```
In [50]:

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

Out[25]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Cou
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	Ur King
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Ur King
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Ur King
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Ur King
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Ur King
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.0	Fra
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	Fra
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	Fra
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	Fra
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	Fra

541909 rows × 8 columns

In [26]: 1 df.head()

Out[26]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

In [26]: 1 df.head()

Out[26]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

In [27]:

1 df.tail()

Out[27]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Cou
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.0	Fra
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	Fra
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	Fra
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	Fra
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	Fra

In [28]: 1 df.shape

Out[28]: (541909, 8)

In [29]: 1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908 Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype
0	InvoiceNo	541909 non-null	object
1	StockCode	541909 non-null	object
2	Description	540455 non-null	object
3	Quantity	541909 non-null	int64
4	InvoiceDate	541909 non-null	datetin

4 InvoiceDate 541909 non-null datetime64[ns] 5 UnitPrice 541909 non-null float64

```
In [29]:
            1 df.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 541909 entries, 0 to 541908
          Data columns (total 8 columns):
                              Non-Null Count
           #
               Column
                                                Dtype
           0
               InvoiceNo
                              541909 non-null
                                                object
                              541909 non-null object
           1
               StockCode
           2
               Description 540455 non-null object
                              541909 non-null int64
               Quantity
           4
               InvoiceDate 541909 non-null datetime64[ns]
                              541909 non-null float64
           5
               UnitPrice
           6
               CustomerID
                              406829 non-null
                                                float64
               Country
                              541909 non-null
                                                object
          dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
          memory usage: 33.1+ MB
In [30]:
            1 df.describe()
Out[30]:
                      Quantity
                                   UnitPrice
                                               CustomerID
           count
                 541909.000000
                               541909.000000
                                            406829.000000
                      9.552250
                                    4.611114
                                              15287.690570
           mean
             std
                    218.081158
                                   96.759853
                                               1713.600303
            min
                  -80995.000000
                               -11062.060000
                                              12346.000000
            25%
                      1.000000
                                    1.250000
                                              13953.000000
            50%
                      3.000000
                                    2.080000
                                              15152.000000
            75%
                     10.000000
                                    4.130000
                                              16791.000000
                  80995.000000
                                             18287 000000
            max
                                38970.000000
In [31]:
            1 df.duplicated().sum()
Out[31]: 5268
In [32]:
            1 df[df.duplicated()]
Out[32]:
                  InvoiceNo StockCode
                                       Description Quantity InvoiceDate UnitPrice CustomerID (
                                            UNION
                                        JACK FLAG
                                                            2010-12-01
              517
                     536409
                                21866
                                                                           1.25
                                                                                    17908.0
                                         LUGGAGE
                                                               11:45:00
                                              TAG
                                            HAND
                                         WARMER
                                                            2010-12-01
              527
                     536409
                                22866
                                          SCOTTY
                                                                           2.10
                                                                                    17908.0
                                                               11:45:00
                                             DOG
                                          DESIGN
                                         SET 2 TEA
                                                            2010-12-01
                                         TOWELS I
              537
                     536409
                                22900
                                                                           2.95
                                                                                    17908.0
                                            I OVE
                                                               11:45:00
                                          LONDON
                                          SCOTTIE
                                         DOG HOT
                                                            2010-12-01
                                                                                    17908.0 k
              539
                     536409
                                 22111
                                                                           4.95
                                           WATER
                                                               11:45:00
            1 df.isnull().sum()
Out[33]: InvoiceNo
                                0
          StockCode
                                0
          Description
                             1454
          Quantity
                                0
          InvoiceDate
                                0
          UnitPrice
                                0
          CustomerID
                           135080
          Country
                                0
          dtype: int64
```

4 165 115

/ !!!!

