Examing Retail in Europe through BigMart

SEABORN GROUP

Data Analysts: Raiyan R., Guy L., Lily O., and Lexi B.

Getting Started

Selecting the data:
“BigMartSalesData.csv”
.CSV file

Pandas Dataframe

Importing helpful libraries
Import Pandas: used for data manipulation and analysis in Python
Import Numpy: used for numerical computing
Import Matplotlib.pyplot: used for creating plots and visualizations
Import Seaborn: used for statistical data visualization

Helpful Information

541,874 entries

Specifc data (from 12 columns):

Invoice Number
Stock Code
Description
Quantity
Unit Price
Amount

Invoice Date
Day
Month
Year
Customer ID
Country

& Data Types

Float64 (3)
Int64 (4)
Object (5)

Statistical Information

Aggregation functions used:

- Count
- Mean
- Standard deviation (std)
- Minimum (min)
- Q1 (25th percentile of each column)
- Q2/median (50th percentile)
- Q3 (75th percentile)
- Maximum (max)

Interpretations from data:

- Average quantity of items purchased: 11.34
- Average unit price: $4.61
- Average amount spent: $21.30
- Min quantity: 1
- Max quantity: 80,995
- Max unit price: $38,970
- Max amount spent: $186,469.60

Items Sold

Used filtering, grouping, aggregating, sorting, and slicing to find the top 10 items sold in the US and in the UK.

SET/10 PINK POLKADOT PARTY CANDLES	96
SET/10 BLUE POLKADOT PARTY CANDLES	96
SET/10 RED POLKADOT PARTY CANDLES	96
SET/10 IVORY POLKADOT PARTY CANDLES	96
SET 12 COLOURING PENCILS DOILY	88
SET OF 5 LUCKY CAT MAGNETS	72
MINI PAINT SET VINTAGE	72
12 PENCILS SMALL TUBE RED RETROSPOT	72
WRAP ENGLISH ROSE	50
BLUE POLKADOT WRAP	50

PAPER CRAFT , LITTLE BIRDIE	161990
MEDIUM CERAMIC TOP STORAGE JAR	151503
WORLD WAR 2 GLIDERS ASSTD DESIGNS	50726
JUMBO BAG RED RETROSPOT	45369
WHITE HANGING HEART T-LIGHT HOLDER	38295
POPCORN HOLDER	35111
ASSORTED COLOUR BIRD ORNAMENT	33791
PACK OF 12 LONDON TISSUES	25355
PACK OF 72 RETROSPOT CAKE CASES	25270
VICTORIAN GLASS HANGING T-LIGHT	24142

Sales by Country

Determining which 10 countries contribute more towards sales using grouping, sorting, and aggregate data, such as the mean to reveal the list:

United Kingdom
Germany
France
EIRE
Spain
Netherlands
Belgium
Switzerland
Portugal
Australia

