In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

	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	Un Kingc
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Un Kingc
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

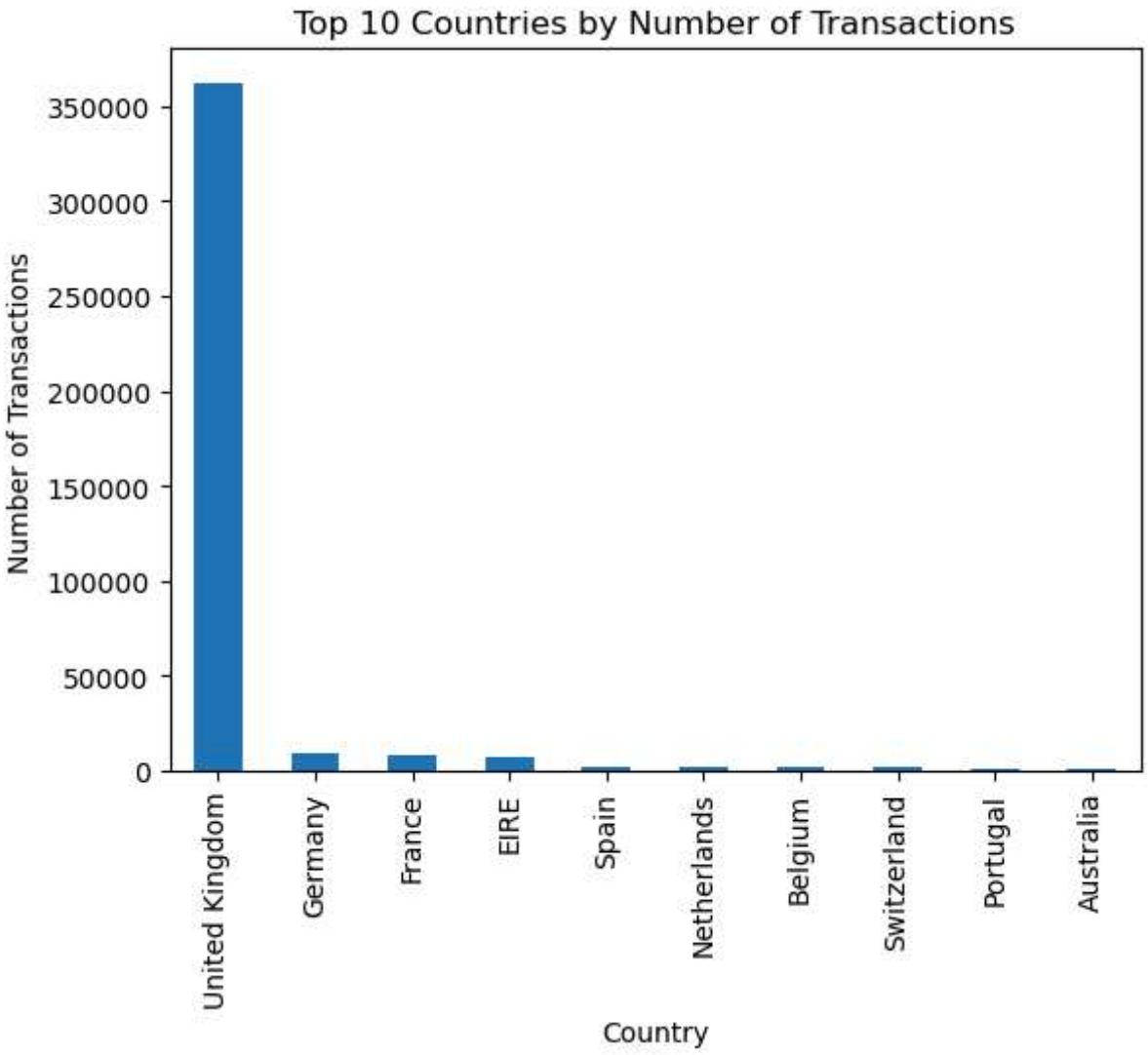
In [5]:
```
df = df.dropna(subset=['CustomerID'])
```