```
1 df.isnull().sum()
Out[33]: InvoiceNo
                                  0
           StockCode
                                  0
           Description
                              1454
           Quantity
                                  a
           InvoiceDate
                                  0
           UnitPrice
                                  0
                            135080
           CustomerID
           Country
                                  0
           dtype: int64
            1 df["Description"].fillna("No Description", inplace=True)
In [41]:
                df["CustomerID"].fillna(-1, inplace=True)
               df
Out[41]:
                   InvoiceNo StockCode
                                          Description Quantity InvoiceDate UnitPrice CustomerID Cou
                                               WHITE
                                            HANGING
                                                                2010-12-01
                                                                                                   Ur
                 0
                      536365
                                 85123A
                                            HEART T-
                                                                                2.55
                                                                                         17850.0
                                                                                                 King
                                                                  08:26:00
                                               LIGHT
                                             HOLDER
                                              WHITE
                                                                2010-12-01
                                                                                                   Ur
                      536365
                                  71053
                                              METAL
                                                                                3.39
                                                                                         17850.0
                                                                                                 King
                                                                  08:26:00
                                            LANTERN
                                              CREAM
                                               CUPID
                                                                2010-12-01
                                                                                                   Ur
                 2
                      536365
                                 84406B
                                             HEARTS
                                                             8
                                                                                2.75
                                                                                         17850.0
                                                                                                 King
                                                                  08:26:00
                                                COAT
                                             HANGER
                                             KNITTED
                                         UNION FLAG
                                                                2010-12-01
                                                                                                   Ur
                                 84029G
                      536365
                                                                                3.39
                                                                                         17850.0
                                          HOT WATER
                                                                  08:26:00
                                                                                                 King
                                             BOTTLE
                                                RED
                                             WOOLLY
                                                                2010-12-01
                                                                                                   Ur
                                                                                         17850.0
                      536365
                                 84029E
                                                             6
                                              HOTTIE
                                                                                3.39
                                                                                                 King
                                                                  08:26:00
                                              WHITE
                                              HEART.
                                          PACK OF 20
                                                                 2011-12-09
            541904
                      581587
                                  22613
                                           SPACEBOY
                                                            12
                                                                                0.85
                                                                                         12680.0
                                                                   12:50:00
                                            NAPKINS
                                          CHILDREN'S
                                                                2011-12-09
            541905
                      581587
                                  22899
                                              APRON
                                                             6
                                                                                2.10
                                                                                         12680.0
                                                                                                   Fra
                                                                  12:50:00
                                          DOLLY GIRL
                                          CHILDRENS
                                                                 2011-12-09
            541906
                      581587
                                  23254
                                            CUTLERY
                                                                                4.15
                                                                                         12680.0
                                                                                                   Fra
                                                                   12:50:00
                                          DOLLY GIRL
                                          CHILDRENS
                                                                 2011-12-09
                                            CUTLERY
            541907
                      581587
                                  23255
                                                                                4.15
                                                                                         12680.0
                                                                                                   Fra
                                             CIRCUS
                                                                   12:50:00
                                             PARADE
                                          BAKING SET
                                                                2011-12-09
            541908
                      581587
                                  22138
                                             9 PIECE
                                                                                4.95
                                                                                         12680.0
                                                                                                  Fra
                                                                  12:50:00
                                         RETROSPOT
           541909 rows × 8 columns
          1 df.info
In [43]:
                                                                                                 Qut[43]:
           <br/>
<br/>
dbundf.mozetlhumahsDataFrame.info of
                                                        InvoiceNo StockCode
           Description Quantity \
                                  85123ACodWHITEDUANGINGIUFART JULIAGETYHOLDERvoiceDate6
           9ndex(['In√39386No',
Out[42]:
                                                 'Country' | WHITE METAL LANTERN
CREAM CUPID HEARTS COAT HANGER
                                                                                              6
           1
                                  '21053merID',
                   'Un$₹₱ネ₠₺e',
                                                                                              8
           2
                  dtyp=36363ect 94406B
                                  84029G
                                           KNITTED UNION FLAG HOT WATER BOTTLE
           3
                      536365
                                                                                              6
           4
                      536365
                                  84029E
                                                 RED WOOLLY HOTTIE WHITE HEART.
                                                                                              6
                                                    PACK OF 20 SPACEBOY NAPKINS
           541904
                      581587
                                   22613
                                                                                             12
           541905
                      581587
                                   22899
                                                   CHILDREN'S APRON DOLLY GIRL
                                                                                              6
```

1/908.0 K

4.90

11:45:00

ᲔᲐ७4UᲧ

541906

581587

23254

CHILDRENS CUTLERY DOLLY GIRL

4

ZZIII

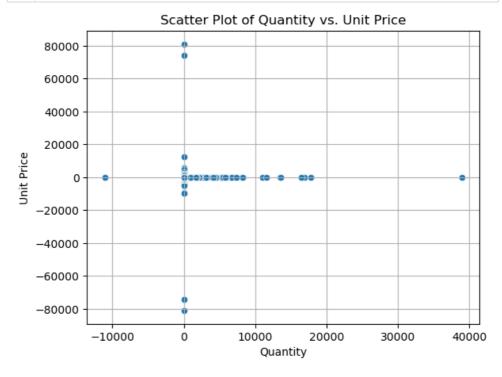
WATER