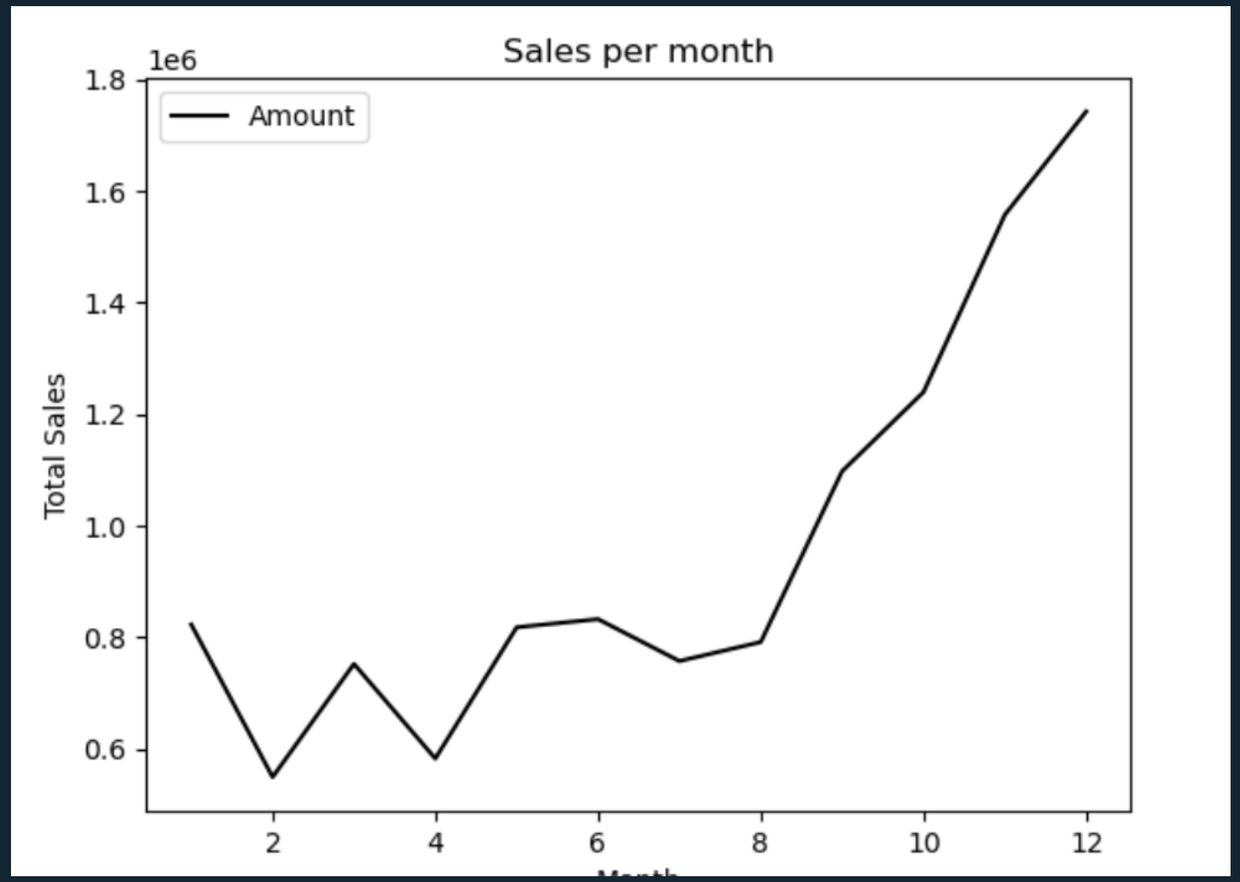
```
df.groupby('Country')['Amount'].count().sort_values(ascending=False)
```

```
Country
United Kingdom    495443
Germany             9495
France              8557
EIRE                8196
Spain               2533
                    ...
Lithuania             35
Brazil                32
Czech Republic        30
Bahrain               19
Saudi Arabia          10
Name: Amount, Length: 38, dtype: int64
```

