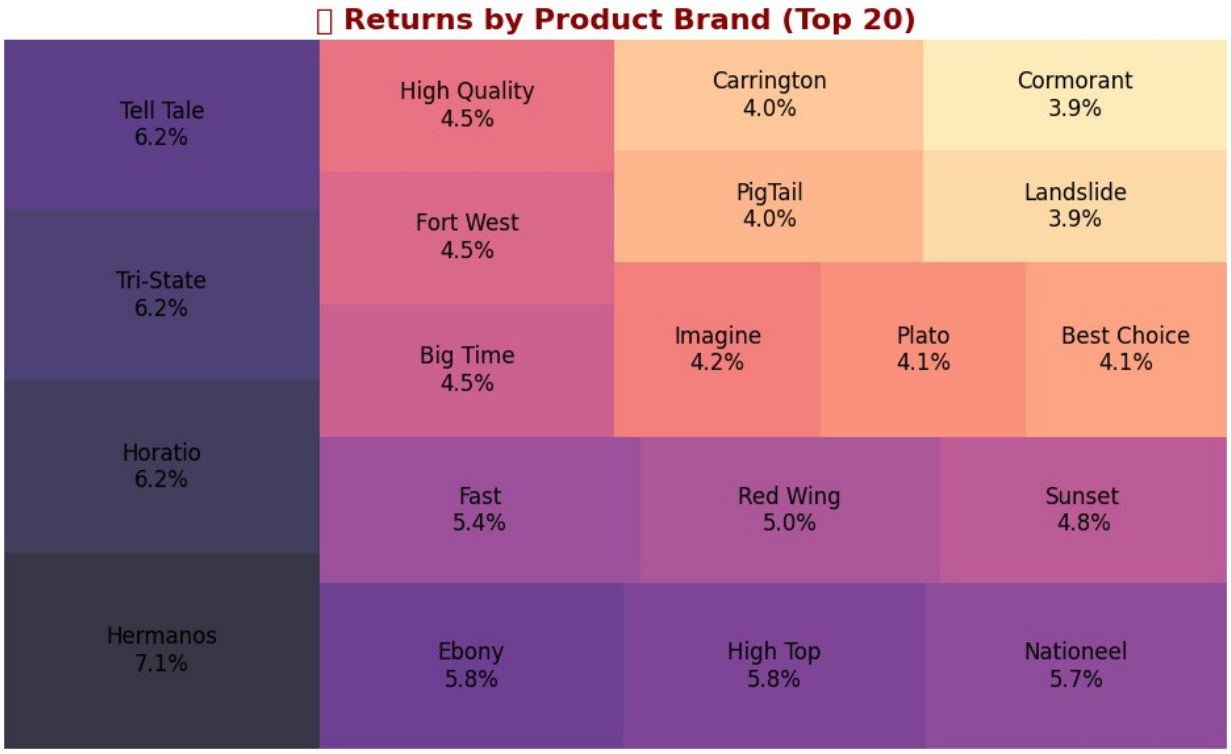
	product_brand	quantity	percentage	label
54	Hermanos	274	7.127992	Hermanos\n7.1%
58	Horatio	240	6.243496	Horatio\n6.2%
107	Tri-State	239	6.217482	Tri-State\n6.2%
100	Tell Tale	238	6.191467	Tell Tale\n6.2%
33	Ebony	223	5.801249	Ebony\n5.8%
56	High Top	222	5.775234	High Top\n5.8%
77	Nationeel	221	5.749220	Nationeel\n5.7%
39	Fast	208	5.411030	Fast\n5.4%
88	Red Wing	194	5.046826	Red Wing\n5.0%
96	Sunset	186	4.838710	Sunset\n4.8%
11	Big Time	174	4.526535	Big Time\n4.5%
42	Fort West	174	4.526535	Fort West\n4.5%
55	High Quality	172	4.474506	High Quality\n4.5%
59	Imagine	160	4.162331	Imagine\n4.2%
80	Plato	159	4.136316	Plato\n4.1%
8	Best Choice	156	4.058273	Best Choice\n4.1%
79	PigTail	153	3.980229	PigTail\n4.0%
20	Carrington	152	3.954214	Carrington\n4.0%

```
69      Landslide       150    3.902185     Landslide\n3.9%
26      Cormorant       149    3.876171     Cormorant\n3.9%
```

