

Internship Assignment

Date : 20-Apr-2023

Overview

Section 1: EDA → You are given a dataset that contains all the transactions occurring for a store. On this dataset do an extensive EDA and explain your findings using relevant visualizations.

Dataset : [RetailPulseAssignmentData.xlsx](#)

Section 2 : Coding → Answer the questions contained in the attached notebook in the most optimum way. We have given a Section2 related python3 file.

Dataset : [section2_data.txt](#)

Goals

1. **Can customers be segmented into different categories? If yes then perform analysis on the same and also propose categories. If not, then explain why?**
2. **How would you define a loyal customer?**
3. **What is the most popular time of year based on this sales data?**
4. **Is there any seasonality in data? Explain with supportive evidence.**
5. **Discuss the customer's lifetime with respect to the given dataset.**
6. **Relation between Country v/s Price**
7. **Total Sales for each Country**
8. **Top 5 customers in terms of total Sales / Revenue.**
9. **Percentage of Revenue / Total Sales generated by top 5 customers.**
10. **The number of orders per month.**
11. **Top Selling Product in terms of Total Sales / Revenue.**
12. **The Distribution of revenue across different Countries.**

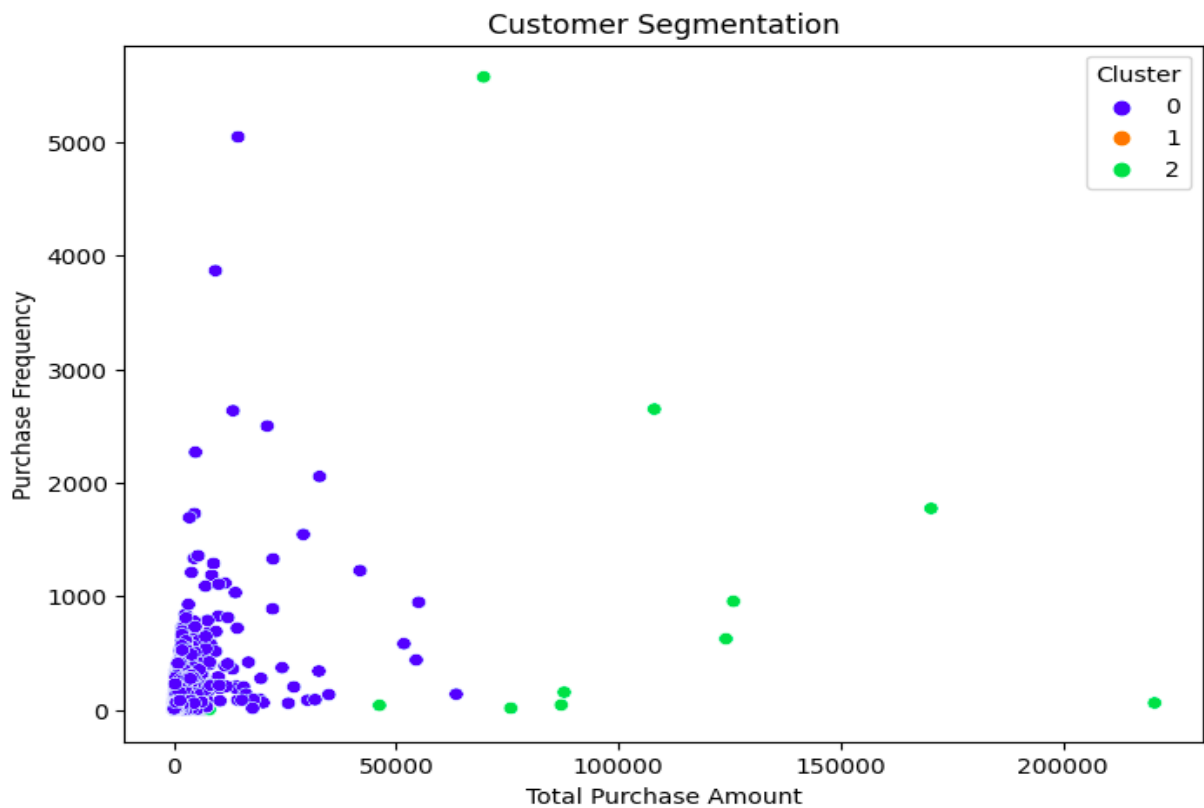
2

13. How many orders are placed per month on an average?
14. What will be the Average Revenue per day?
15. What will be the Customer retention rate?