Sales Per Month

Key factors

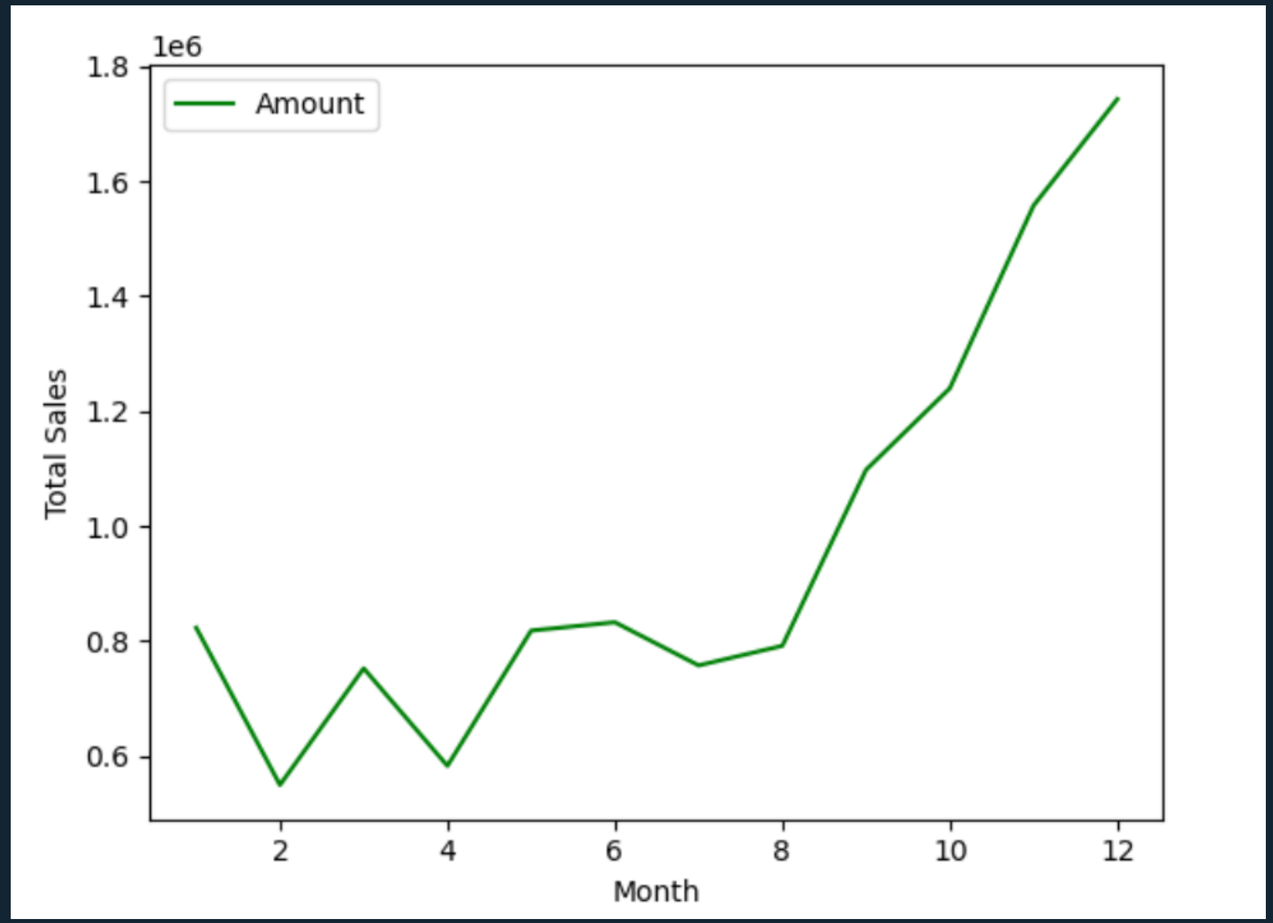
- Line graph shows an increase in sales starting at Month 0 and continuing to Month 12
- Y-axis displaying total sales uses scale, as actual values are larger
- Plot includes: axes labels, legend, color, and graph title