Treemap visualization

```
plt.figure(figsize=(12, 7))
squarify.plot(
    sizes=returns_category['quantity'],
    label=returns_category['label'],
    alpha=0.8,
    color=sns.color_palette("magma", len(returns_category)),
    text_kwargs={'fontsize': 12}  # Adjust text size for readability
)
# Customizing labels & title
plt.title(' Returns by Product Brand (Top 20)', fontsize=16,
fontweight='bold', color='darkred')
plt.axis('off')  # Hide axes for cleaner look

# Show plot

plt.show()
```

```
D:\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning:
Glyph 128230 (\N{PACKAGE}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



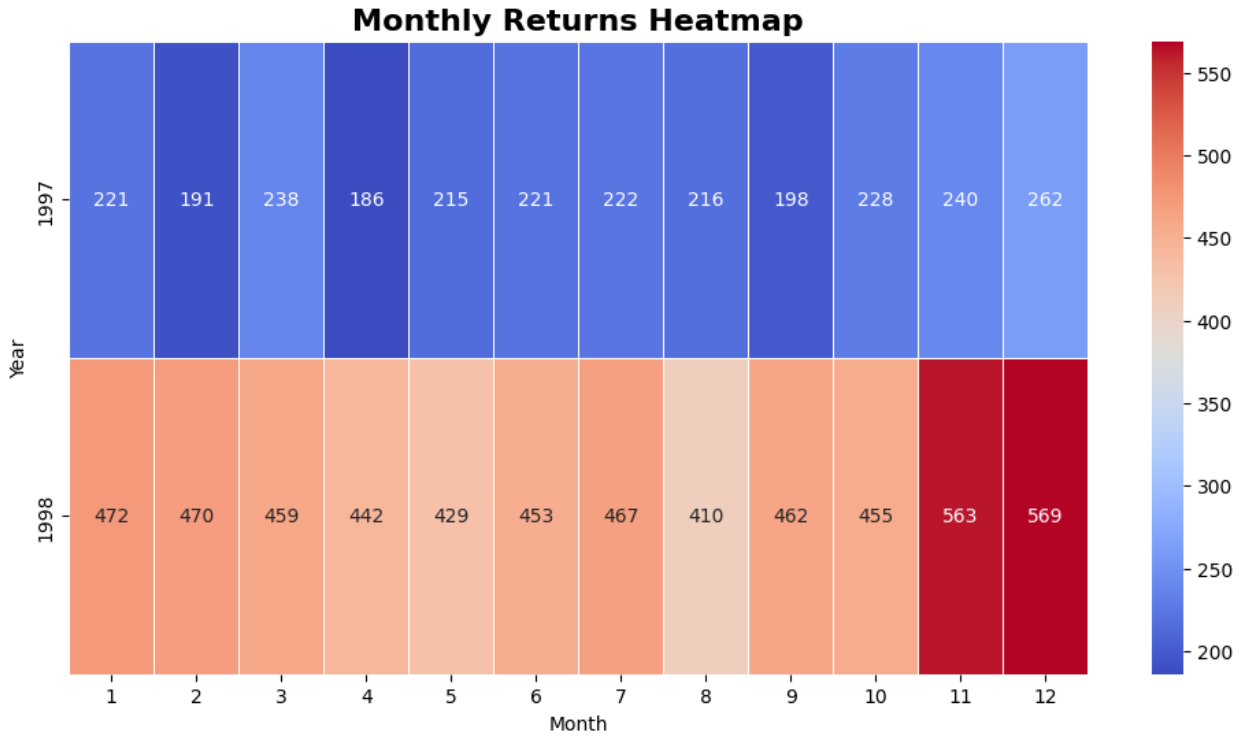

```
pip install squarify
```

```
Requirement already satisfied: squarify in d:\lib\site-packages
(0.4.4)
Note: you may need to restart the kernel to use updated packages.


[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip
```

Heatmap of Sales Returns Per Month

```


# Extract Month and Year
returns['return_date'] = pd.to_datetime(returns['return_date'])
returns['Year'] = returns['return_date'].dt.year
returns['Month'] = returns['return_date'].dt.month

# Aggregate return quantity
returns_monthly = returns.groupby(['Year', 'Month'])
['quantity'].sum().unstack()

# Plot heatmap
plt.figure(figsize=(12, 6))
```

```
sns.heatmap(returns_monthly, cmap="coolwarm", annot=True, fmt=".0f",
linewidths=0.5)