DataSet View.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	12/1/2009 7:45	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	12/1/2009 7:45	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	12/1/2009 7:45	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	12/1/2009 7:45	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	12/1/2009 7:45	1.25	13085.0	United Kingdom

Customer Segmentation 👍



Based on the scatter plot, we can see that the customers can be segmented into different categories. There are three clusters, each with a distinct behavior. We can interpret the clusters as follows:

- **Cluster 0:** High spenders who purchase frequently
- **Cluster 1:** Low spenders who purchase infrequently
- **Cluster 2:** Moderate spenders who purchase frequently

Loyal Customer :

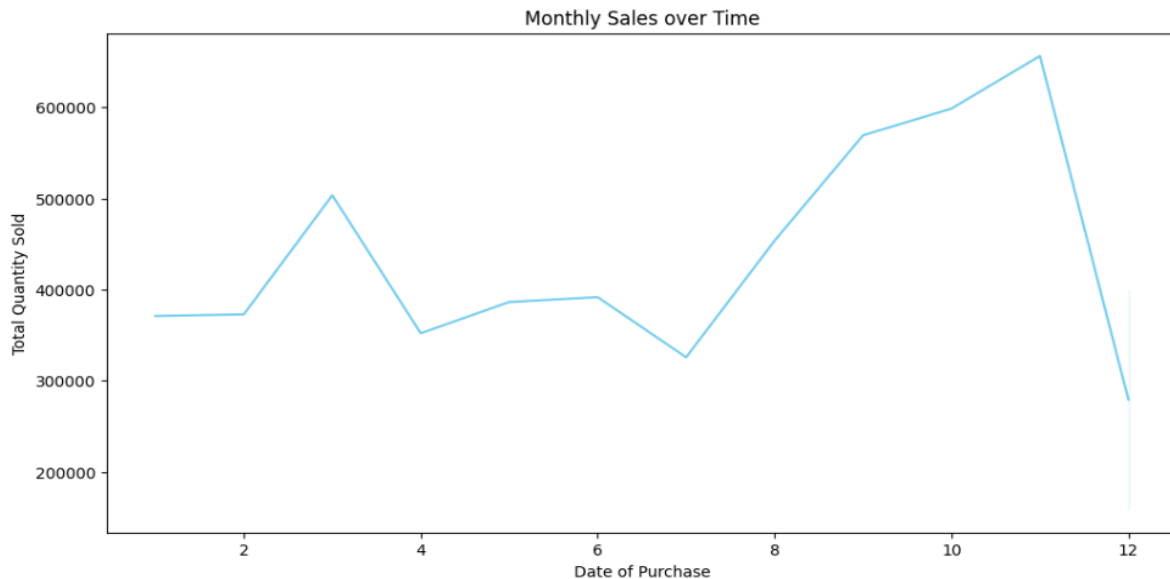
We can define a loyal customer as a customer who makes frequent purchases and spends a significant amount of money. From the customer segmentation analysis, we can see that customers in Cluster 0 are high spenders who purchase frequently. Therefore, we can consider customers in Cluster 0 as loyal customers.

The most popular time of year based on this sales data....