In [6]:
```
df.info()
```

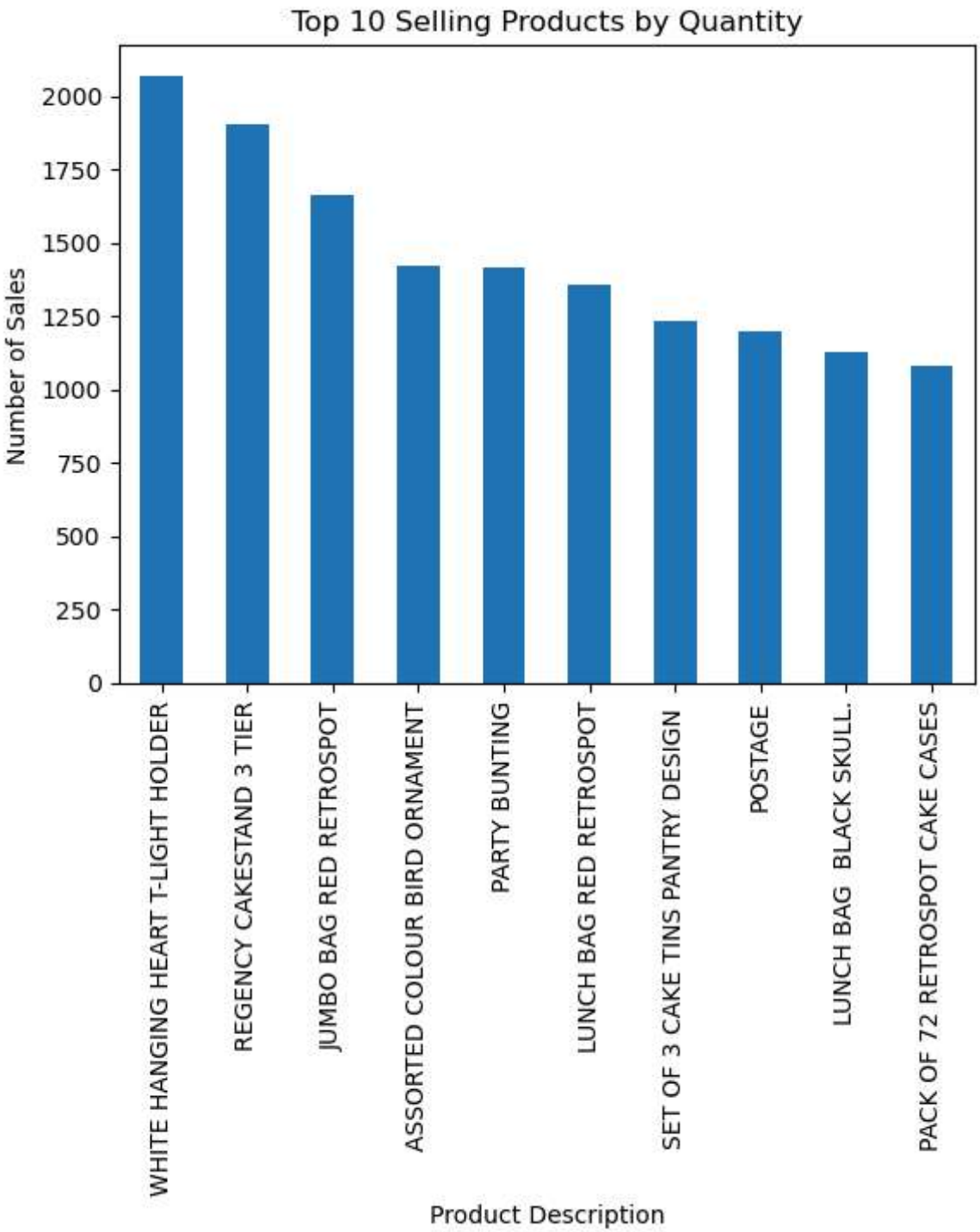
```
<class 'pandas.core.frame.DataFrame'>
Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    406829 non-null  object
 1   StockCode    406829 non-null  object
 2   Description  406829 non-null  object
 3   Quantity     406829 non-null  int64
 4   InvoiceDate  406829 non-null  datetime64[ns]
 5   UnitPrice    406829 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      406829 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

In [7]:
```
df['TotalPrice'] = df['Quantity'] * df['UnitPrice']
```

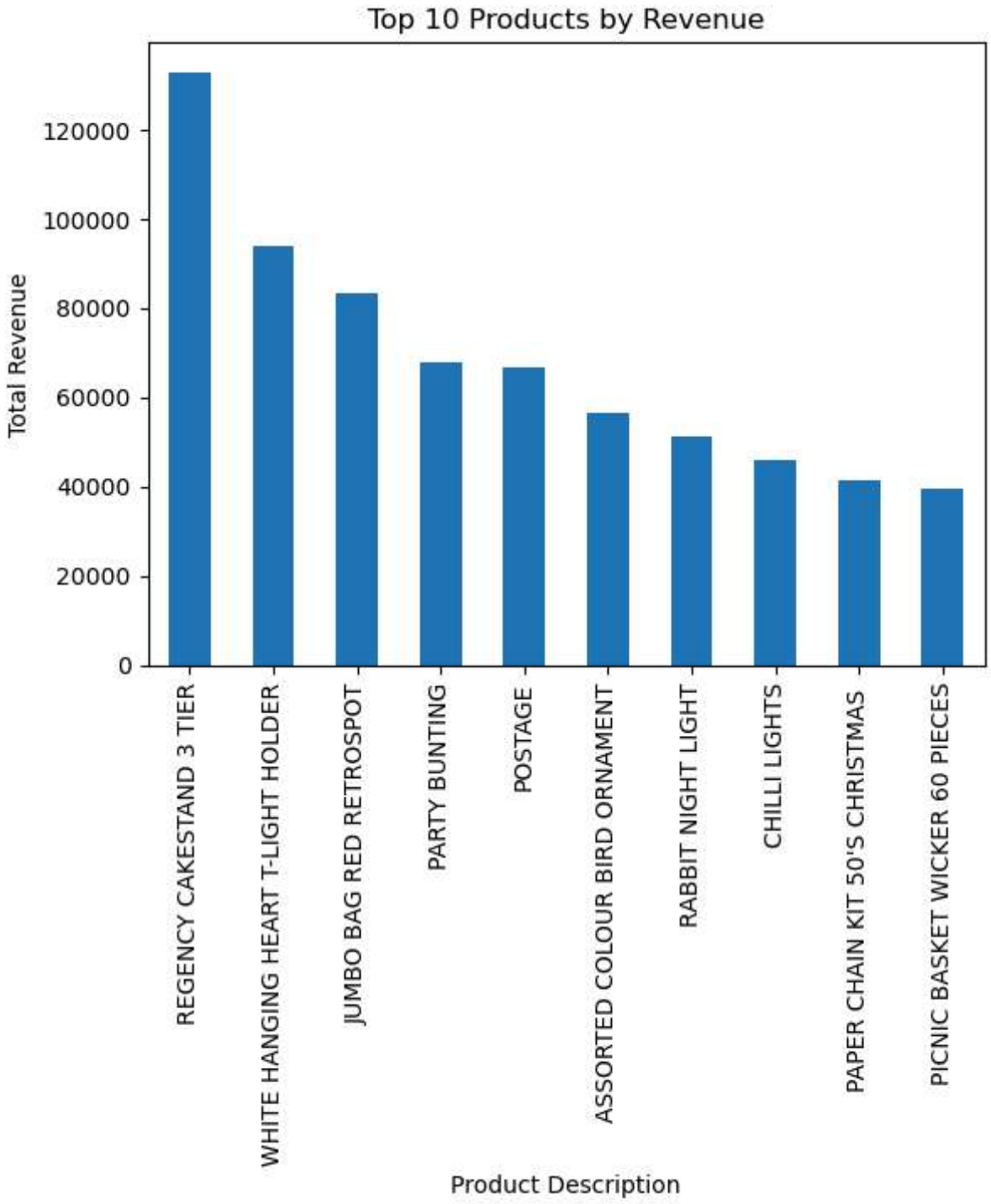
In [8]:
```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Countries by Number of Transactions')
plt.xlabel('Country')
plt.ylabel('Number of Transactions')
plt.show()
```



In [9]:
```
df['Description'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Selling Products by Quantity')
plt.xlabel('Product Description')
plt.ylabel('Number of Sales')
plt.show()
```



In [10]:
```
df.groupby('Description')['TotalPrice'].sum().sort_values(ascending=False).head(10)
plt.title('Top 10 Products by Revenue')
plt.xlabel('Product Description')
plt.ylabel('Total Revenue')
plt.show()
```


Top 10 Products by Revenue
Total Revenue
0
20000
40000
60000
80000
100000
120000
REGENCY CAKESTAND 3 TIER
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
POSTAGE
ASSORTED COLOUR BIRD ORNAMENT
RABBIT NIGHT LIGHT
CHILLI LIGHTS
PAPER CHAIN KIT 50'S CHRISTMAS
PICNIC BASKET WICKER 60 PIECES
Product Description

In [ ]:

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

	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	Un Kingc
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Un Kingc
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

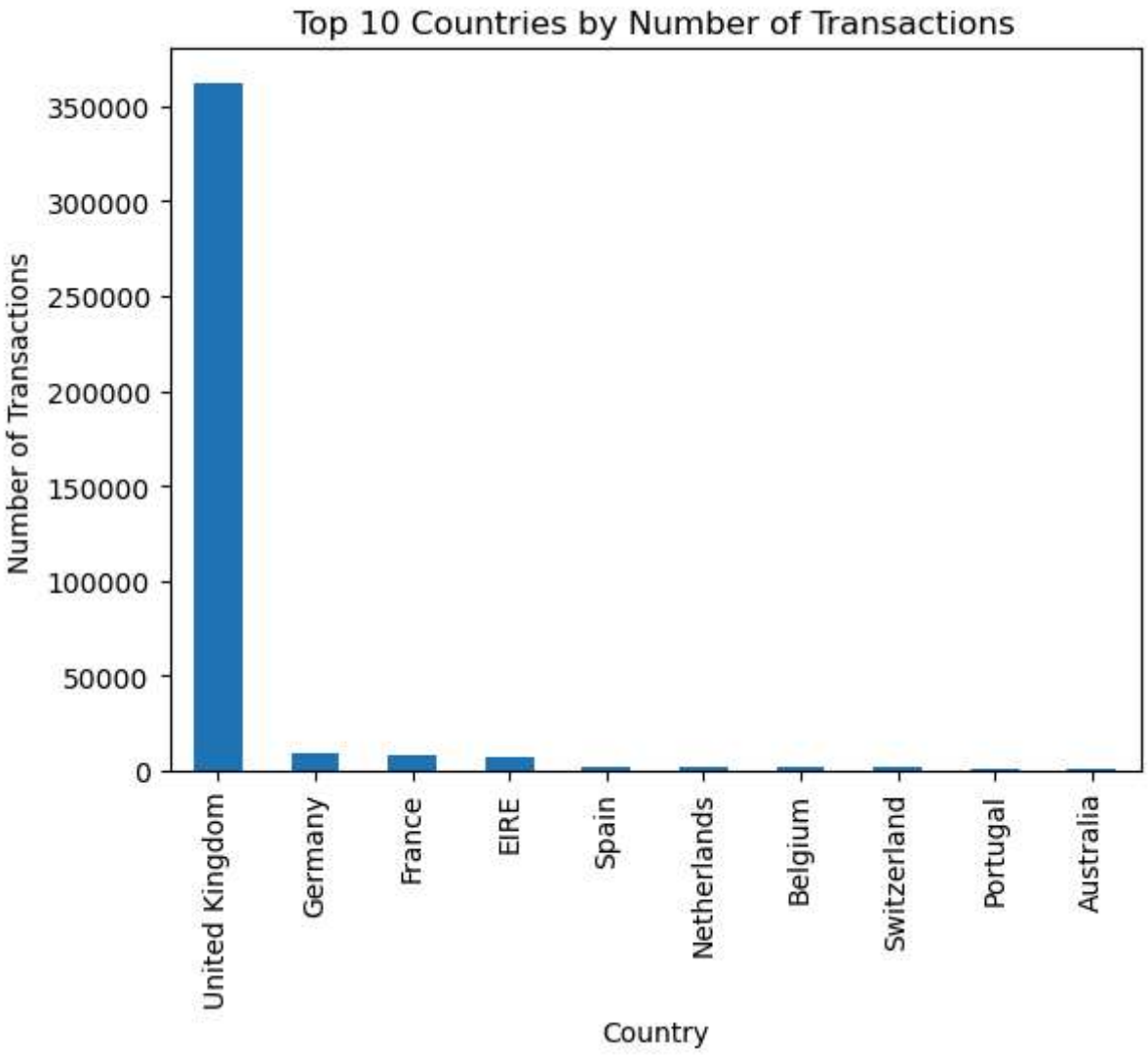
In [5]:
```
df = df.dropna(subset=['CustomerID'])
```

In [6]:
```
df.info()
```

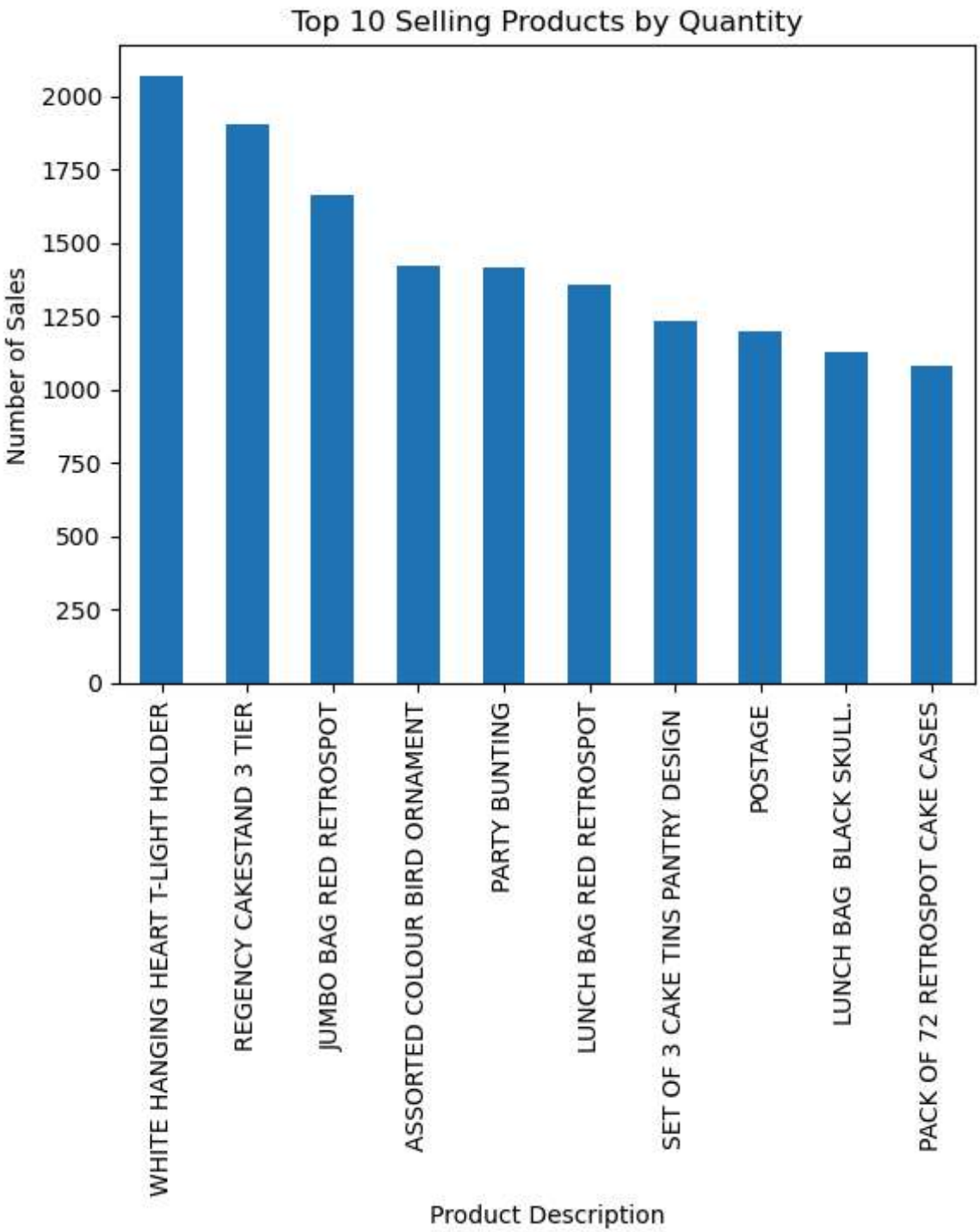
```
<class 'pandas.core.frame.DataFrame'>
Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    406829 non-null  object
 1   StockCode    406829 non-null  object
 2   Description  406829 non-null  object
 3   Quantity     406829 non-null  int64
 4   InvoiceDate  406829 non-null  datetime64[ns]
 5   UnitPrice    406829 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      406829 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