plt.title('Monthly Returns Heatmap', fontsize=16, fontweight='bold')
plt.xlabel('Month')
plt.ylabel('Year')
plt.show()
```

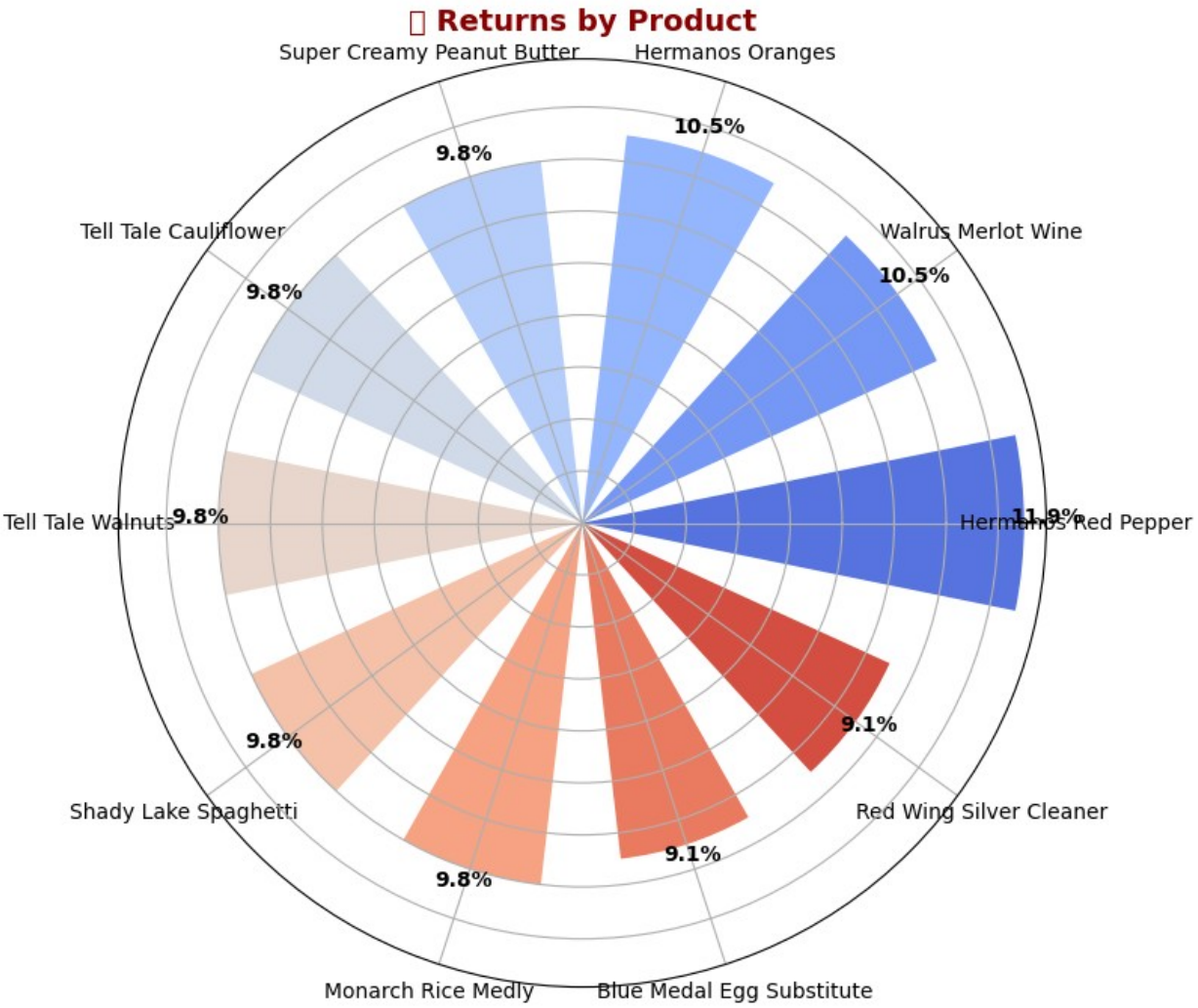


Radial Bar Chart – Returns by Product

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Prepare data
top_returns = returns_summary.nlargest(10, 'quantity')
angles = np.linspace(0, 2 * np.pi, len(top_returns), endpoint=False)
top_returns['percentage'] = top_returns['quantity'] /
top_returns['quantity'].sum() * 100

# Create polar plot
fig, ax = plt.subplots(figsize=(8, 8), subplot_kw={'projection':
'polar'})
bars = ax.bar(angles, top_returns['quantity'], width=0.4,
color=sns.color_palette("coolwarm", len(top_returns)))
```

```
# Add percentage labels
for bar, angle, percent in zip(bars, angles,
top_returns['percentage']):
    ax.text(angle, bar.get_height() * 1.05, f"{percent:.1f}%",
ha='center', fontsize=10, fontweight="bold")

# Set labels & title
ax.set_xticks(angles)
ax.set_xticklabels(top_returns['product_name'], fontsize=10)
ax.set_yticklabels([])
plt.title("□ Returns by Product", fontsize=14, fontweight="bold",
color="darkred")

plt.show()
```

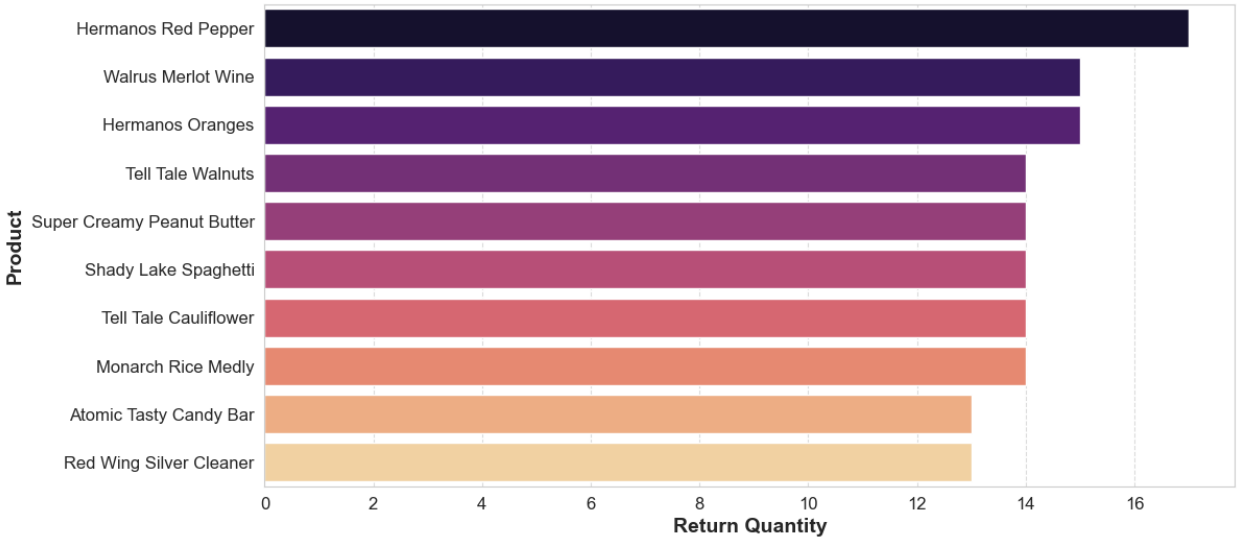
```
D:\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarning:
Glyph 128202 (\N{BAR CHART}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



Identify most returned products