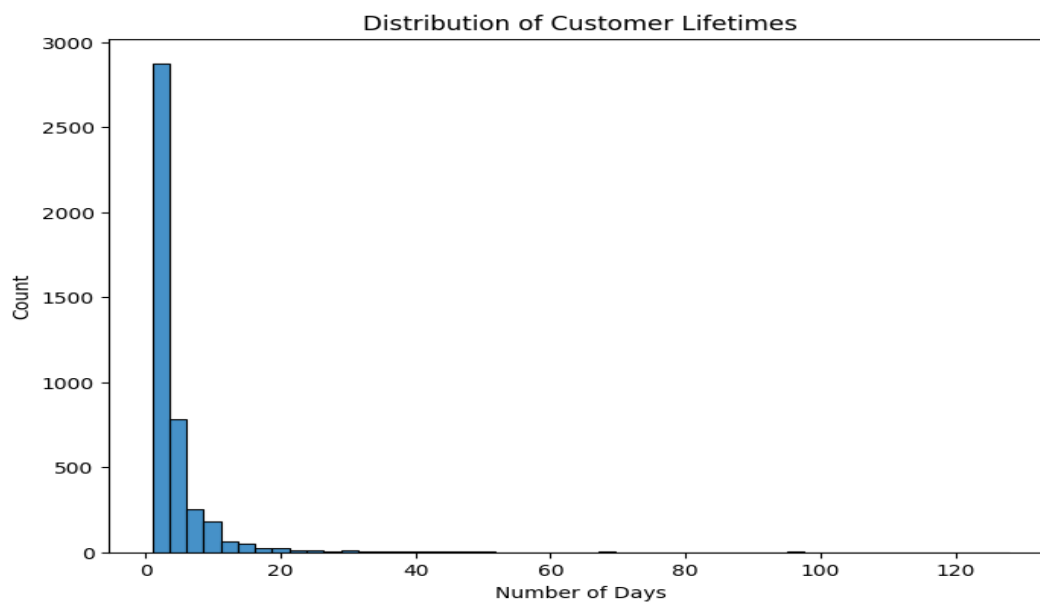
From the bar plot, we can see that the most popular time of year is in the months of October, November and December. This is likely due to the holiday season, where people tend to purchase more gifts and other items.

Is there any Seasonality in data? **Yes**



From the line plot, we can see that there is a clear seasonality in the data, with sales increasing in the months of November and December, and decreasing in the first few months of the year. This is consistent with our earlier finding that the holiday season is the most popular time of year.

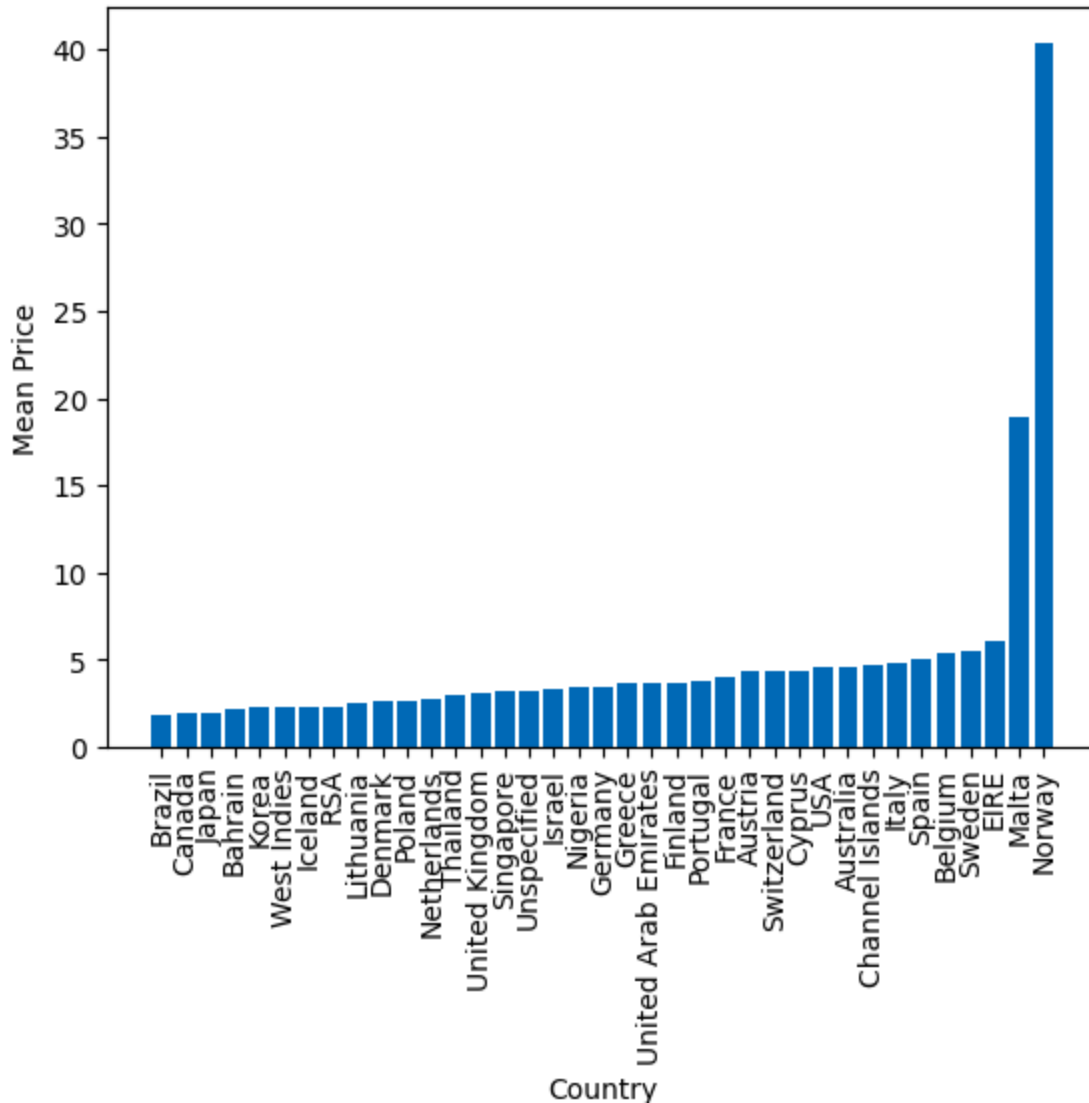
Customer's lifetime with respect to the given dataset