In [7]:
```
df['TotalPrice'] = df['Quantity'] * df['UnitPrice']
```

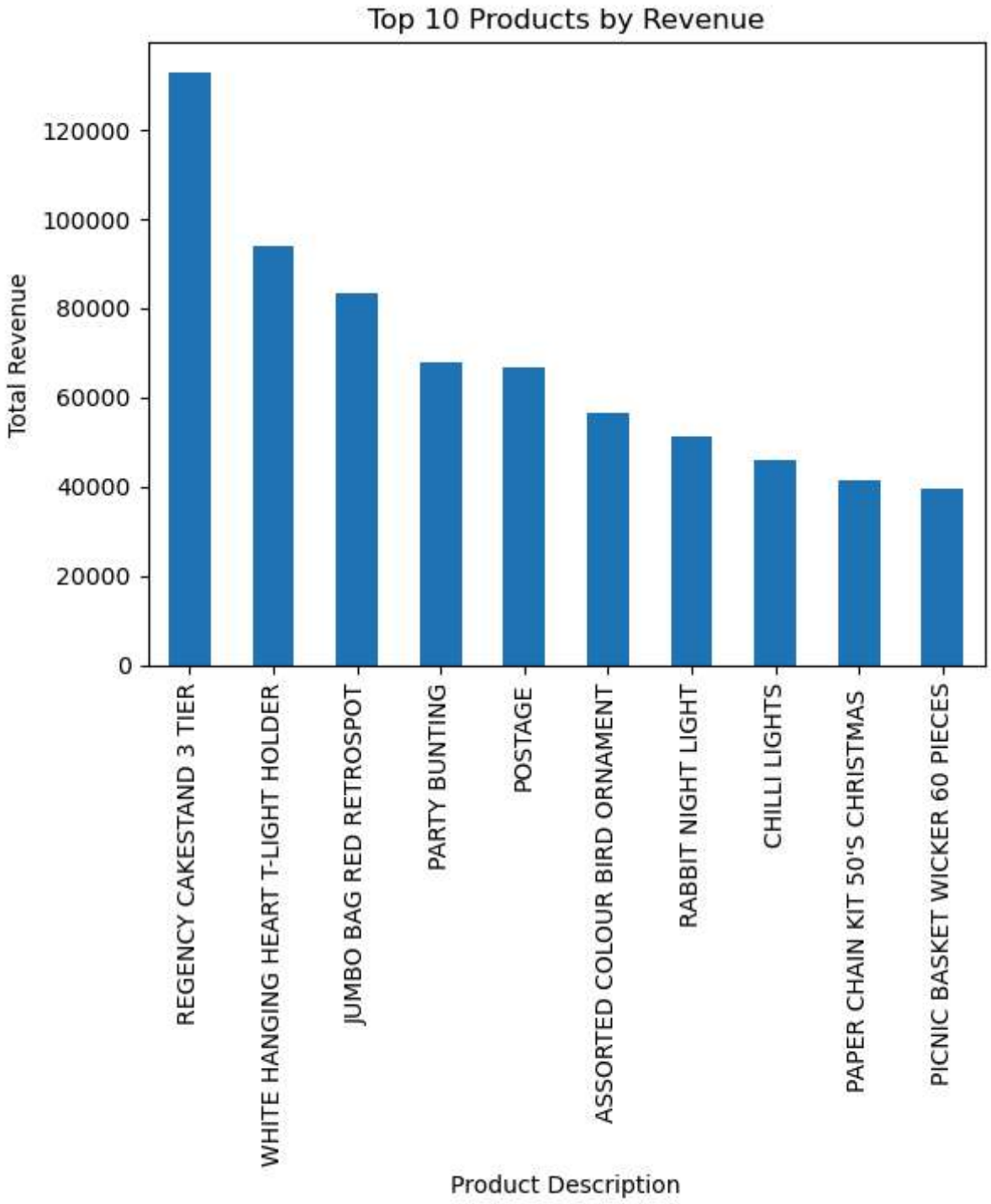
In [8]:
```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Countries by Number of Transactions')
plt.xlabel('Country')
plt.ylabel('Number of Transactions')
plt.show()
```