```
returns_summary = returns.groupby('product_id')
['quantity'].sum().reset_index()
returns_summary = returns_summary.merge(products, on='product_id',
how='left')
returns_summary = returns_summary.sort_values(by='quantity',
ascending=False).head(10)

# Unique and Modern Returns Visualization - Horizontal Bar Chart
plt.figure(figsize=(12,6))
sns.set_style("whitegrid")
```

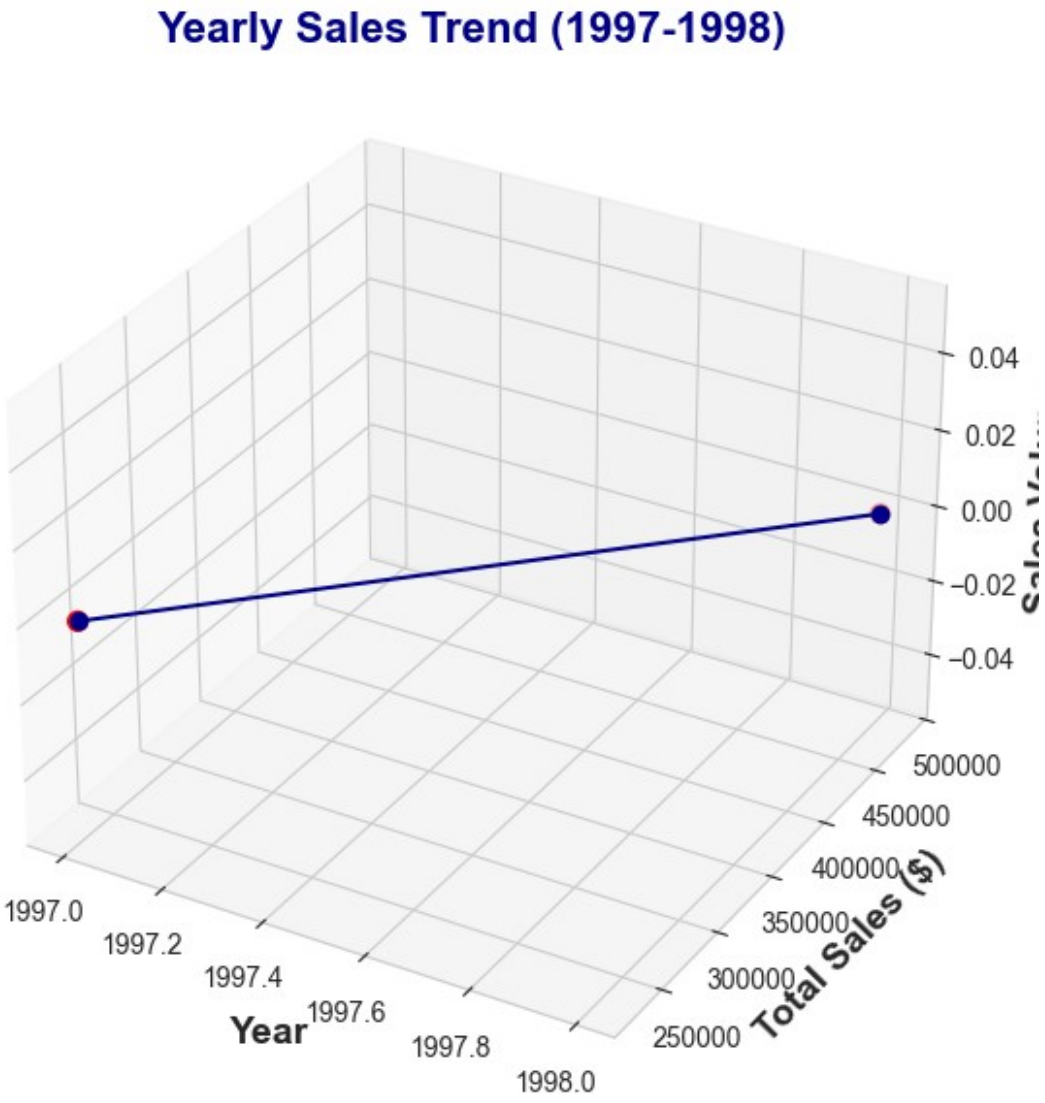
```
ax = sns.barplot(x=returns_summary['quantity'],
y=returns_summary['product_name'],
hue=returns_summary['product_name'], palette='magma', legend=False)

# Customizing labels & title
plt.xlabel('Return Quantity', fontsize=14, fontweight='bold')
plt.ylabel('Product', fontsize=14, fontweight='bold')
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.grid(axis='x', linestyle='--', alpha=0.7)

# Show the plot
plt.show()
```



Sales Visualization - 3D Line Chart

```
fig = plt.figure(figsize=(12, 7))
ax = fig.add_subplot(111, projection='3d')

years = sales_by_year.index
sales = sales_by_year.values

ax.plot(years, sales, zs=0, zdir='z', label='Yearly Sales',
color='darkblue', marker='o', linestyle='-')
ax.scatter(years, sales, zs=0, zdir='z', color='red', s=50)

# Customizing labels & title
ax.set_title('Yearly Sales Trend (1997-1998)', fontsize=16,
fontweight='bold', color='darkblue')
ax.set_xlabel('Year', fontsize=14, fontweight='bold')
ax.set_ylabel('Total Sales ($)', fontsize=14, fontweight='bold')
```