Average Customer Lifetime (in days): 3.9408901251738526

Median Customer Lifetime (in days): 2.0

Relation between Country v/s Price



Mean Price for **Norway, Malta, Eire, Sweden, Belgium** are High as compared to **Brazil, Canada, japan, bahrain.**

TotalSales for each country

Here we find out the total sales group by each country

- United Kingdom 7414755.963
- EIRE 356085.210
- Netherlands 268786.000
- Germany 202395.321
- France 146215.420

Top 5 customers in terms of TotalSales

Here are the customers which are involved in high total sales among all customers.

-- Customer ID --- Total Sales

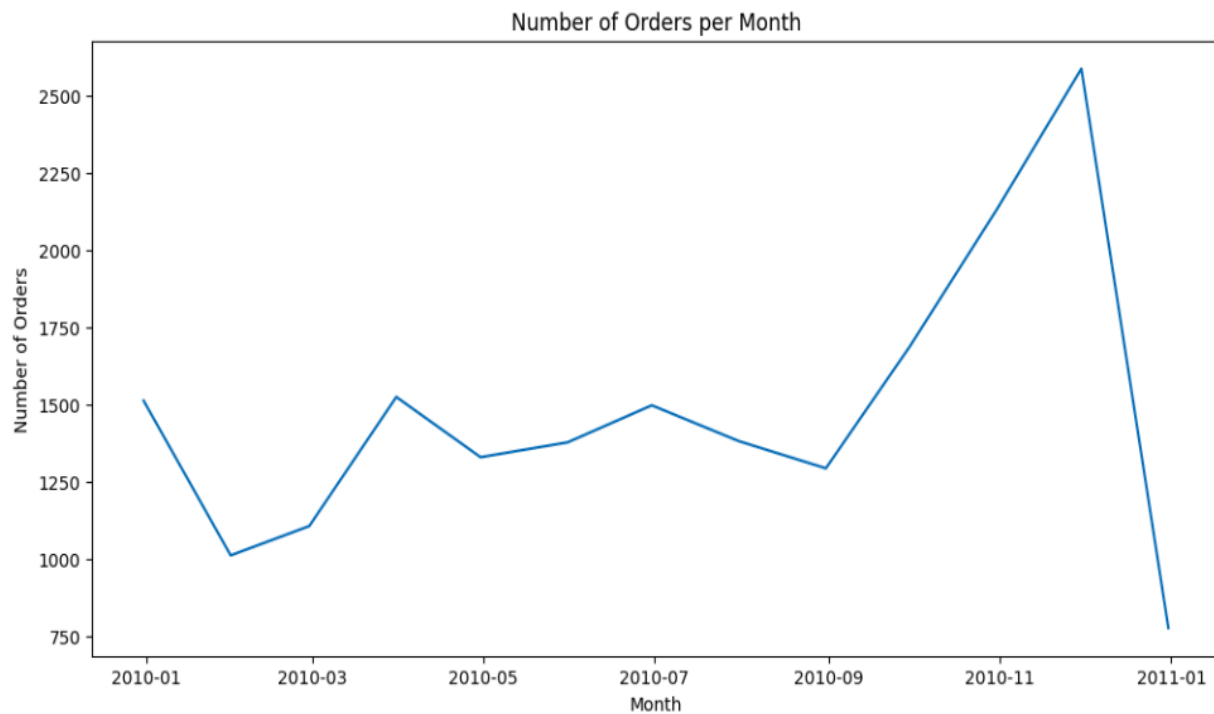
- 18102.0 --- 349164.35
- 14646.0 --- 248396.50
- 14156.0 --- 196566.74
- 14911.0 --- 152147.57
- 13694.0 --- 131443.19