```
In [43]: 1 df.info
Out[42]: <bbu noter.noceofthwardh sDataFrame.info of
                                                  InvoiceNo StockCode
         Description Quantity \
Out[42]: Index(['Intotention',
                              85132ACodWHJTEDUSNGINGIURART JūLARGUTyHOLDERvoiceDate6
                              'Clatamerid', 'Country'] WHITE METAL LANTERN
'&4406B CREAM CUPID HEARTS COAT HANGER
                                                                                    6
                 'Un146795e',
                dtyp=36365ect 84406B
          2
                                                                                    8
                    536365
          3
                              84029G KNITTED UNION FLAG HOT WATER BOTTLE
                                                                                    6
                                           RED WOOLLY HOTTIE WHITE HEART.
                    536365
                              84029E
          4
                                                                                    6
                     . . .
                                . . .
          541904
                    581587
                                               PACK OF 20 SPACEBOY NAPKINS
                               22613
                                                                                   12
                                             CHILDREN'S APRON DOLLY GIRL
          541905
                    581587
                               22899
                                                                                    6
          541906
                                             CHILDRENS CUTLERY DOLLY GIRL
                    581587
                               23254
                                                                                    4
          541907
                    581587
                               23255
                                           CHILDRENS CUTLERY CIRCUS PARADE
                                                                                    4
          541908
                    581587
                               22138
                                             BAKING SET 9 PIECE RETROSPOT
                                                                                    3
                         InvoiceDate UnitPrice CustomerID
                                                                      Country
         0
                 2010-12-01 08:26:00
                                         2.55
                                                 17850.0 United Kingdom
                                                  United Kingdom 17850.0 United Kingdom 17850.0 United V
                                            3.39
         1
                 2010-12-01 08:26:00
          2
                 2010-12-01 08:26:00
                                           2.75
                 2010-12-01 08:26:00
                                           3.39
          3
                                                     17850.0 United Kingdom
          4
                 2010-12-01 08:26:00
                                           3.39
                                            . . .
                                                        . . .
          541904 2011-12-09 12:50:00
                                            0.85 12680.0
                                                                      France
          541905 2011-12-09 12:50:00
                                           2.10 12680.0
                                                                      France
          541906 2011-12-09 12:50:00
                                           4.15
                                                    12680.0
                                                                      France
          541907 2011-12-09 12:50:00
                                           4.15
                                                     12680.0
                                                                      France
          541908 2011-12-09 12:50:00
                                           4.95
                                                     12680.0
                                                                       France
          [541909 rows x 8 columns]>
In [44]:
          1 df['Description'].value_counts()
Out[44]: WHITE HANGING HEART T-LIGHT HOLDER
                                                  2369
         REGENCY CAKESTAND 3 TIER
                                                  2200
          JUMBO BAG RED RETROSPOT
                                                  2159
          PARTY BUNTING
                                                  1727
          LUNCH BAG RED RETROSPOT
                                                  1638
         Missing
          historic computer difference?....se
                                                     1
         DUSTY PINK CHRISTMAS TREE 30CM
                                                     1
          WRAP BLUE RUSSIAN FOLKART
                                                     1
          PINK BERTIE MOBILE PHONE CHARM
                                                     1
         Name: Description, Length: 4224, dtype: int64
```

visualization

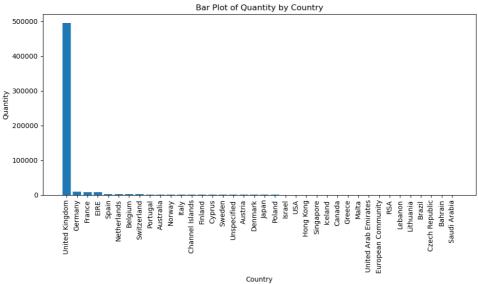
```
In [56]: 1 #scatter plot
2 
3 sns.scatterplot(x='UnitPrice', y='Quantity',data=df)
4 plt.title('Scatter Plot of Quantity vs. Unit Price')
5 plt.xlabel('Quantity')
6 plt.ylabel('Unit Price')
7 plt.grid(True)
```

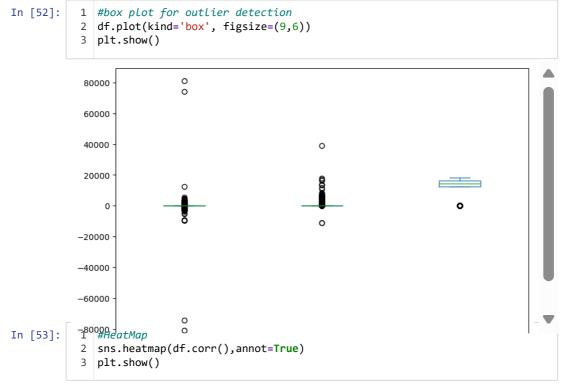
visualization



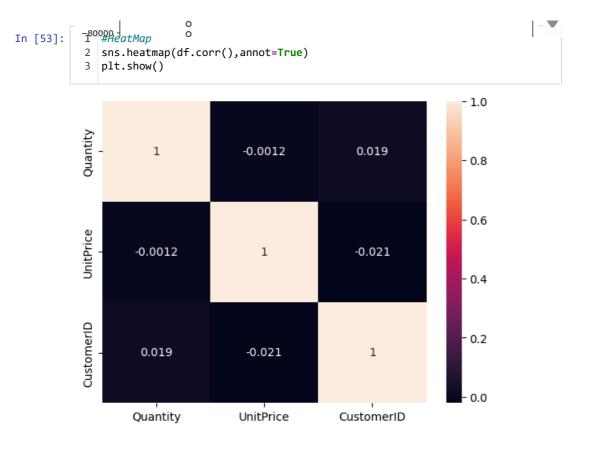
```
In [58]: 1 #Barplot
2
3 country_counts = df['Country'].value_counts()
4 countries = country_counts.index
5 quantity_values = country_counts.values
6
7 plt.figure(figsize=(10, 6))
8 plt.bar(countries, quantity_values)
9 plt.title('Bar Plot of Quantity by Country')
10 plt.xlabel('Country')
11 plt.ylabel('Quantity')
12 plt.xticks(rotation=90)
13 plt.tight_layout()
```





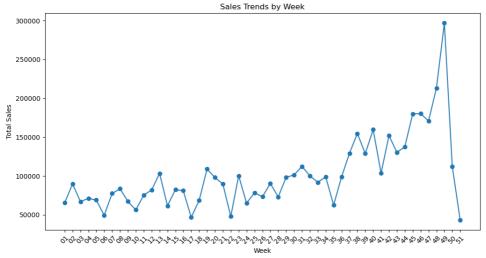