```
ax.set_zlabel('Sales Volume', fontsize=14, fontweight='bold')
ax.grid(True)

# Show the plot
plt.show()
```




```
!pip install scikit-learn

Requirement already satisfied: scikit-learn in d:\lib\site-packages
(1.6.1)
Requirement already satisfied: numpy>=1.19.5 in d:\lib\site-packages
(from scikit-learn) (2.2.3)
Requirement already satisfied: scipy>=1.6.0 in d:\lib\site-packages
(from scikit-learn) (1.15.2)
Requirement already satisfied: joblib>=1.2.0 in d:\lib\site-packages
```

```
(from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in d:\lib\site-
packages (from scikit-learn) (3.5.0)


[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: D:\python.exe -m pip install --upgrade pip
```

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error
```

Combine transaction datasets

```
transactions = pd.concat([transactions_1997, transactions_1998])
transactions
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1997  12/31/1996         869         3449         6

1              1/1/1997  12/31/1996        1472         3449         6

2              1/1/1997  12/28/1996          76         3449         6

3              1/1/1997  12/26/1996         320         3449         6

4              1/1/1997  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

182878       12/30/1998  12/29/1998        1521         7197        11

182879       12/30/1998  12/23/1998        1536         5223        10

182880       12/30/1998  12/23/1998        1542         8077        10

182881       12/30/1998  12/28/1998        1544         4485        10

182882       12/30/1998  12/29/1998        1549         5223        10


        quantity
0              5
1              3
2              4
3              3
4              4
```

```
...          ...
182878         3
182879         2
182880         4
182881         2
182882         3

[269720 rows x 6 columns]
```

Combine transaction datasets

```
transactions = pd.concat([transactions_1997, transactions_1998])
transactions
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0              1/1/1997  12/31/1996         869         3449         6

1              1/1/1997  12/31/1996        1472         3449         6

2              1/1/1997  12/28/1996          76         3449         6

3              1/1/1997  12/26/1996         320         3449         6

4              1/1/1997  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

182878       12/30/1998  12/29/1998        1521         7197        11

182879       12/30/1998  12/23/1998        1536         5223        10

182880       12/30/1998  12/23/1998        1542         8077        10

182881       12/30/1998  12/28/1998        1544         4485        10

182882       12/30/1998  12/29/1998        1549         5223        10


        quantity
0              5
1              3
2              4
3              3
4              4
...          ...
182878         3
182879         2
182880         4
```

```
182881         2
182882         3

[269720 rows x 6 columns]
```

Step 6: One-Hot Encoding for Categorical Variables

```
df = pd.get_dummies(df, columns=['store_type', 'product_brand'],
drop_first=True)
df
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0            1997-01-01  12/31/1996         869         3449         6

1            1997-01-01  12/31/1996        1472         3449         6

2            1997-01-01  12/28/1996          76         3449         6

3            1997-01-01  12/26/1996         320         3449         6

4            1997-01-01  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       1998-12-30  12/29/1998        1521         7197        11

269716       1998-12-30  12/23/1998        1536         5223        10

269717       1998-12-30  12/23/1998        1542         8077        10

269718       1998-12-30  12/28/1998        1544         4485        10

269719       1998-12-30  12/29/1998        1549         5223        10


        quantity                        product_name  product_sku  \
0              5          Nationeel Grape Fruit Roll  52382137179
1              3             Fort West Fudge Cookies  37276054024
2              4           Red Spade Sliced Chicken  62054644227
3              3          Excellent Cranberry Juice  36570182442
4              4              Washington Cream Soda  64412155747
...          ...                                ...          ...
269715         3  Top Measure White Zinfandel Wine  19118239283
269716         2                    CDR White Sugar  92759070992
269717         4                  CDR Tomato Sauce  95931351780
269718         2                CDR Grape Preserves  65897241234
```

```
269719         3                            CDR Salt  44236547350

        product_retail_price  product_cost  ...
product_brand_Thresher  \
0                       2.12          0.91  ...
False
1                       2.20          0.90  ...
False
2                       1.69          0.69  ...
False
3                       3.26          1.08  ...
False
4                       3.64          1.64  ...
False
...                      ...           ...  ...                       ..
.
269715                  0.70          0.26  ...
False
269716                  1.33          0.48  ...
False
269717                  0.95          0.41  ...
False
269718                  1.46          0.57  ...
False
269719                  1.52          0.74  ...
False

        product_brand_Tip Top  product_brand_Token  product_brand_Top
Measure  \
0                       False                False
False
1                       False                False
False
2                       False                False
False
3                       False                False
False
4                       False                False
False
...                       ...                  ...
...
269715                  False                False
True
269716                  False                False
False
269717                  False                False
False
269718                  False                False
False
```

```
269719                   False                False
False

       product_brand_Toretti product_brand_Toucan product_brand_Tri-
State  \
0                      False                False
False
1                      False                False
False
2                      False                False
False
3                      False                False
False
4                      False                False
False
...                      ...                  ...
...
269715                 False                False
False
269716                 False                False
False
269717                 False                False
False
269718                 False                False
False
269719                 False                False
False

       product_brand_Urban product_brand_Walrus
product_brand_Washington
0                    False                False
False
1                    False                False
False
2                    False                False
False
3                    False                False
False
4                    False                False
True
...                    ...                  ...                      .
..
269715               False                False
False
269716               False                False
False
269717               False                False
False
269718               False                False
```

```
False
269719               False                False
False

[269720 rows x 140 columns]
```