Sales Amount Per Month

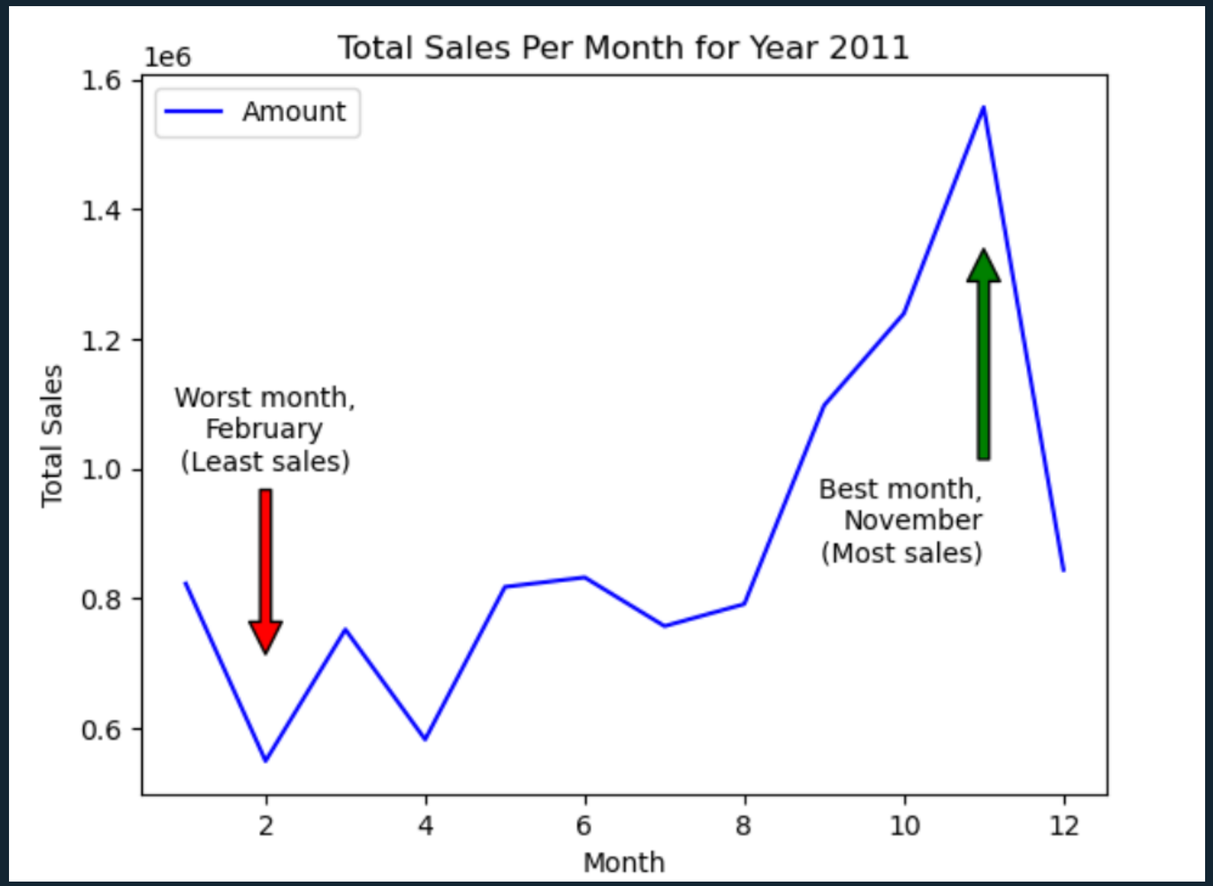
Key factors

- Increasing amount of sales per month
- Change in color of graph
- Plt.show() displays graph