```
Analyze sales trends over time
In [70]:
           1 df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate'])
In [78]:
              df['Month'] = df['InvoiceDate'].dt.month
              df['Hour'] = df['InvoiceDate'].dt.hour
           4 | sales_by_month = df.groupby('Month')['Quantity'].sum()
           1 # Plot sales by month
In [79]:
           2
              plt.figure(figsize=(12, 6))
              plt.plot(sales_by_month.index, sales_by_month.values, marker='o')
             plt.title('Sales Trends by Month')
             plt.xlabel('Month')
              plt.ylabel('Total Sales')
           8
              plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul'
           9
              plt.show()
          10
                                               Sales Trends by Month
            700000
            600000
          हूँ। #Plot sales by day of the week
In [80]:
           _3 |df['Week'] = df['InvoiceDate'].dt.strftime('%U')
              sales_by_week = df.groupby('Week')['Quantity'].sum()
           4
             plt.figure(figsize=(12, 6))
           7 plt.plot(sales_by_week.index, sales_by_week.values, marker='o')
800001t.title('Sales_Trends_by_Week')
           9 plt.xlabel('Week')
          10 plt.ylabel(fabrotal arSales)
          11 plt.xticks(rotation=45)
          12 plt.show()
```

```
In [80]:
             #Plot sales by day of the week
          = #P
2500000
              df['Week'] = df['InvoiceDate'].dt.strftime('%U')
           3
              sales_by_week = df.groupby('Week')['Quantity'].sum()
              plt.figure(figsize=(12, 6))
             plt.plot(sales_by_week.index, sales_by_week.values, marker='o')
           8000 Tt.title('Sales Trends by Week')
           9 plt.xlabel('Week')
                                                   Jul
Month
          10 plt.ylabel(febTotalMarSaleSpr)
                                                                         Oct
                                                                               Nov
                                                                                     Dec
          plt.xticks(rotation=45)
          12 plt.show()
```



Top selling product and countries

```
In [86]:
           1 # Plot top-selling products
           plt.figure(figsize=(10, 6))
           3 top_selling_products.plot(kind='bar', color='skyblue')
           4 plt.title('Top-Selling Products by Quantity')
             plt.xlabel('Product')
plt.ylabel('Total Quantity Sold')
           5
           6
           7
              plt.xticks(rotation=45)
           8
             plt.show()
           9
          10 # Plot top-selling countries
          plt.figure(figsize=(10, 6))
          12 top_selling_countries.plot(kind='bar', color='lightcoral')
          13 plt.title('Top-Selling Countries by Quantity')
```

```
In [86]:
            1 # Plot top-selling products
               plt.figure(figsize=(10, 6))
               top_selling_products.plot(kind='bar', color='skyblue')
               plt.title('Top-Selling Products by Quantity')
plt.xlabel('Product')
               plt.ylabel('Total Quantity Sold')
            7
               plt.xticks(rotation=45)
            8
               plt.show()
            9
           10
              # Plot top-selling countries
           11
               plt.figure(figsize=(10, 6))
               top_selling_countries.plot(kind='bar', color='lightcoral')
              plt.title('Top-Selling Countries by Quantity')
plt.xlabel('Country')
           13
               plt.ylabel('Total Quantity Sold')
           16
               plt.xticks(rotation=45)
           17
               plt.show()
           18
```