Step 7: Define Features and Target Variable

```
# Create 'Sales' column by multiplying 'Quantity_Sold' and 'Price'
df['Sales'] = df['quantity'] * df['product_cost']
df
```

```
       transaction_date  stock_date  product_id  customer_id  store_id
\
0            1997-01-01  12/31/1996         869         3449         6

1            1997-01-01  12/31/1996        1472         3449         6

2            1997-01-01  12/28/1996          76         3449         6

3            1997-01-01  12/26/1996         320         3449         6

4            1997-01-01  12/25/1996           4         3449         6

...                 ...         ...         ...          ...       ...

269715       1998-12-30  12/29/1998        1521         7197        11

269716       1998-12-30  12/23/1998        1536         5223        10

269717       1998-12-30  12/23/1998        1542         8077        10

269718       1998-12-30  12/28/1998        1544         4485        10

269719       1998-12-30  12/29/1998        1549         5223        10


        quantity                        product_name  product_sku  \
0              5          Nationeel Grape Fruit Roll  52382137179
1              3             Fort West Fudge Cookies  37276054024
2              4           Red Spade Sliced Chicken  62054644227
3              3          Excellent Cranberry Juice  36570182442
4              4              Washington Cream Soda  64412155747
...          ...                                ...          ...
269715         3  Top Measure White Zinfandel Wine  19118239283
269716         2                    CDR White Sugar  92759070992
269717         4                  CDR Tomato Sauce  95931351780
269718         2                CDR Grape Preserves  65897241234
269719         3                            CDR Salt  44236547350
```

```
        product_retail_price  product_cost  ...  product_brand_Tip Top
\
0                       2.12          0.91  ...                   False

1                       2.20          0.90  ...                   False

2                       1.69          0.69  ...                   False

3                       3.26          1.08  ...                   False

4                       3.64          1.64  ...                   False

...                      ...           ...  ...                     ...

269715                  0.70          0.26  ...                   False

269716                  1.33          0.48  ...                   False

269717                  0.95          0.41  ...                   False

269718                  1.46          0.57  ...                   False

269719                  1.52          0.74  ...                   False


        product_brand_Token  product_brand_Top Measure
product_brand_Toretti  \
0                     False                      False
False
1                     False                      False
False
2                     False                      False
False
3                     False                      False
False
4                     False                      False
False
...                     ...                        ...
...
269715                False                       True
False
269716                False                      False
False
269717                False                      False
False
269718                False                      False
False
269719                False                      False
False
```

```
       product_brand_Toucan product_brand_Tri-State
product_brand_Urban  \
0                     False                     False
False
1                     False                     False
False
2                     False                     False
False
3                     False                     False
False
4                     False                     False
False
...                     ...                       ...                   ..
.
269715                False                     False
False
269716                False                     False
False
269717                False                     False
False
269718                False                     False
False
269719                False                     False
False

       product_brand_Walrus product_brand_Washington Sales
0                     False                      False  4.55
1                     False                      False  2.70
2                     False                      False  2.76
3                     False                      False  3.24
4                     False                       True  6.56
...                     ...                        ...   ...
269715                False                      False  0.78
269716                False                      False  0.96
269717                False                      False  1.64
269718                False                      False  1.14
269719                False                      False  2.22

[269720 rows x 141 columns]
```

Define Features and Target Variable

```
X = df.drop(columns=['Sales'])  # Independent variables
y = df['Sales']  # Target variable
y

0         4.55
1         2.70
```

```
2         2.76
3         3.24
4         6.56
          ...
269715    0.78
269716    0.96
269717    1.64
269718    1.14
269719    2.22
Name: Sales, Length: 269720, dtype: float64
```

Step 8: Split Data into Train & Test Sets

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
X_train, X_test, y_train, y_test
```

```
(       transaction_date  stock_date  product_id  customer_id
store_id  \
 142214       1998-04-25   4/22/1998        1120         1136
24
 198444       1998-08-17   8/11/1998         681         4419
12
 162017       1998-06-05    6/2/1998        1176         4693
3
 406          1997-01-03  12/29/1996        1443         5736
7
 26003        1997-04-18   4/14/1997         301         1235
11
 ...                 ...         ...         ...          ...       ..
.
 119879       1998-03-08    3/6/1998         913         9009
19
 259178       1998-12-11   12/6/1998         472         5348
12
 131932       1998-04-03    4/1/1998         800         5542
16
 146867       1998-05-04   4/28/1998        1530         6559
3
 121958       1998-03-11   3/10/1998        1130         7152
12

         quantity               product_name  product_sku  \
 142214         3  Tri-State Elephant Garlic  50807433724
 198444         2            Gorilla 1% Milk  22163082957
 162017         5         Horatio Beef Jerky  90692868828
 406            4         Hermanos Firm Tofu  74774398207
 26003          3         Super Apple Butter  76410352587
```