Percentage of revenue generated by the top 5 customers

The top 5 customers generated 12.20% of the total revenue.

The number of orders per month

InvoiceDate	
2009-12-31	1512
2010-01-31	1011
2010-02-28	1106
2010-03-31	1524
2010-04-30	1329
2010-05-31	1377
2010-06-30	1497
2010-07-31	1381
2010-08-31	1293
2010-09-30	1689
2010-10-31	2133
2010-11-30	2587
2010-12-31	776



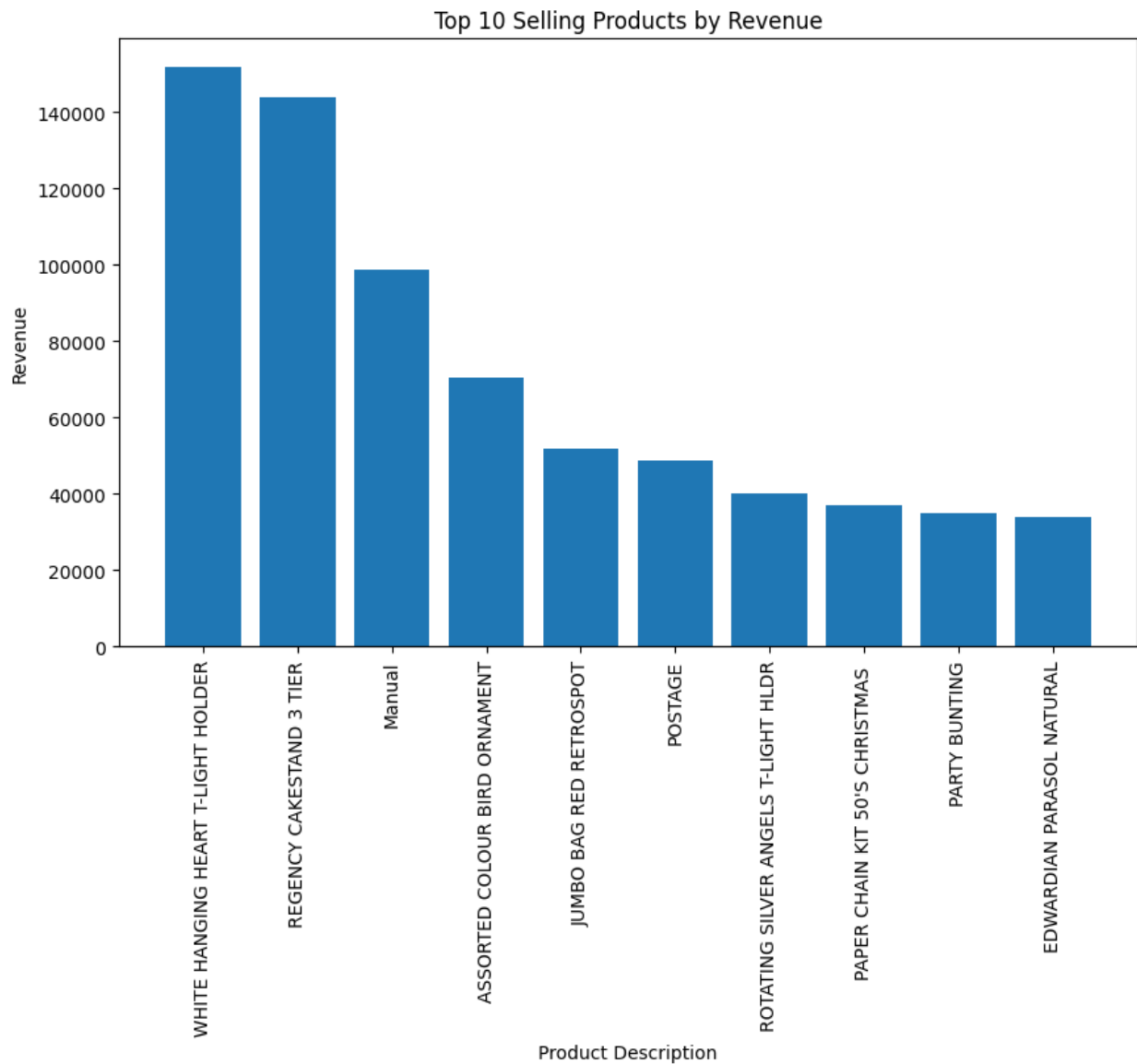
Here we can analyze that max orders are placed in the months of **September, October , November.**