In [9]:
```
df['Description'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Selling Products by Quantity')
plt.xlabel('Product Description')
plt.ylabel('Number of Sales')
plt.show()
```



In [10]:
```
df.groupby('Description')['TotalPrice'].sum().sort_values(ascending=False).head(10)
plt.title('Top 10 Products by Revenue')
plt.xlabel('Product Description')
plt.ylabel('Total Revenue')
plt.show()
```


Top 10 Products by Revenue
Total Revenue
0
20000
40000
60000
80000
100000
120000
REGENCY CAKESTAND 3 TIER
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
POSTAGE
ASSORTED COLOUR BIRD ORNAMENT
RABBIT NIGHT LIGHT
CHILLI LIGHTS
PAPER CHAIN KIT 50'S CHRISTMAS
PICNIC BASKET WICKER 60 PIECES
Product Description

In [ ]:

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

	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	Un Kingc
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Un Kingc
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

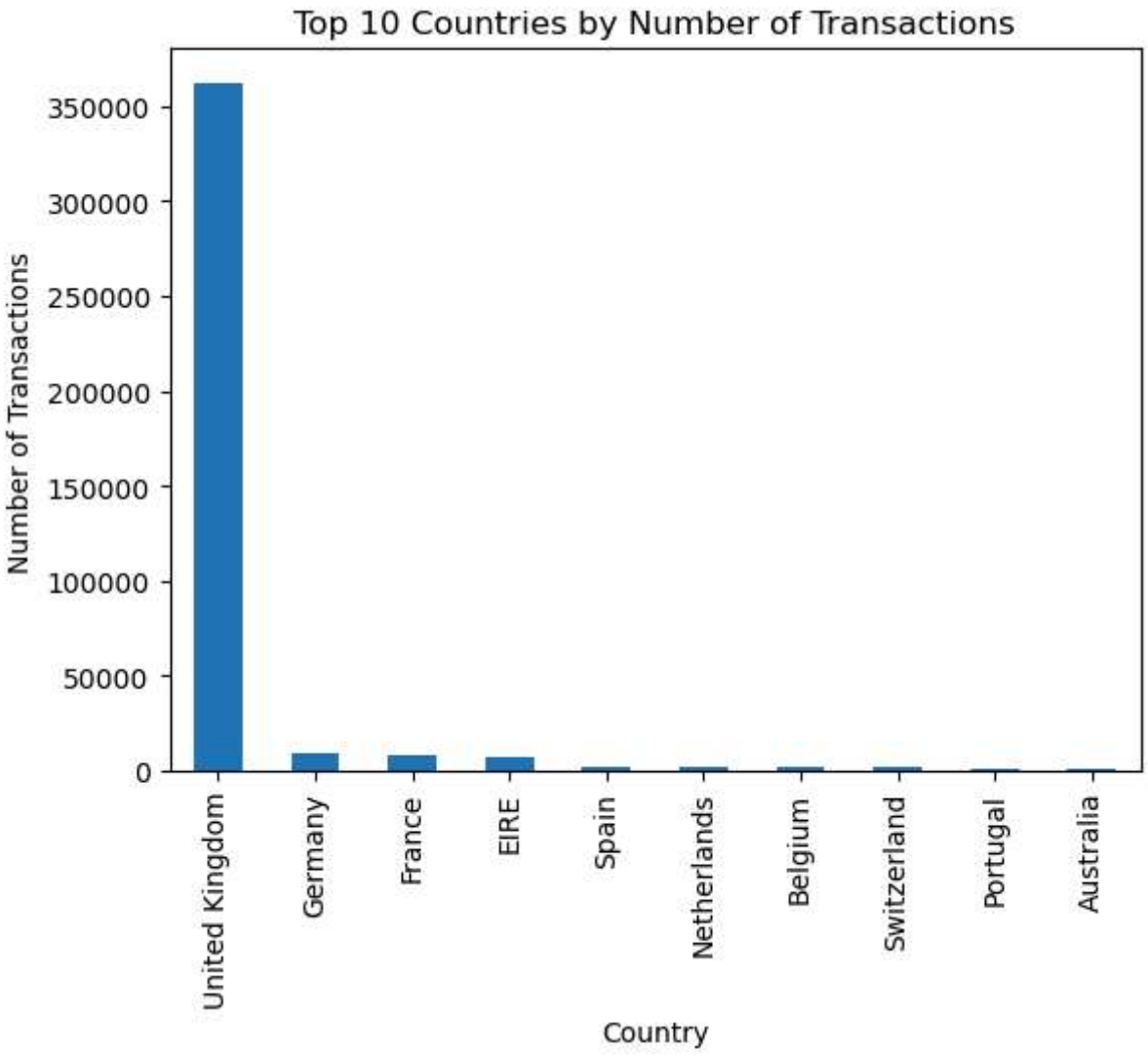
In [5]:
```
df = df.dropna(subset=['CustomerID'])
```