```
 ...          ...                          ...           ...
 119879         2       BBB Best Brown Sugar  88933951258
 259178         4          Red Wing Scissors  57052045464
 131932         3               Ebony Onions  81383071541
 146867         5   Modell Blueberry Muffins  40483739972
 121958         4           Tri-State Squash  22650697827

         product_retail_price  product_cost  ...
product_brand_Thresher  \
 142214                  1.71          0.86  ...
False
 198444                  2.30          1.10  ...
False
 162017                  1.71          0.84  ...
False
 406                     2.17          0.87  ...
False
 26003                   0.53          0.23  ...
False
 ...                      ...           ...  ...                       .
..
 119879                  1.78          0.62  ...
False
 259178                  0.82          0.37  ...
False
 131932                  3.26          1.43  ...
False
 146867                  1.27          0.51  ...
False
 121958                  1.55          0.56  ...
False

         product_brand_Tip Top  product_brand_Token  product_brand_Top
Measure  \
 142214                  False                False
False
 198444                  False                False
False
 162017                  False                False
False
 406                     False                False
False
 26003                   False                False
False
 ...                       ...                  ...
...
 119879                  False                False
False
 259178                  False                False
```

```
False
 131932                   False                False
False
 146867                   False                False
False
 121958                   False                False
False

        product_brand_Toretti product_brand_Toucan product_brand_Tri-
State  \
 142214                 False                False
True
 198444                 False                False
False
 162017                 False                False
False
 406                    False                False
False
 26003                  False                False
False
 ...                      ...                  ...
...
 119879                 False                False
False
 259178                 False                False
False
 131932                 False                False
False
 146867                 False                False
False
 121958                 False                False
True

        product_brand_Urban product_brand_Walrus
product_brand_Washington
 142214               False                False
False
 198444               False                False
False
 162017               False                False
False
 406                  False                False
False
 26003                False                False
False
 ...                    ...                  ...
...
 119879               False                False
False
```

```
 259178                False                False
False
 131932                False                False
False
 146867                False                False
False
 121958                False                False
False

 [215776 rows x 140 columns],
        transaction_date  stock_date  product_id  customer_id
store_id  \
 169304       1998-06-20   6/17/1998         194         5138
24
 72719        1997-11-11   11/8/1997          73         8477
3
 52213        1997-08-12    8/7/1997         370         3921
23
 142944       1998-04-27   4/25/1998         152         3682
13
 208593       1998-09-09    9/7/1998         947         6844
11
 ...                 ...         ...         ...          ...       ..
.
 199383       1998-08-19   8/15/1998         173         3884
2
 113629       1998-02-22   2/19/1998         811         1431
19
 86160        1997-12-28  12/27/1997        1467         3418
3
 243456       1998-11-16  11/10/1998         416         8846
4
 64438        1997-10-07   10/1/1997        1425         4320
6

         quantity                 product_name  product_sku  \
 169304         3           High Top Firm Tofu  31548164486
 72719          2     Red Spade Sliced Turkey  86320835947
 52213          3              Carlson 1% Milk  24180286526
 142944         3                Denny Tissues  15871854424
 208593         2    Fabulous Cranberry Juice  57837085127
 ...          ...                          ...          ...
 199383         2           High Top Asparagus  72946912646
 113629         4             Ebony Baby Onion  48186540007
 86160          3   Fort West BBQ Potato Chips  81177968382
 243456         2  Big Time Frozen Cauliflower  27328963875
 64438          4               Hermanos Beets  95661703944

         product_retail_price  product_cost  ...
```

```
product_brand_Thresher  \
 169304                   3.36          1.28  ...
False
 72719                    2.33          0.82  ...
False
 52213                    0.92          0.36  ...
False
 142944                   2.89          1.04  ...
False
 208593                   0.65          0.22  ...
False
 ...                       ...           ...  ...                        .
..
 199383                   1.43          0.60  ...
False
 113629                   2.97          0.95  ...
False
 86160                    3.88          1.67  ...
False
 243456                   3.32          1.33  ...
False
 64438                    2.18          0.72  ...
False

         product_brand_Tip Top  product_brand_Token  product_brand_Top
Measure  \
 169304                   False                False
False
 72719                    False                False
False
 52213                    False                False
False
 142944                   False                False
False
 208593                   False                False
False
 ...                        ...                  ...
...
 199383                   False                False
False
 113629                   False                False
False
 86160                    False                False
False
 243456                   False                False
False
 64438                    False                False
False
```

```
        product_brand_Toretti product_brand_Toucan product_brand_Tri-
State  \
 169304                 False                False
False
 72719                  False                False
False
 52213                  False                False
False
 142944                 False                False
False
 208593                 False                False
False
 ...                      ...                  ...
...
 199383                 False                False
False
 113629                 False                False
False
 86160                  False                False
False
 243456                 False                False
False
 64438                  False                False
False

        product_brand_Urban product_brand_Walrus
product_brand_Washington
 169304               False                False
False
 72719                False                False
False
 52213                False                False
False
 142944               False                False
False
 208593               False                False
False
 ...                    ...                  ...
...
 199383               False                False
False
 113629               False                False
False
 86160                False                False
False
 243456               False                False
False
 64438                False                False
False
```

```
[53944 rows x 140 columns],
142214    2.58
198444    2.20
162017    4.20
406       3.48
26003     0.69
          ...
119879    1.24
259178    1.48
131932    4.29
146867    2.55
121958    2.24
Name: Sales, Length: 215776, dtype: float64,
169304    3.84
72719     1.64
52213     1.08
142944    3.12
208593    0.44
          ...
199383    1.20
113629    3.80
86160     5.01
243456    2.66
64438     2.88
Name: Sales, Length: 53944, dtype: float64)
```

Step 9: Scale the Data (optional but improves performance)