Top-selling products in terms of Total Sales?

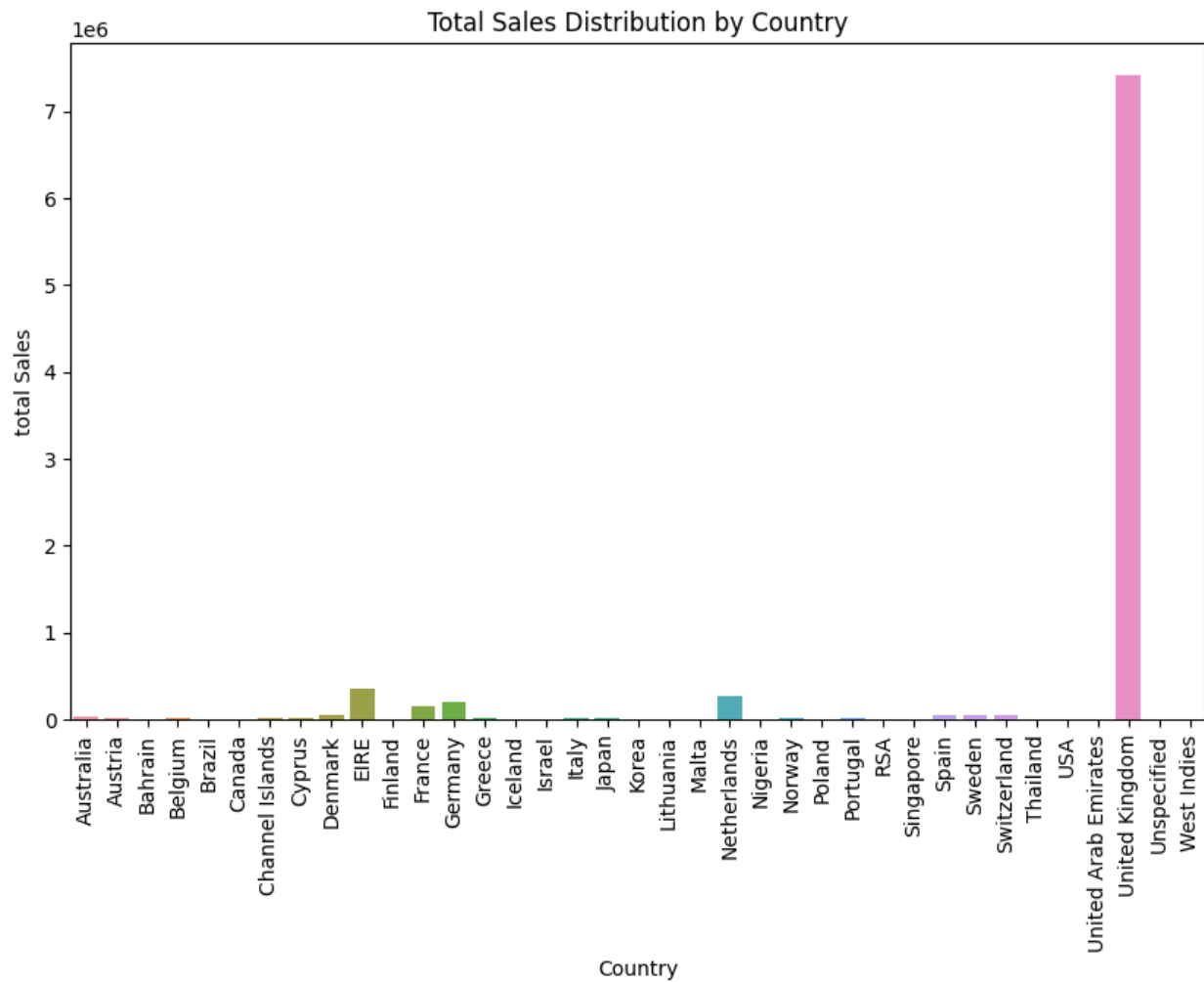
```

StockCode
85123A    151624.31
22423     143893.35
M         98560.64
85099B    85395.90
84879     70493.83
POST      48741.08
21843     41718.34
48138     41275.99
84347     40186.65
22086     36933.50
Name: TotalSales, dtype: float64