In [6]:
```
df.info()
```

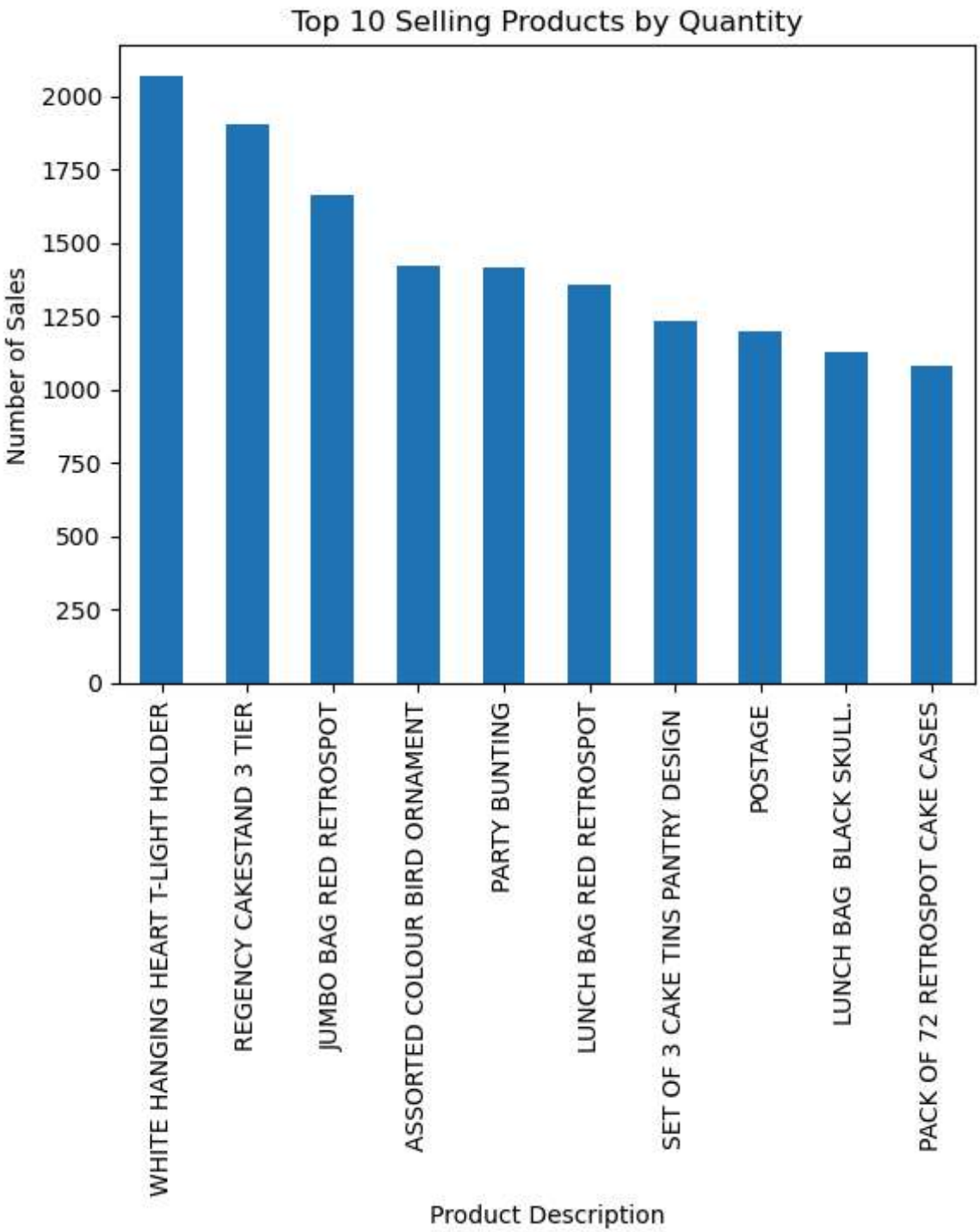
```
<class 'pandas.core.frame.DataFrame'>
Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    406829 non-null  object
 1   StockCode    406829 non-null  object
 2   Description  406829 non-null  object
 3   Quantity     406829 non-null  int64
 4   InvoiceDate  406829 non-null  datetime64[ns]
 5   UnitPrice    406829 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      406829 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

In [7]:
```
df['TotalPrice'] = df['Quantity'] * df['UnitPrice']
```

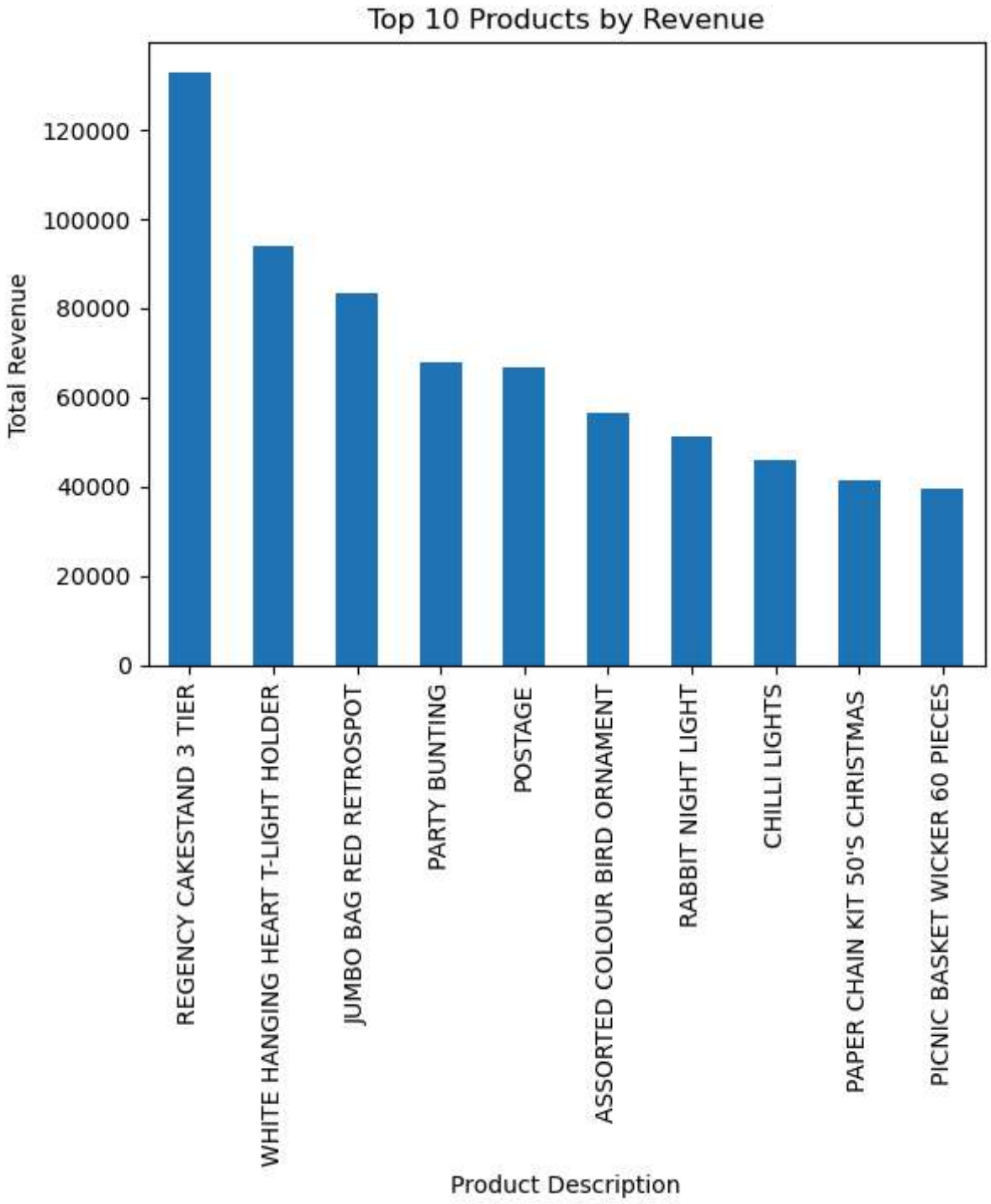
In [8]:
```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Countries by Number of Transactions')
plt.xlabel('Country')
plt.ylabel('Number of Transactions')
plt.show()
```



In [9]:
```
df['Description'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Selling Products by Quantity')
plt.xlabel('Product Description')
plt.ylabel('Number of Sales')
plt.show()
```



In [10]:
```
df.groupby('Description')['TotalPrice'].sum().sort_values(ascending=False).head(10)
plt.title('Top 10 Products by Revenue')
plt.xlabel('Product Description')
plt.ylabel('Total Revenue')