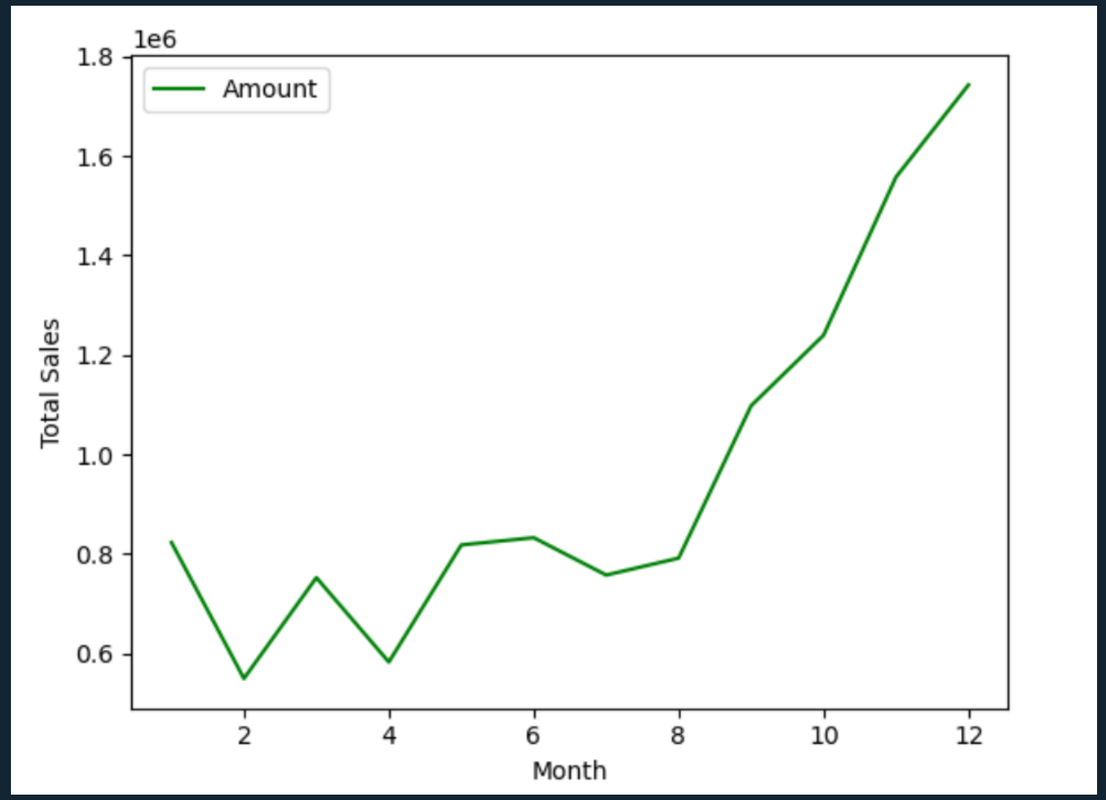


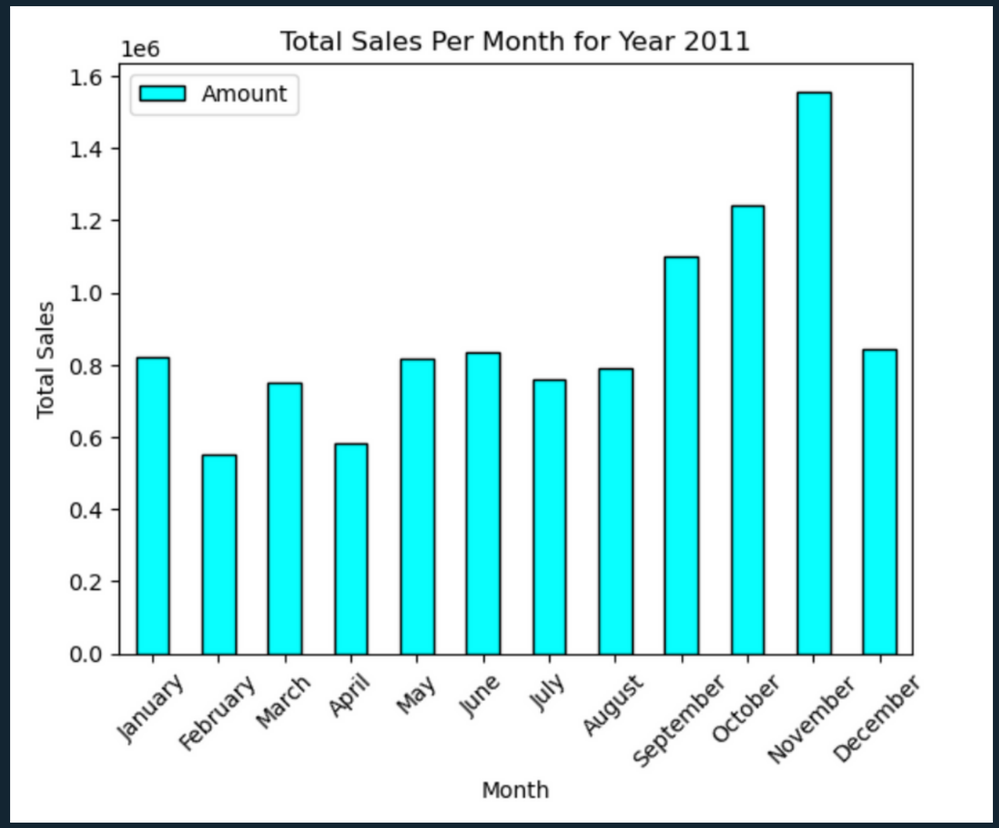
Changes in Sales in 2011

- Calculate and plot the total sales by month
- Identify the best and worst sales months
- Values can be used to analyze trends, seasonality, and overall sales patterns over the course of the year