```
from sklearn.preprocessing import StandardScaler

# Ensure transaction_date is dropped
X_train = X_train.drop(columns=['transaction_date'], errors='ignore')
X_test = X_test.drop(columns=['transaction_date'], errors='ignore')

# Select only numeric columnss
numeric_cols = X_train.select_dtypes(include=['number']).columns

# Initialize and apply StandardScaler only on numeric columns
scaler = StandardScaler()
X_train[numeric_cols] = scaler.fit_transform(X_train[numeric_cols])
X_test[numeric_cols] = scaler.transform(X_test[numeric_cols])

# Display results
print(X_train.head())
print(X_test.head())
```

```
        stock_date  product_id  customer_id  store_id  quantity  \
142214   4/22/1998    0.751226    -1.372643  1.774284 -0.107547
198444   8/11/1998   -0.229220    -0.241693 -0.116270 -1.305402
162017    6/2/1998    0.876295    -0.147304 -1.534186  2.288163
406     12/29/1996    1.472603     0.211996 -0.904001  1.090308
26003    4/14/1997   -1.077898    -1.338539 -0.273816 -0.107547

                     product_name  product_sku
product_retail_price  \
142214  Tri-State Elephant Garlic    -0.185504             -0.438748

198444            Gorilla 1% Milk    -1.296063              0.195146

162017         Horatio Beef Jerky     1.360879             -0.438748

406            Hermanos Firm Tofu     0.743710              0.055475

26003          Super Apple Butter     0.807137             -1.706537


        product_cost  product_weight  ...  product_brand_Thresher  \
142214      0.013650        0.498212  ...                   False
198444      0.608556       -0.944538  ...                   False
162017     -0.035926       -0.901277  ...                   False
406         0.038437       -1.478809  ...                   False
26003      -1.547979       -0.648201  ...                   False

        product_brand_Tip Top  product_brand_Token product_brand_Top
Measure  \
142214                  False                False
False
198444                  False                False
False
162017                  False                False
False
406                     False                False
False
26003                   False                False
False

       product_brand_Toretti product_brand_Toucan product_brand_Tri-
State  \
142214                 False                False
True
198444                 False                False
False
162017                 False                False
False
406                    False                False
False
26003                  False                False
```

```
False

       product_brand_Urban product_brand_Walrus
product_brand_Washington
142214               False                False
False
198444               False                False
False
162017               False                False
False
406                  False                False
False
26003                False                False
False

[5 rows x 139 columns]
       stock_date  product_id  customer_id  store_id  quantity  \
169304  6/17/1998   -1.316868     0.005993  1.774284 -0.107547
72719   11/8/1997   -1.587105     1.156234 -1.534186 -1.305402
52213    8/7/1997   -0.923796    -0.413247  1.616738 -0.107547
142944  4/25/1998   -1.410669    -0.495580  0.041276 -0.107547
208593   9/7/1998    0.364855     0.593687 -0.273816 -1.305402

                    product_name  product_sku  product_retail_price  \
169304        High Top Firm Tofu    -0.932198              1.334008
72719    Red Spade Sliced Turkey     1.191372              0.227378
52213             Carlson 1% Milk    -1.217855             -1.287522
142944              Denny Tissues    -1.539978              0.829041
208593  Fabulous Cranberry Juice     0.087040             -1.577609

        product_cost  product_weight  ...  product_brand_Thresher  \
169304      1.054735       -0.475157  ...                   False
72719      -0.085501        1.147124  ...                   False
52213      -1.225738       -1.550189  ...                   False
142944      0.459829        0.736146  ...                   False
208593     -1.572766        0.476581  ...                   False

        product_brand_Tip Top  product_brand_Token product_brand_Top
Measure  \
169304                  False                False
False
72719                   False                False
False
52213                   False                False
False
142944                  False                False
False
208593                  False                False
False
```

```
       product_brand_Toretti product_brand_Toucan product_brand_Tri-
State  \
169304                  False                 False
False
72719                   False                 False
False
52213                   False                 False
False
142944                  False                 False
False
208593                  False                 False
False

       product_brand_Urban product_brand_Walrus
product_brand_Washington
169304                False                 False
False
72719                 False                 False
False
52213                 False                 False
False
142944                False                 False
False
208593                False                 False
False

[5 rows x 139 columns]

print("Columns in DataFrame:", df.columns)

target_column = "Sales"
if target_column not in df.columns:
    raise ValueError(f"Column '{target_column}' not found in
dataset!")

X = df.drop(columns=[target_column])
y = df[target_column]

X = pd.get_dummies(X).fillna(X.mean())

if X.empty or y.empty:
    raise ValueError("Feature matrix (X) or target variable (y) is
empty after preprocessing!")

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

```python
y_pred = model.predict(X_test)

print("First 5 Predictions:", y_pred[:5])

print("□ Model Performance:")
print("MAE:", mean_absolute_error(y_test, y_pred))
print("MSE:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))


Columns in DataFrame: Index(['transaction_date', 'stock_date',
'product_id', 'customer_id',
       'store_id', 'quantity', 'product_name', 'product_sku',
       'product_retail_price', 'product_cost',
       ...
       'product_brand_Tip Top', 'product_brand_Token',
       'product_brand_Top Measure', 'product_brand_Toretti',
       'product_brand_Toucan', 'product_brand_Tri-State',
       'product_brand_Urban', 'product_brand_Walrus',
       'product_brand_Washington', 'Sales'],
      dtype='object', length=141)
```