plt.show()
```


Top 10 Products by Revenue
Total Revenue
0
20000
40000
60000
80000
100000
120000
REGENCY CAKESTAND 3 TIER
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
POSTAGE
ASSORTED COLOUR BIRD ORNAMENT
RABBIT NIGHT LIGHT
CHILLI LIGHTS
PAPER CHAIN KIT 50'S CHRISTMAS
PICNIC BASKET WICKER 60 PIECES
Product Description

In [ ]:

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

	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	Un Kingc
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Un Kingc
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

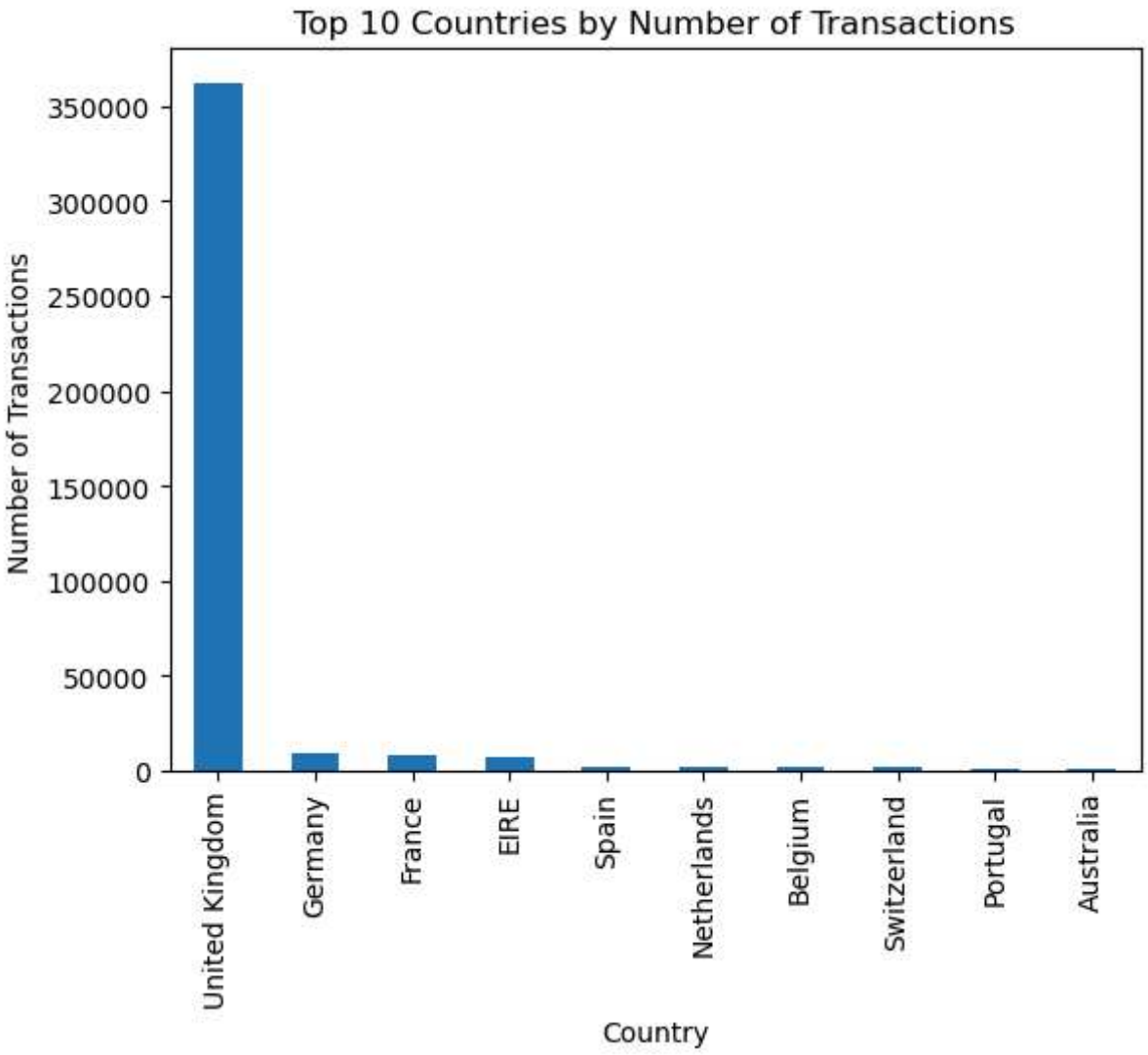
In [5]:
```
df = df.dropna(subset=['CustomerID'])
```

In [6]:
```
df.info()
```

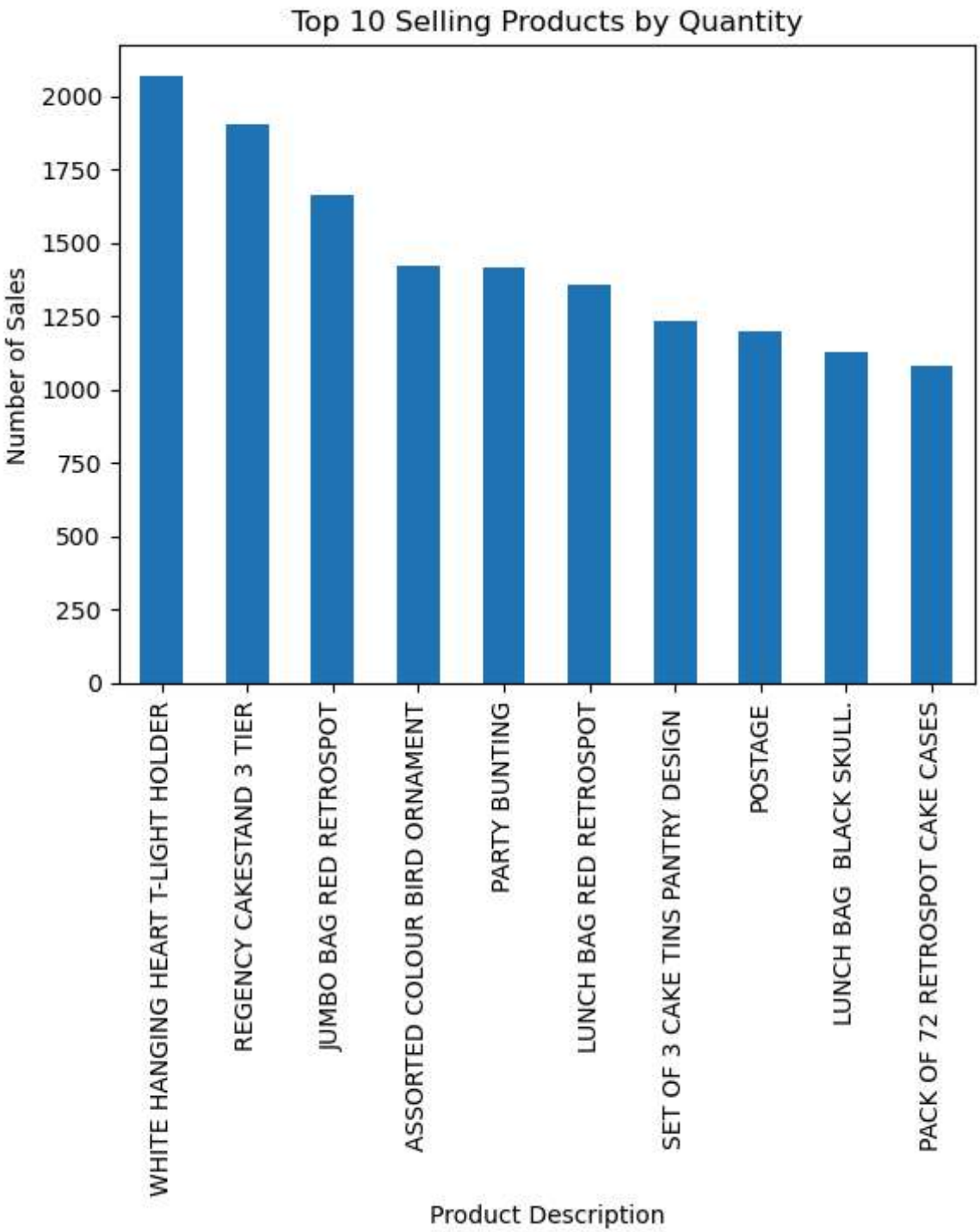
```
<class 'pandas.core.frame.DataFrame'>
Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    406829 non-null  object
 1   StockCode    406829 non-null  object
 2   Description  406829 non-null  object
 3   Quantity     406829 non-null  int64
 4   InvoiceDate  406829 non-null  datetime64[ns]
 5   UnitPrice    406829 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      406829 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