Choosing the Visualization: Simple vs. Bar Plot





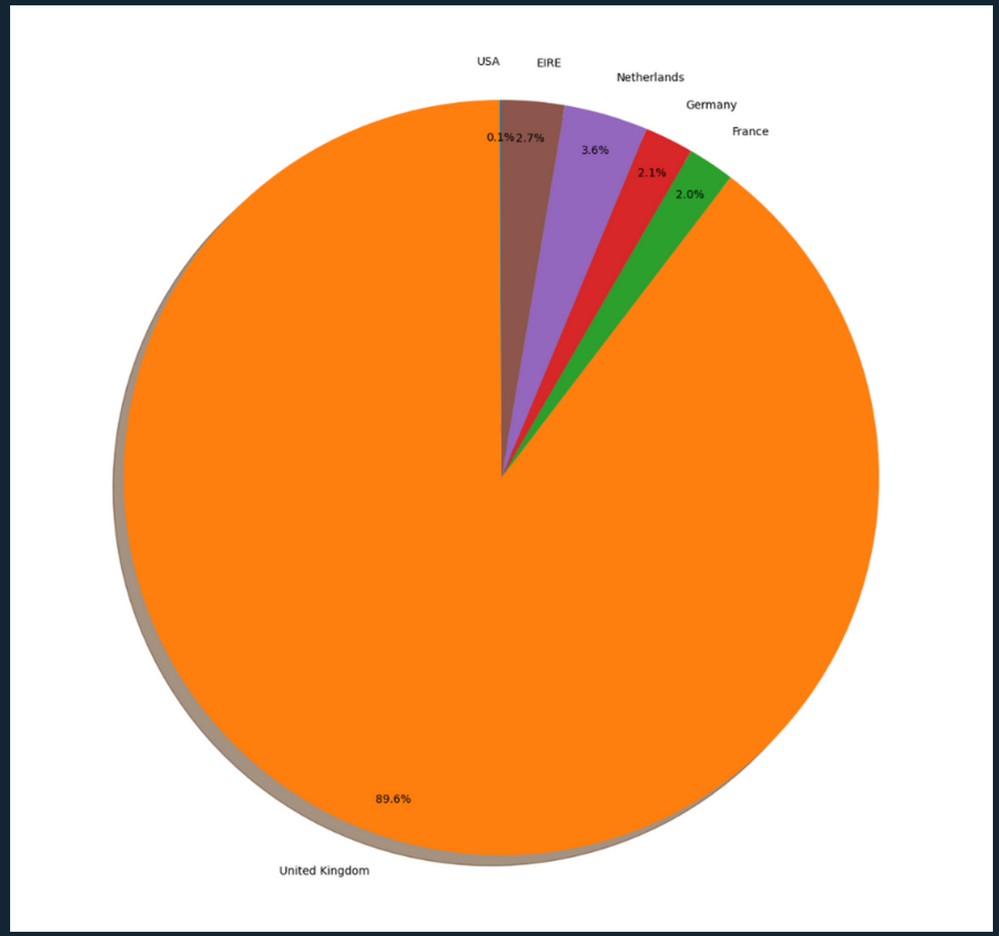
The bar plot could be easier to read as every individual month is labeled, unlike the simple plot which labels every other month.

Competition Among Countries

The UK contributes the most towards sales (89.6%), followed by...

- Netherlands - 3.6%
- EIRE (Ireland) - 2.7%
- Germany - 2.1%
- France 2%
- USA - 0.1%

Foundations used: lists, data aggregation, plotting, pie chart, and layout



Thank you for listening!