```



The distribution of revenue across different countries?



We can see that the **United Kingdom, EIRE, Netherlands, Germany, France** have the highest sales Distribution among all countries in the given dataset.

How many orders are placed per month on average

Average orders per month : **1601.25**

Average revenue per order

Average revenue per order: **459.64107593026495**

The customer retention rate

```

Month
1      1.404167
2      1.428941
3      1.441816
4      1.410828
5      1.425466
6      1.438040
7      1.488147
8      1.419319
9      1.475109
10     1.424850
11     1.609832
12     1.772270
dtype: float64

```

Customer retention rate as the ratio of repeat customers to total customers. Here we have calculated the CRR month wise which gives 1 means if Customer purchases anything from the store he will be satisfied with the services and make the purchase again.

Milestones

After performing the EDA on the given dataset of store transactions, we have come up with some interesting insights:

- Customers can be segmented into different categories based on their purchase behavior such as frequent buyers, occasional buyers, and one-time buyers. The categories can help in creating targeted marketing strategies.
- Loyal customers can be defined as those who make repeated purchases over a period of time. They are an important asset to the business as they contribute significantly to the revenue.
- The most popular time of year for sales is during the months of November and December, which can be attributed to the holiday season.
- The data shows seasonality with the highest sales occurring in November and December, followed by a decline in January and February.

- The customer lifetime value can be calculated by analyzing the purchasing behavior of customers and their loyalty towards the store.
- The top selling products in terms of revenue are 85099B, M, 22423 and 85123A. Their names are :“ Assorted color bird ornament”, “Manual”, “Regency Cake Stand 3-tier”, “White hanging heart t-light holder”.
- The box plot analysis of top-selling products shows that the product 22423 has a higher median and a larger spread compared to other products.
- The average price of the products sold by the store is around 3.5 GBP.
- Average orders per month: 1601.25
- Average revenue per order: 459.64107593026495

Based on these insights, the retail business can make better decisions regarding inventory management, marketing strategies, and customer retention programs.