In [7]:
```
df['TotalPrice'] = df['Quantity'] * df['UnitPrice']
```

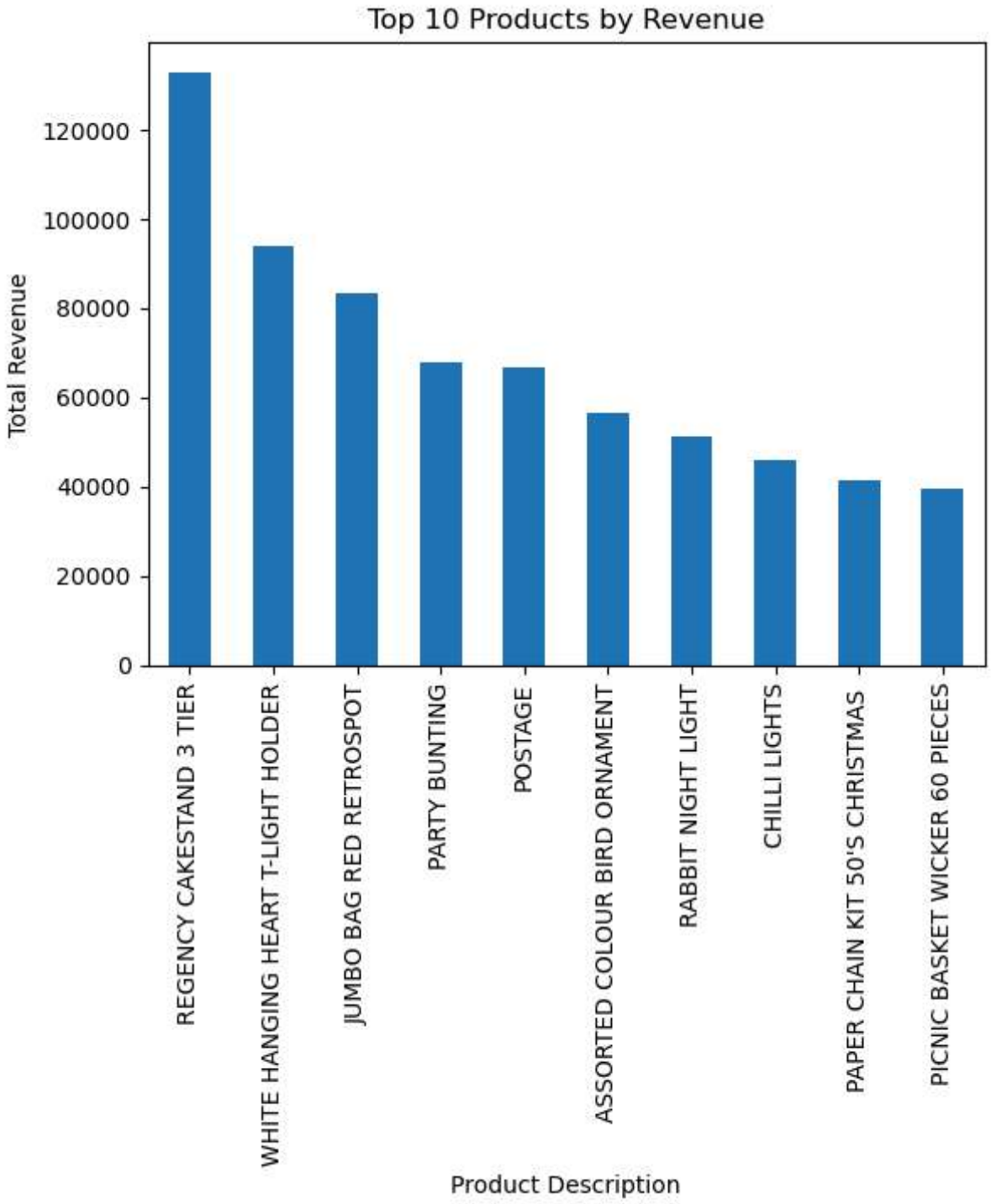
In [8]:
```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Countries by Number of Transactions')
plt.xlabel('Country')
plt.ylabel('Number of Transactions')
plt.show()
```



In [9]:
```
df['Description'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Selling Products by Quantity')
plt.xlabel('Product Description')
plt.ylabel('Number of Sales')
plt.show()
```



In [10]:
```
df.groupby('Description')['TotalPrice'].sum().sort_values(ascending=False).head(10)
plt.title('Top 10 Products by Revenue')
plt.xlabel('Product Description')
plt.ylabel('Total Revenue')
plt.show()
```


Top 10 Products by Revenue
Total Revenue
0
20000
40000
60000
80000
100000
120000
REGENCY CAKESTAND 3 TIER
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
POSTAGE
ASSORTED COLOUR BIRD ORNAMENT
RABBIT NIGHT LIGHT
CHILLI LIGHTS
PAPER CHAIN KIT 50'S CHRISTMAS
PICNIC BASKET WICKER 60 PIECES
Product Description

In [ ]:

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

	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	Un Kingc
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Un Kingc
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Un Kingc

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

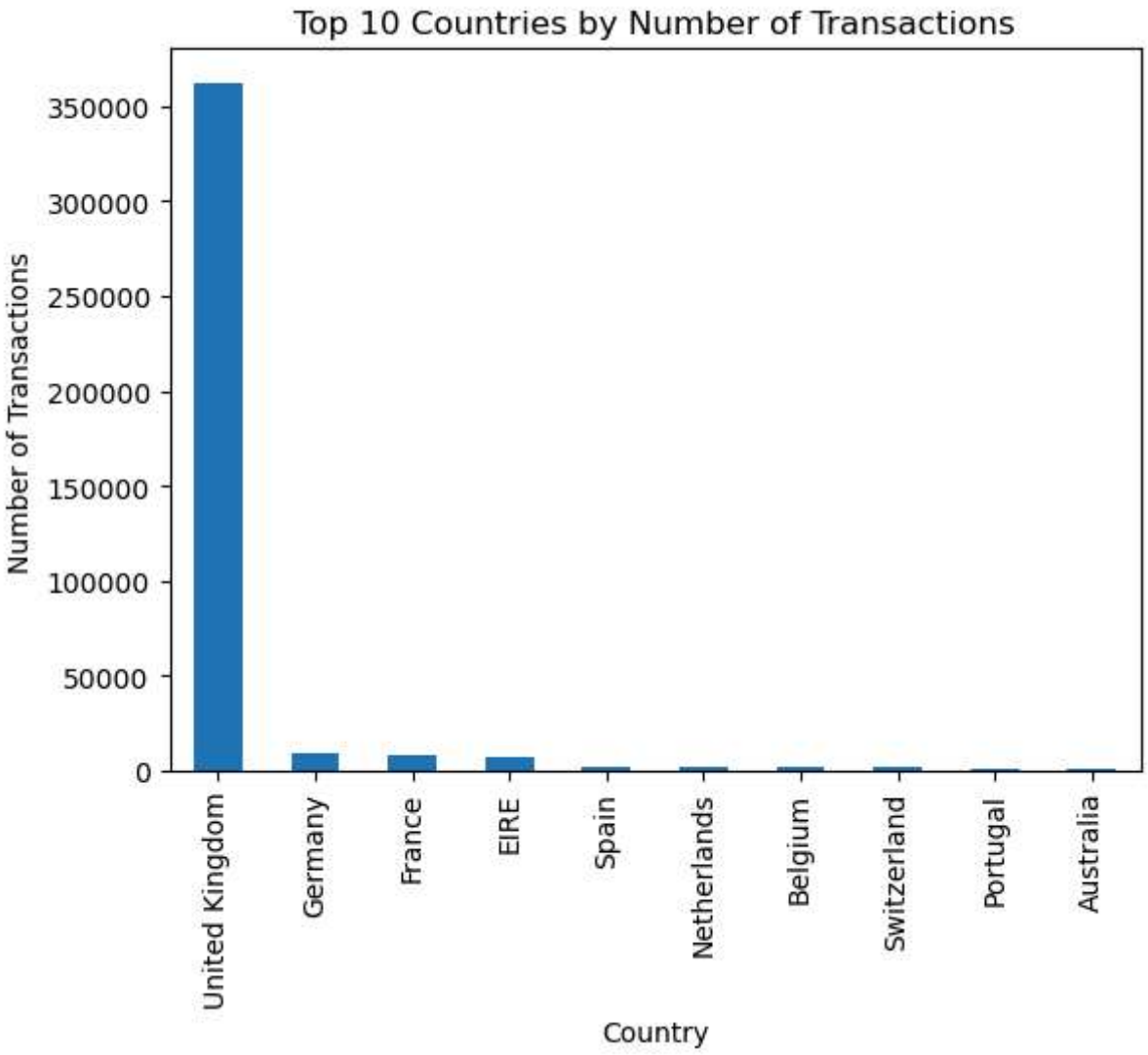
In [5]:
```
df = df.dropna(subset=['CustomerID'])
```

In [6]:
```
df.info()
```

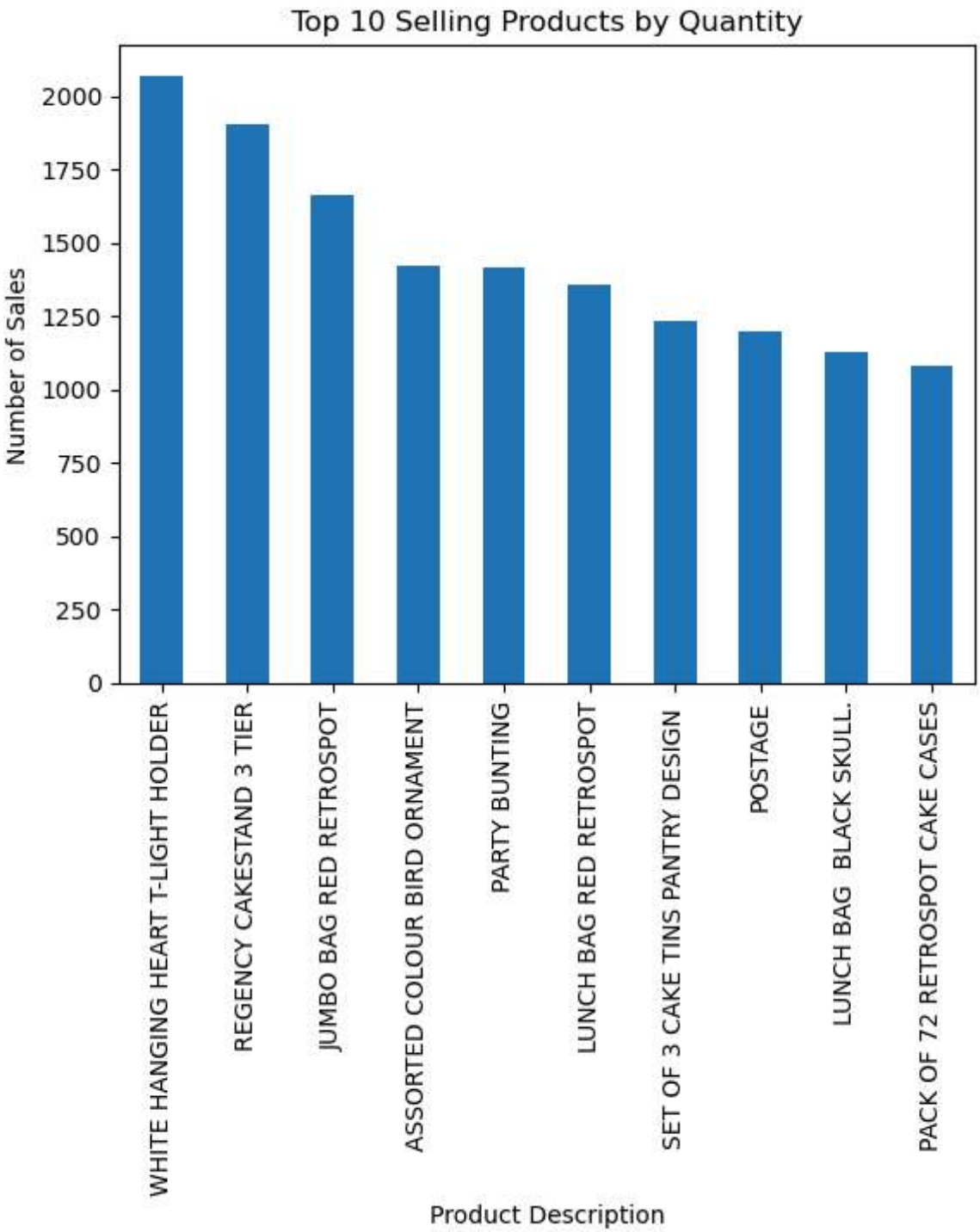
```
<class 'pandas.core.frame.DataFrame'>
Index: 406829 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    406829 non-null  object
 1   StockCode    406829 non-null  object
 2   Description  406829 non-null  object
 3   Quantity     406829 non-null  int64
 4   InvoiceDate  406829 non-null  datetime64[ns]
 5   UnitPrice    406829 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      406829 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.9+ MB
```

In [7]:
```
df['TotalPrice'] = df['Quantity'] * df['UnitPrice']
```

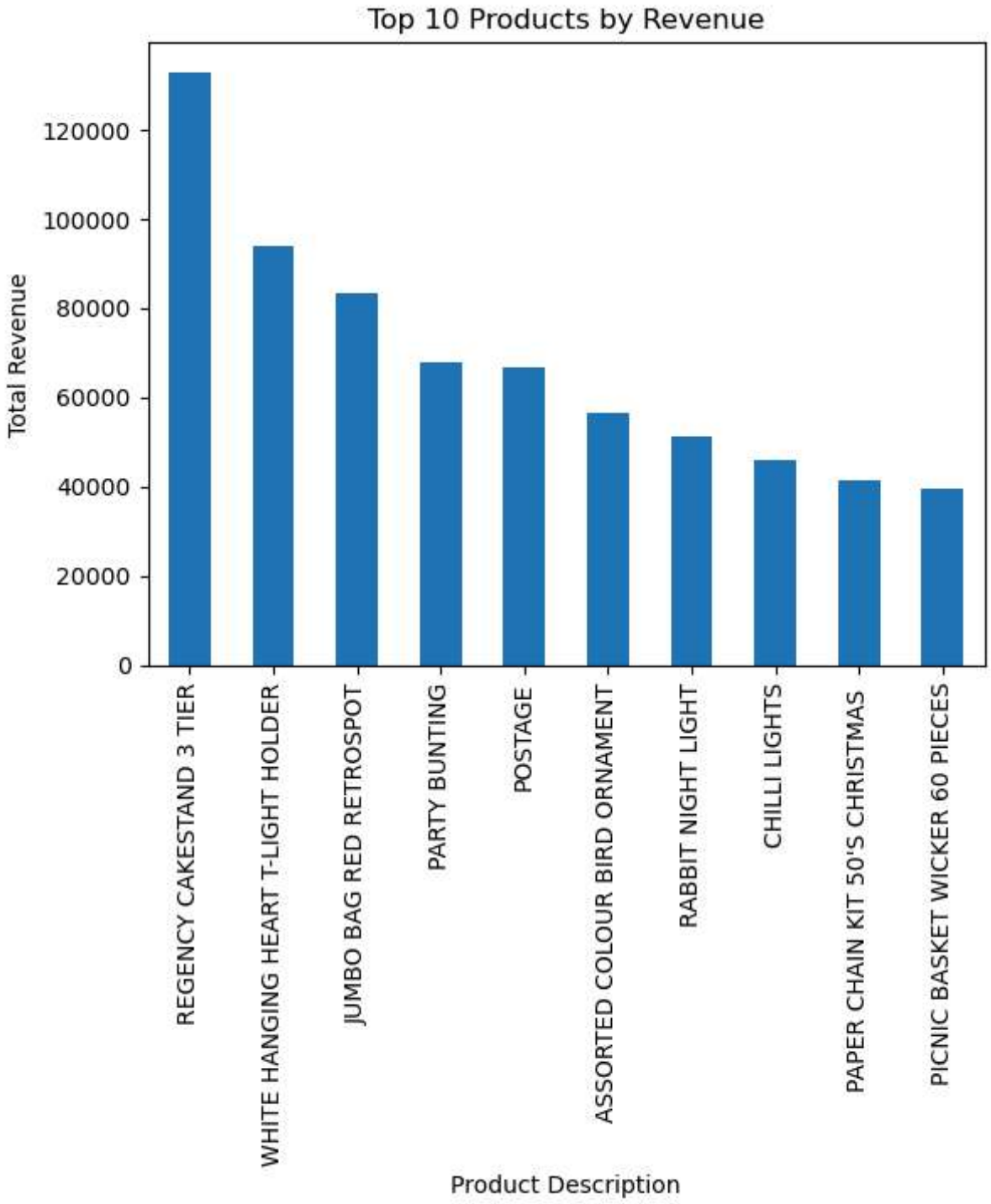
In [8]:
```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Countries by Number of Transactions')
plt.xlabel('Country')
plt.ylabel('Number of Transactions')
plt.show()
```



In [9]:
```
df['Description'].value_counts().head(10).plot(kind='bar')
plt.title('Top 10 Selling Products by Quantity')
plt.xlabel('Product Description')
plt.ylabel('Number of Sales')
plt.show()
```



In [10]:
```
df.groupby('Description')['TotalPrice'].sum().sort_values(ascending=False).head(10)
plt.title('Top 10 Products by Revenue')
plt.xlabel('Product Description')
plt.ylabel('Total Revenue')
plt.show()
```


Top 10 Products by Revenue
Total Revenue
0
20000
40000
60000
80000
100000
120000
REGENCY CAKESTAND 3 TIER
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
POSTAGE
ASSORTED COLOUR BIRD ORNAMENT
RABBIT NIGHT LIGHT
CHILLI LIGHTS
PAPER CHAIN KIT 50'S CHRISTMAS
PICNIC BASKET WICKER 60 PIECES
Product Description

In [ ]: