# CUSTOMERS TRANSACTION ANALYSIS

## OVERVIEW

This report uses customers transactions data to gain insights into customer behavior and optimize sales, revenue, customer retention, and churn.

## PROBLEM STATEMENT

The purpose is to analyze customer purchasing transactions and gain insight into customer behavior to efficiently and proactively target customers, with the goal of increasing sales/revenue, improving customer retention, and reducing churn.

## DATASET EXPLORING

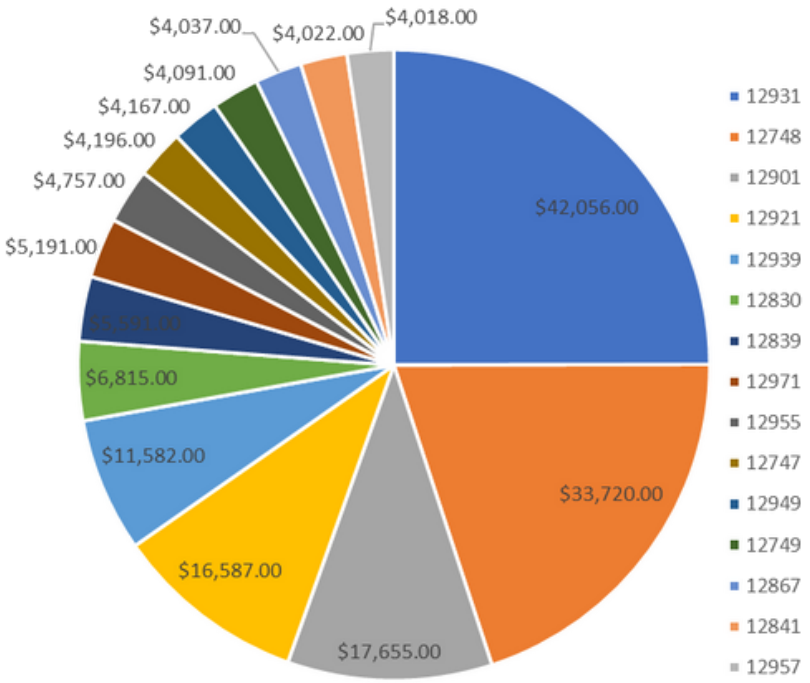
The OnlineRetail dataset contains 12858 rows of retail transactions data. Each row represents a purchase made by a customer and includes information such as the invoice number, stock code, quantity, invoice date, price, customer ID, and country.

| STOCKCODE | INVOICE | QUANTITY | INVOICEDATE | PRICE | CUSTOMER_ID | COUNTRY |
|---|---|---|---|---|---|---|
| 85124C | 537215 | 12 | 12/5/2010 15:38 | 2.55 | 12747 | United Kingdom |
| 85124B | 537215 | 6 | 12/5/2010 15:38 | 2.55 | 12747 | United Kingdom |
| 84879 | 537215 | 16 | 12/5/2010 15:38 | 1.69 | 12747 | United Kingdom |
| 85062 | 537215 | 24 | 12/5/2010 15:38 | 1.65 | 12747 | United Kingdom |
| 85064 | 537215 | 6 | 12/5/2010 15:38 | 5.45 | 12747 | United Kingdom |
| 82484 | 537215 | 36 | 12/5/2010 15:38 | 5.55 | 12747 | United Kingdom |
| 21136 | 537215 | 8 | 12/5/2010 15:38 | 1.69 | 12747 | United Kingdom |
| 22795 | 538537 | 16 | 12/13/2010 10:41 | 5.95 | 12747 | United Kingdom |
| 48138 | 538537 | 2 | 12/13/2010 10:41 | 7.95 | 12747 | United Kingdom |
| 82494L | 538537 | 24 | 12/13/2010 10:41 | 2.55 | 12747 | United Kingdom |
| 84879 | 538537 | 24 | 12/13/2010 10:41 | 1.69 | 12747 | United Kingdom |
| 85062 | 538537 | 12 | 12/13/2010 10:41 | 1.65 | 12747 | United Kingdom |

## 1- The total revenue generated by each customer:

This information is important to understand the contribution of each customer to the overall revenue.
It helps the business to identify its most valuable customers and focuses its marketing strategies and promotions to retain them.
It can also identify customers who are not contributing much to the revenue and develop targeted strategies to increase their purchases.

| CUSTOMER_ID | REVENUE |
|---|---|
| 12931 | 42055.96 |
| 12748 | 33719.73 |
| 12901 | 17654.54 |
| 12921 | 16587.09 |
| 12939 | 11581.8 |
| 12830 | 6814.64 |
| 12839 | 5591.42 |
| 12971 | 5190.74 |
| 12955 | 4757.16 |
| 12747 | 4196.01 |
| 12949 | 4167.22 |
| 12749 | 4090.88 |
| 12867 | 4036.82 |



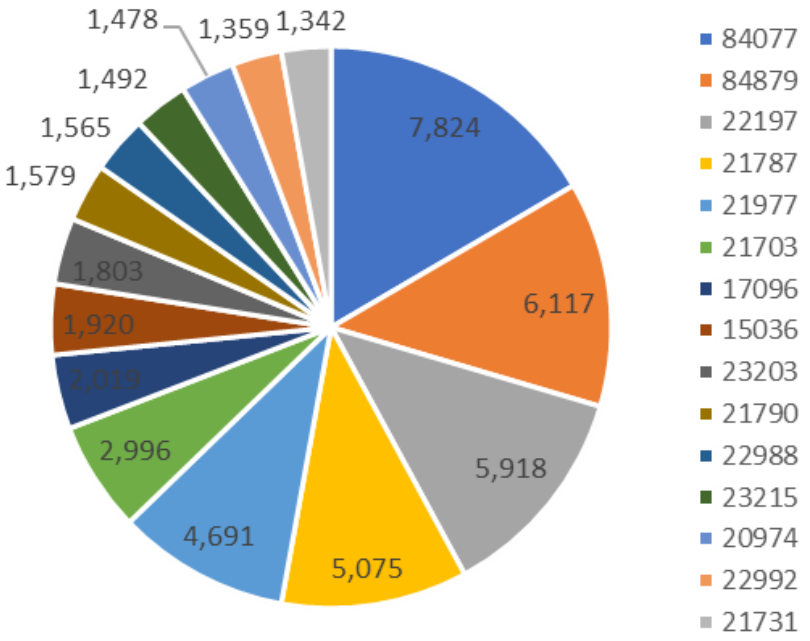

**$255,722.00**
**Total Revenue**

The chart shows the top 15 customers per revenue from total of 110 customers and their revenue percentage to the total revenue.

## 2- The top-selling products by quantity:

This information is important to identify the top-selling products by quantity.
It helps to provide insights into customer preferences and demand.
The business can determine which products are most and less popular and focus the promotions according to them.
It can also help with pricing strategies.
In addition, it can help the business to forecast demand and plan for future inventory and purchasing needs.

| STOCKCODE | TOTAL_QUANTITY |
|---|---|
| 84077 | 7824 |
| 84879 | 6117 |
| 22197 | 5918 |
| 21787 | 5075 |
| 21977 | 4691 |
| 21703 | 2996 |
| 17096 | 2019 |
| 15036 | 1920 |
| 23203 | 1803 |
| 21790 | 1579 |
| 22988 | 1565 |
| 23215 | 1492 |



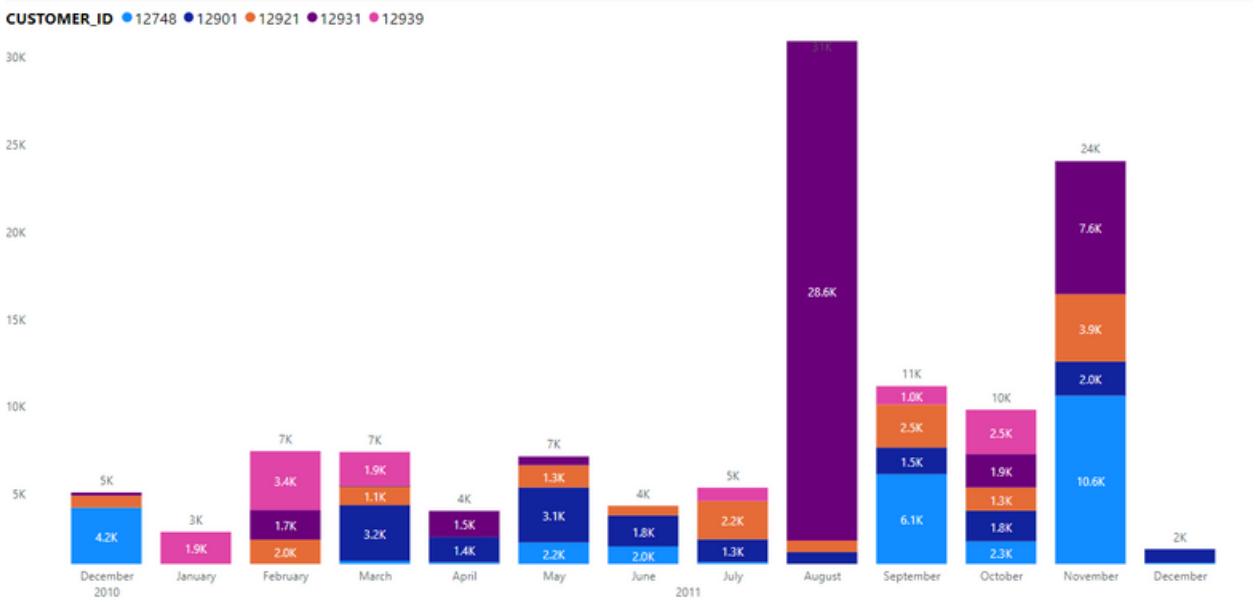

**176,653**
**Total Quantities**

The chart shows the top selling 15 products per quantity and their percentage to the total quantities. The business may consider offering a discount on these products to encourage more sales.

## 3- The monthly revenue for the top 5 customers:

This information is important to provide insight into the monthly revenue generated by the top 5 customers of a business.
Businesses can identify trends and patterns in customer behavior, and use this information to make data-driven decisions on how to optimize revenue generation from these high-value customers.

| CUSTOMER_ID | ORDERMONTH | REVENUEPERMONTH | TOTALREVENUE | RNK |
|---|---|---|---|---|
| 12931 | 02-2011 | 1696.4 | 42055.96 | 1 |
| 12931 | 03-2011 | 62.5 | 42055.96 | 1 |
| 12931 | 04-2011 | 1488 | 42055.96 | 1 |
| 12931 | 05-2011 | 496.8 | 42055.96 | 1 |
| 12931 | 08-2011 | 28610 | 42055.96 | 1 |
| 12931 | 10-2011 | 1909.36 | 42055.96 | 1 |
| 12931 | 11-2011 | 7615.9 | 42055.96 | 1 |
| 12931 | 12-2010 | 177 | 42055.96 | 1 |
| 12748 | 01-2011 | 418.77 | 33719.73 | 2 |
| 12748 | 02-2011 | 389.64 | 33719.73 | 2 |
| 12748 | 03-2011 | 1179.37 | 33719.73 | 2 |
| 12748 | 04-2011 | 1100.37 | 33719.73 | 2 |
| 12748 | 05-2011 | 2234.5 | 33719.73 | 2 |

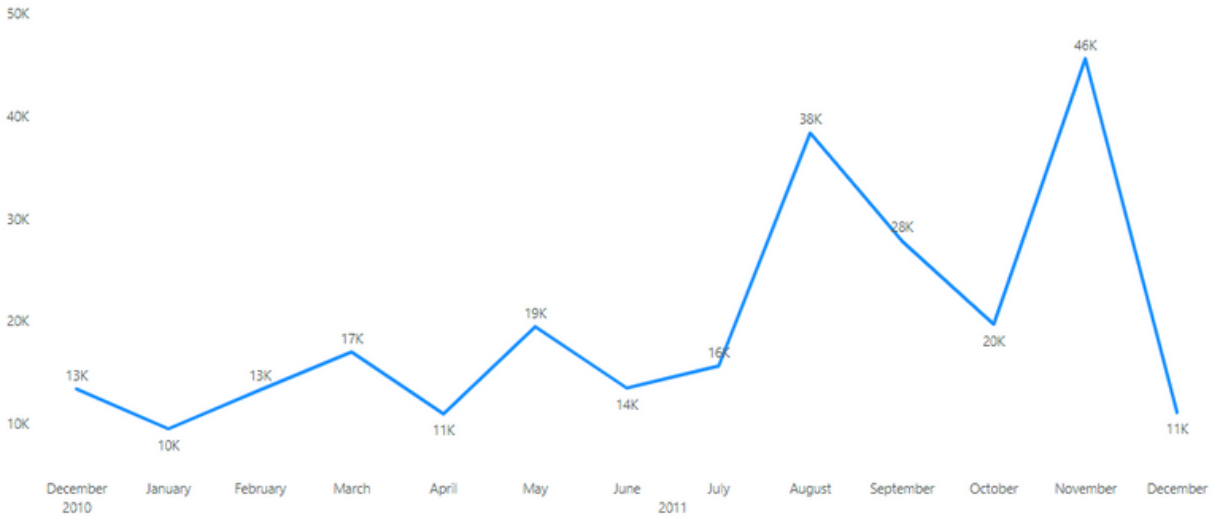



The chart shows the monthly revenue for the top 5 customers from December 2010 to December 2011. The average sales for all months are in the close range, but some months have very low sales averages and no sales from one or more of the high-value customers. The business needs to review the products and promotions that are provided in these months.

## 4- The total revenue per month:

This information is important to provide insights into how revenue varies from month to month in the period from December 2010 to December 2011. It can help identify any trends or patterns that may impact business decisions.

| ORDERMONTH | REVENUEPERMONTH |
|---|---|
| 2010-12 | 13422.96 |
| 2011-01 | 9541.29 |
| 2011-02 | 13336.84 |
| 2011-03 | 17038.01 |
| 2011-04 | 10980.51 |
| 2011-05 | 19496.18 |
| 2011-06 | 13517.01 |
| 2011-07 | 15664.54 |
| 2011-08 | 38374.64 |
| 2011-09 | 27853.82 |
| 2011-10 | 19735.07 |
| 2011-11 | 45633.38 |
| 2011-12 | 11124.13 |

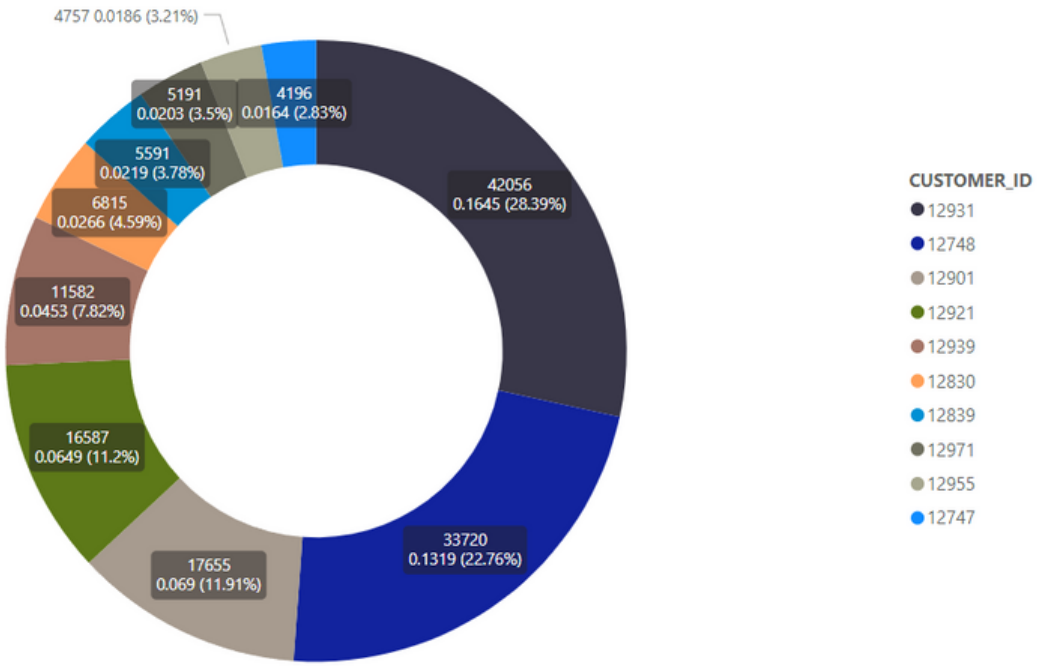



The chart shows the monthly revenue from December 2010 to December 2011. The average sales for all months are in the close range. Some months have dropped in sales averages. The business needs to investigate whether there were any issues with the supply chain or marketing strategies during these months.

## 5- The percentage of total revenue by each customer:

This information is important to provide valuable insights into the company's customer base and its contribution to overall revenue. By identifying the customers who generate the most revenue, businesses can optimize their marketing strategies and tailor promotions to retain their most valuable customers. Additionally, this analysis can help businesses understand their customer segmentation, including which segments contribute the most to their revenue and which segments may need further attention.

| CUSTOMER_ID | TOTAL_REVENUE | PERCENTAGE_OF_TOTAL_REVENUE |
|---|---|---|
| 12931 | 42055.96 | 16.45% |
| 12748 | 33719.73 | 13.19% |
| 12901 | 17654.54 | 6.9% |
| 12921 | 16587.09 | 6.49% |
| 12939 | 11581.8 | 4.53% |
| 12830 | 6814.64 | 2.66% |
| 12839 | 5591.42 | 2.19% |
| 12971 | 5190.74 | 2.03% |
| 12955 | 4757.16 | 1.86% |
| 12747 | 4196.01 | 1.64% |
| 12949 | 4167.22 | 1.63% |
| 12749 | 4090.88 | 1.6% |
| 12867 | 4036.82 | 1.58% |

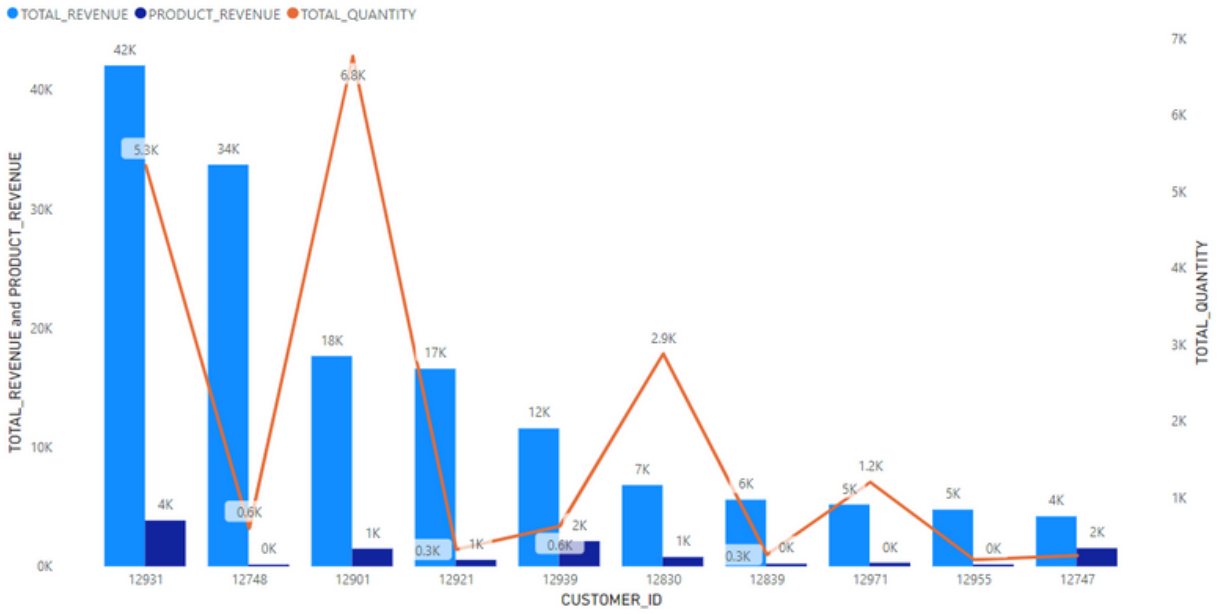



The chart shows the percentage revenue of each customer.

## 6- The top 10 customers by revenue and their most popular product:

This information is important to help the business identify the most profitable customers and the products that are raising their revenue, allowing them to make decisions on marketing and product offerings to maximize revenue, in addition, to making good marketing strategies or finding similar products that can attract them.

| CUSTOMER_ID | TOTAL_REVENUE | MOST_POPULAR_PRODUCT | TOTAL_QUANTITY | PRODUCT_REVENUE |
|---|---|---|---|---|
| 12931 | 42055.96 | 22197 | 5340 | 3844.8 |
| 12748 | 33719.73 | 21135 | 595 | 147.31 |
| 12901 | 17654.54 | 84077 | 6768 | 1482.72 |
| 12921 | 16587.09 | 84879 | 320 | 540.8 |
| 12939 | 11581.8 | 22570 | 624 | 2115.36 |
| 12830 | 6814.64 | 21703 | 2880 | 777.6 |
| 12839 | 5591.42 | 22197 | 252 | 214.2 |
| 12971 | 5190.74 | 40016 | 1200 | 300 |
| 12955 | 4757.16 | 15039 | 188 | 159.8 |
| 12747 | 4196.01 | 82484 | 240 | 1519.2 |

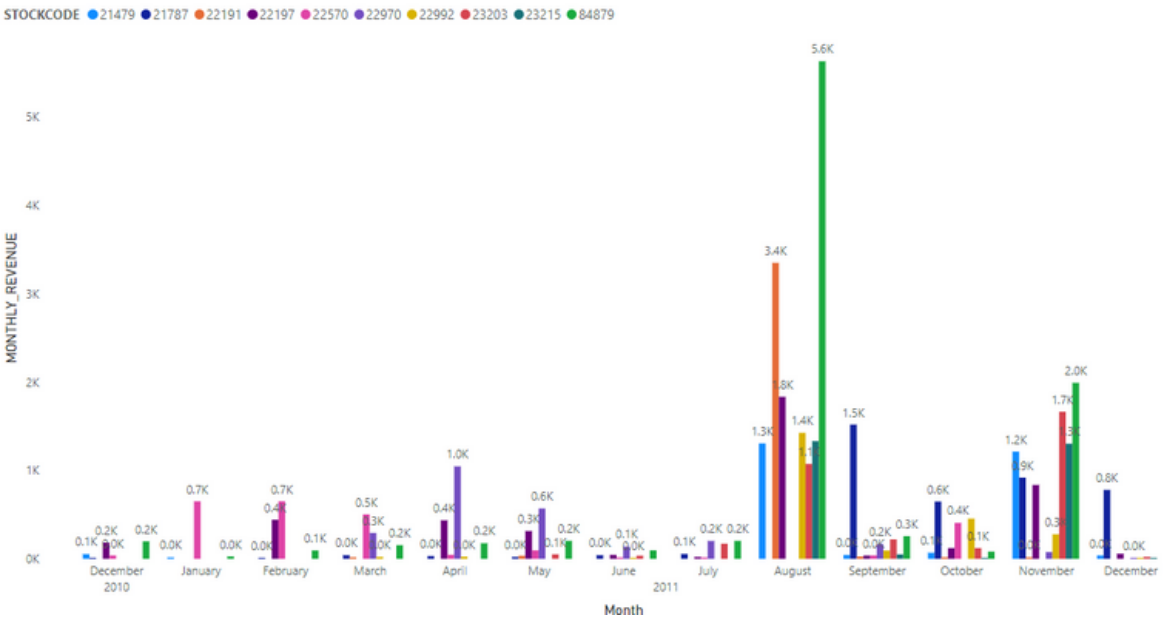



The chart shows the top 10 customers and their total revenue compared to the revenue generated from their top product, in addition to the quantity they bought from that product.

## 7- The top 10 products by revenue and their monthly revenue:

This information is important to allow the business to focus on the products that generate the most revenue and track their performance over time. By analyzing the monthly revenue of these products, the business can make data-driven decisions regarding inventory management, marketing strategies, and product pricing.

| STOCKCODE | MONTH | MONTHLY_REVENUE | TOTAL_REVENUE |
|---|---|---|---|
| 21479 | 2010-12 | 52.5 | 2736.01 |
| 21479 | 2011-01 | 15 | 2736.01 |
| 21479 | 2011-08 | 1305.51 | 2736.01 |
| 21479 | 2011-09 | 42.5 | 2736.01 |
| 21479 | 2011-10 | 68 | 2736.01 |
| 21479 | 2011-11 | 1214.25 | 2736.01 |
| 21479 | 2011-12 | 38.25 | 2736.01 |
| 21787 | 2010-12 | 3.4 | 4059.35 |
| 21787 | 2011-02 | 0.85 | 4059.35 |
| 21787 | 2011-03 | 40.8 | 4059.35 |
| 21787 | 2011-04 | 31.2 | 4059.35 |
| 21787 | 2011-05 | 20.4 | 4059.35 |
| 21787 | 2011-06 | 40.8 | 4059.35 |
| 21787 | 2011-07 | 55.25 | 4059.35 |
| 21787 | 2011-09 | 1519.8 | 4059.35 |
| 21787 | 2011-10 | 648 | 4059.35 |
| 21787 | 2011-11 | 918.85 | 4059.35 |
| 21787 | 2011-12 | 780 | 4059.35 |
| 22191 | 2011-03 | 17 | 3461.2 |

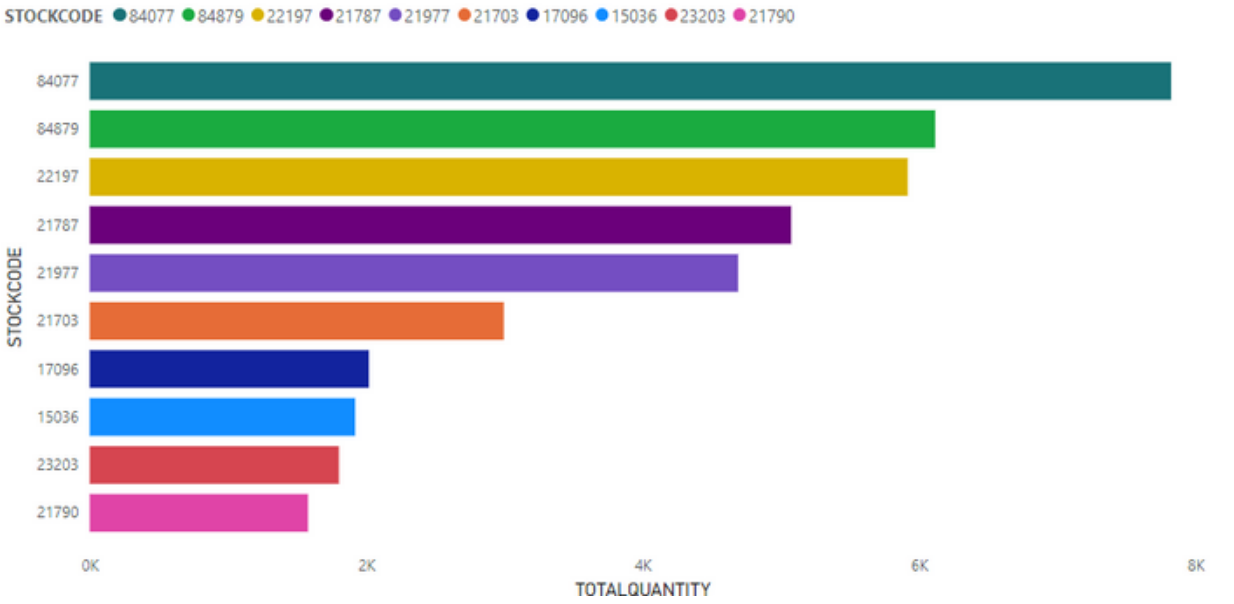



The chart shows the top 10 products and their monthly revenue. It shows that some months have a very low sales range.

## 8- The top 10 most frequently purchased items and their total revenue:

This information is important to identify the products that are in high demand among customers and generate significant revenue for the business. It can help the business make informed decisions about inventory management, marketing strategies, and product development. By understanding which products are selling the most, the business can focus on meeting customer demand and maximizing profitability.

| STOCKCODE | TOTALQUANTITY | TOTALREVENUE |
|---|---|---|
| 84077 | 7824 | 1788.96 |
| 84879 | 6117 | 9114.69 |
| 22197 | 5918 | 4323.1 |
| 21787 | 5075 | 4059.35 |
| 21977 | 4691 | 2063.69 |
| 21703 | 2996 | 826.32 |
| 17096 | 2019 | 343.23 |
| 15036 | 1920 | 1329.36 |
| 23203 | 1803 | 3357.44 |
| 21790 | 1579 | 1011.67 |

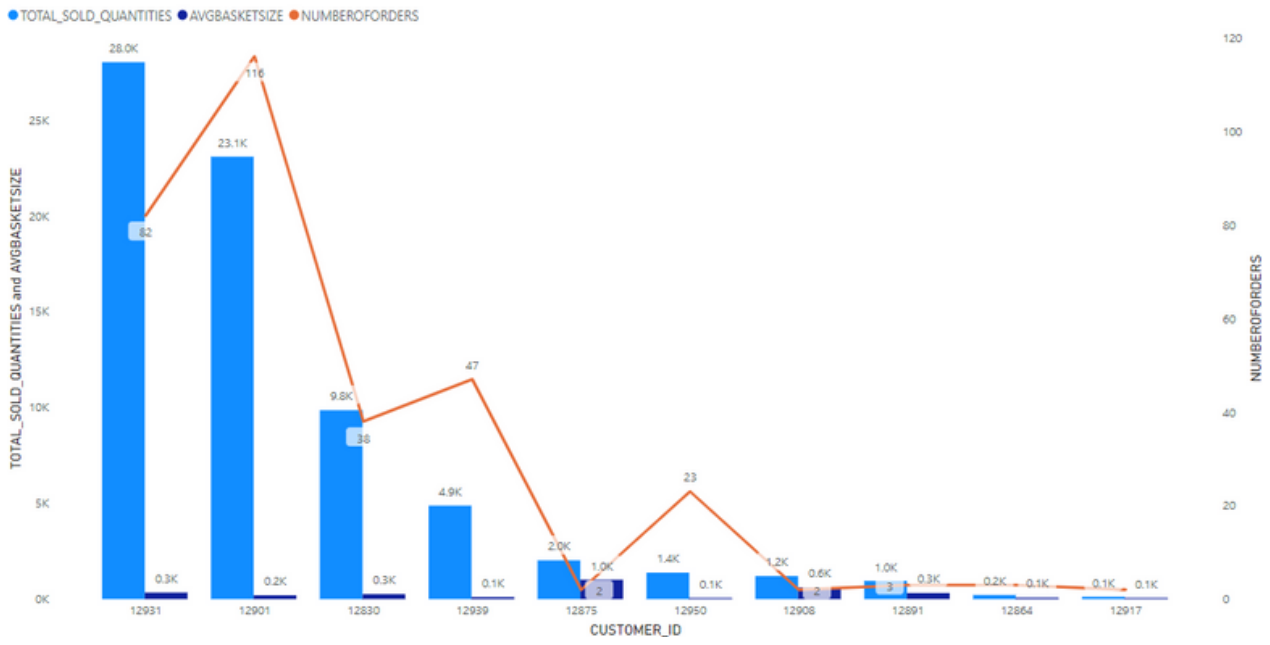



The chart shows the top 10 most purchased products and their sold quantity.

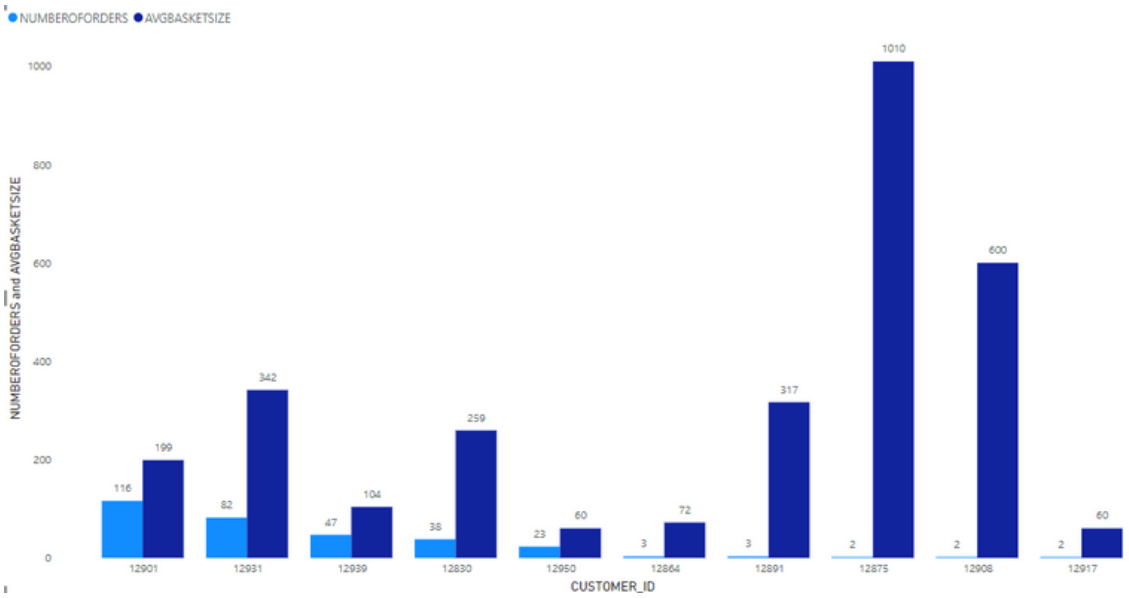
## 9- The average basket size (average number of items purchased per transaction) for each customer:

This information is important to analyze customer purchasing behavior and preferences. It helps identify high-value customers, predict future sales, improve customer experience, and optimize inventory management.

| CUSTOMER_ID | NUMBEROFORDERS | TOTAL_SOLD_QUANTITIES | AVGBASKETSIZE |
|---|---|---|---|
| 12875 | 2 | 2019 | 1009.5 |
| 12908 | 2 | 1200 | 600 |
| 12931 | 82 | 28004 | 341.51 |
| 12891 | 3 | 950 | 316.67 |
| 12830 | 38 | 9848 | 259.16 |
| 12901 | 116 | 23075 | 198.92 |
| 12939 | 47 | 4876 | 103.74 |
| 12864 | 3 | 216 | 72 |
| 12917 | 2 | 120 | 60 |
| 12950 | 23 | 1380 | 60 |
| 12823 | 5 | 230 | 46 |
| 12863 | 5 | 188 | 37.6 |
| 12879 | 10 | 370 | 37 |
| 12829 | 11 | 376 | 34.18 |
| 12873 | 4 | 128 | 32 |
| 12913 | 50 | 1419 | 28.38 |
| 12912 | 24 | 670 | 27.92 |
| 12906 | 63 | 1749 | 27.76 |




The chart shows the top 10 customers by average basket size and the total quantities that they bought compared to the average basket size and the number of orders.




The chart shows the top 10 customers by average basket size compared to the number of orders. It shows that some customers have a high average basket size, which suggests that the customer is a frequent shopper and can be offered personalized discounts or promotions. On the other hand, some customers have a low average basket size, which indicates that the customer needs more incentives to increase their spending and offers them promotions or discounts to encourage them to purchase more.

## MONETARY MODEL

Implementing Mentory Model or  RFM Segmentation is important to separate a group of customers into subgroups of customers according to their behavior for product purchasing. This analysis makes it easier for business to tailor their products and services to meet each segment's needs.

This model segments each customer into the following groups:
**Champions** - **Loyal Customers** - **Potential Loyalists** – **Recent Customers** – **Promising** - **Customers Needing Attention** - **At Risk** - **Cant Lose Them** – **Hibernating** – **Lost**

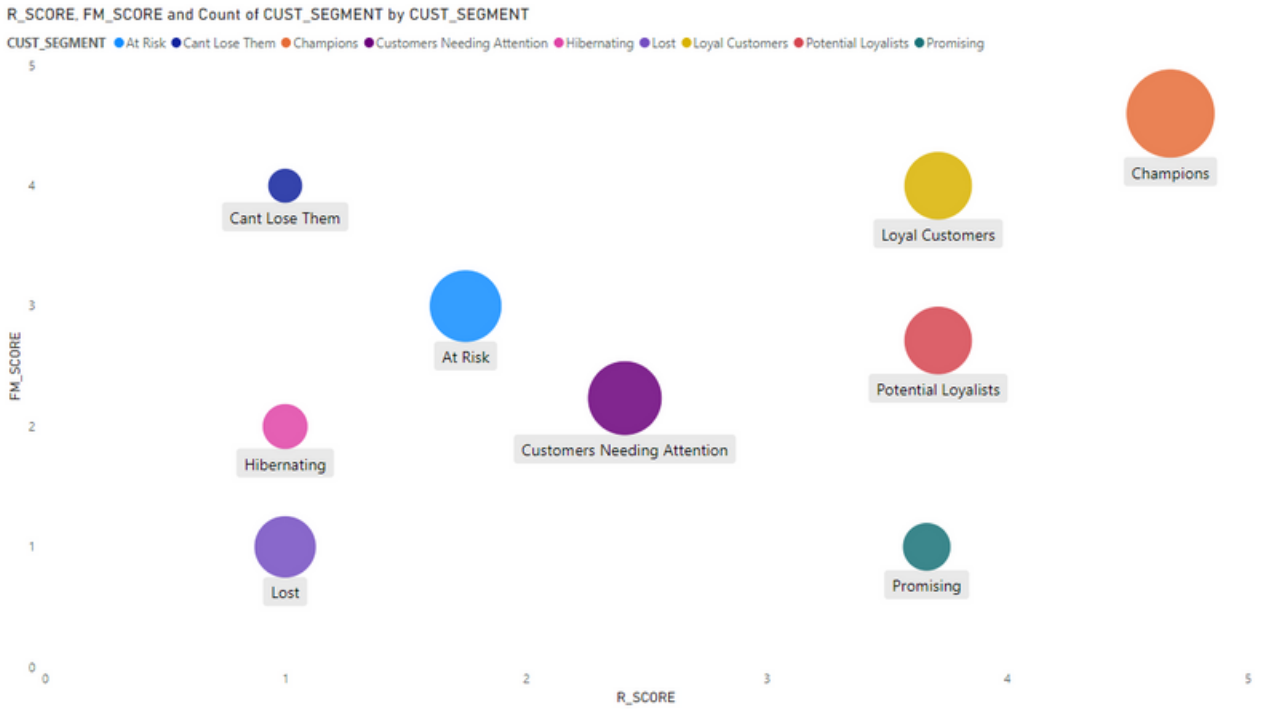
The customers will be grouped based on 3 main values
- **Recency** => how recent the last transaction is (Hint: choose a reference date, which is the most recent purchase in the dataset ).
- **Frequency** => how many times the customer has bought from our store.
- **Monetary** => how much each customer has paid for our products.

As there are many groups for each of the R, F, and M features, there are also many potential permutations, this number is too much to manage in terms of marketing strategies.
For this, so the permutations would be decreased by getting the average scores of the frequency and monetary (as both of them are indicative of purchasing volume anyway).

Label each customer based on the below values:

| Group name | Recency score | AVG(Frequency & Monetary ) score |
|---|---|---|
| Champions | 5 | 5 |
| | 5 | 4 |
| | 4 | 5 |
| Potential Loyalists | 5 | 2 |
| | 4 | 2 |
| | 3 | 3 |
| | 4 | 3 |
| Loyal Customers | 5 | 3 |
| | 4 | 4 |
| | 3 | 5 |
| | 3 | 4 |
| Recent Customers | 5 | 1 |
| Promising | 4 | 1 |
| | 3 | 1 |
| Customers Needing Attention | 3 | 2 |
| | 2 | 3 |
| | 2 | 2 |
| At Risk | 2 | 5 |
| | 2 | 4 |
| | 1 | 3 |
| Cant Lose Them | 1 | 5 |
| | 1 | 4 |
| Hibernating | 1 | 2 |
| Lost | 1 | 1 |

| CUSTOMER_ID | RECENCY | FREQUENCY | MONETARY | R_SCORE | FM_SCORE | CUST_SEGMENT |
|---|---|---|---|---|---|---|
| 12864 | 138 | 1 | 147.12 | 2 | 1 | At Risk |
| 12923 | 64 | 1 | 176.97 | 3 | 1 | Promising |
| 12966 | 9 | 1 | 160.18 | 4 | 1 | Promising |
| 12831 | 262 | 1 | 215.05 | 1 | 1 | Lost |
| 12837 | 173 | 1 | 134.1 | 2 | 1 | At Risk |
| 12929 | 311 | 1 | 117.85 | 1 | 1 | Lost |
| 12968 | 112 | 1 | 135.95 | 2 | 1 | At Risk |
| 12834 | 282 | 1 | 312.38 | 1 | 1 | Lost |
| 12902 | 264 | 1 | 138.68 | 1 | 1 | Lost |
| 12852 | 294 | 1 | 311.55 | 1 | 1 | Lost |
| 12821 | 214 | 1 | 92.72 | 1 | 1 | Lost |
| 12893 | 30 | 1 | 188.14 | 3 | 1 | Promising |
| 12884 | 88 | 1 | 309.05 | 2 | 1 | At Risk |
| 12956 | 306 | 1 | 108.07 | 1 | 1 | Lost |
| 12851 | 96 | 1 | 135.18 | 2 | 1 | At Risk |
| 12829 | 336 | 2 | 293 | 1 | 1 | Lost |
| 12881 | 275 | 1 | 298 | 1 | 1 | Lost |
| 12873 | 282 | 1 | 374 | 1 | 1 | Lost |
| 12938 | 25 | 1 | 114.14 | 4 | 1 | Promising |
| 12890 | 24 | 1 | 380.47 | 4 | 1 | Promising |




The chart shows the customer segments based on the r_socre and fm_score, and the size of each segment (number of customers).

## DAILY PURCHASING CUSTOMERS' TRANSACTIONS:

The DailyCustomers dataset contains 574396 rows of daily purchasing transactions data for customers. Each row represents a purchase made by a customer and includes information such as the customer ID, purchasing date, and the amount.

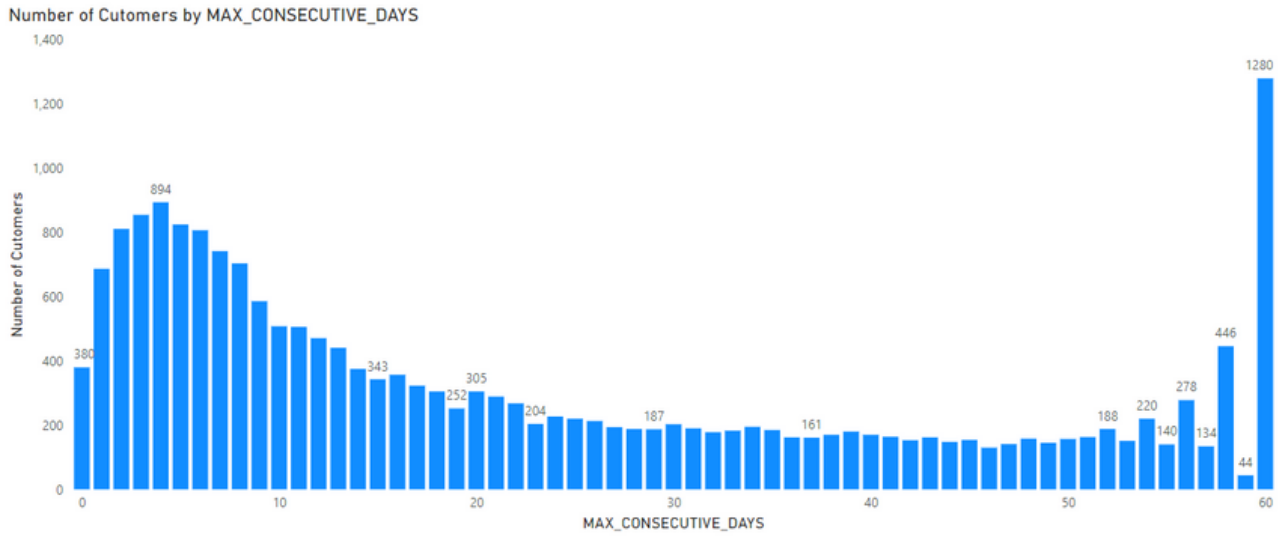
| CUST_ID | CALENDAR_DT | AMT_LE |
|---|---|---|
| 174868729 | 11/02/2019 | 0.34 |
| 182935636 | 11/01/2019 | 0 |
| 186715137 | 11/01/2019 | 1.6 |
| 152418025 | 11/01/2019 | 0.64 |
| 69562373 | 11/01/2019 | 9.62 |
| 175731818 | 11/01/2019 | 0.04 |
| 165892302 | 11/01/2019 | 0.2 |
| 31220765 | 11/01/2019 | 35.03 |
| 36721367 | 11/02/2019 | 0 |
| 122126398 | 11/01/2019 | 163.01 |
| 180133927 | 11/01/2019 | 121.05 |
| 102922885 | 11/01/2019 | 0.59 |
| 139473408 | 11/08/2019 | 12.56 |
| 176391069 | 11/01/2019 | 211.35 |
| 180056254 | 11/01/2019 | 3.86 |

## 1- The maximum number of consecutive days a customer made purchases:

This information is important because it provides valuable insights for businesses that aim to improve customer retention and increase revenue. It can recognize its loyal customers and tailor their marketing strategies to retain them. Businesses can also use this information to identify customers who may have stopped purchasing from them and try to re-engage them with targeted promotions or offers.

In addition, It can help businesses evaluate the effectiveness of their marketing campaigns and promotional activities. If a customer's maximum number of consecutive days of purchase increases after a particular marketing campaign, it indicates the campaign's success in engaging the customer.

| CUST_ID | MAX_CONSECUTIVE_DAYS |
|---|---|
| 625527 | 37 |
| 646553 | 8 |
| 664981 | 5 |
| 811892 | 43 |
| 1073321 | 3 |
| 1331618 | 12 |
| 1357409 | 20 |
| 1444226 | 3 |
| 1490919 | 3 |
| 1505992 | 34 |
| 1519955 | 60 |
| 1798791 | 5 |
| 1807965 | 13 |
| 1988054 | 18 |

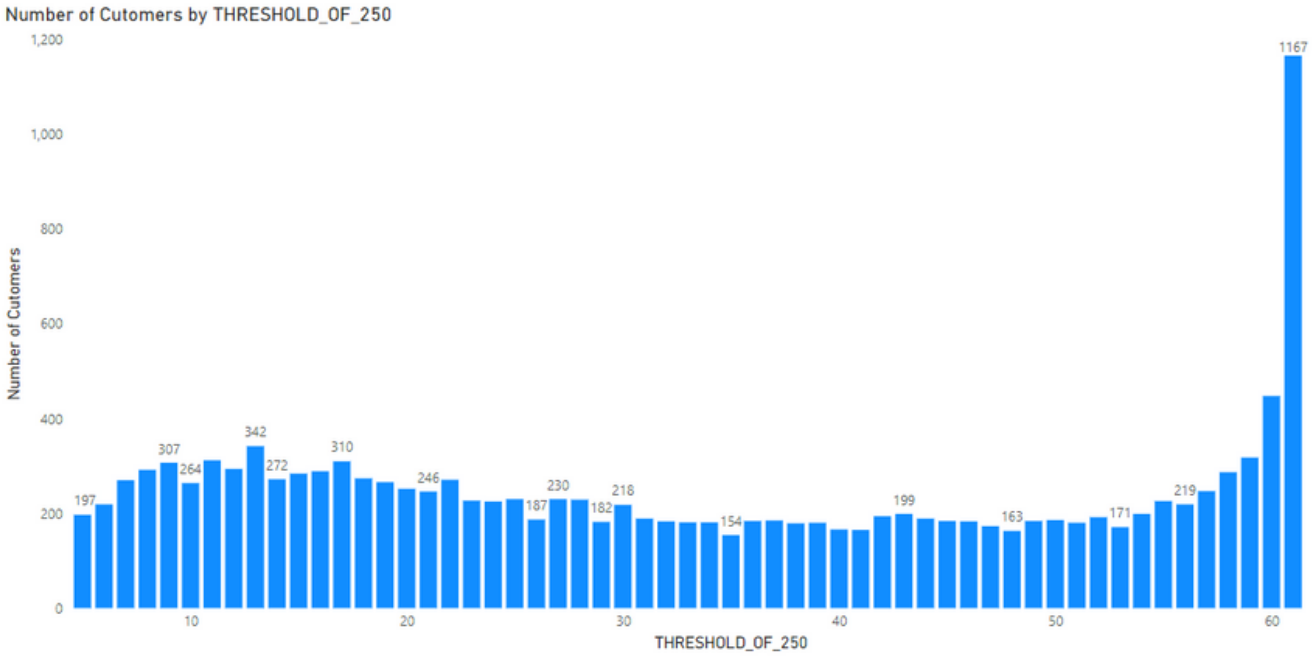


The chart shows the number of consecutive days starting from 0 to the maximum consecutive numbers for all customers. It also counts the customers that have a similar number of consecutive days.
It shows that the maximum consecutive days of 60 have the biggest number of customers which is 1280.

## 2- The number of days or transactions it takes a customer to reach a spent threshold of 250 L.E:

This information provides insight into customer behavior and spending patterns. It can help businesses understand how quickly their customers are willing to spend a certain amount and can be used to develop targeted marketing strategies and loyalty programs. Additionally, it can help identify customers who may be more likely to spend larger amounts and help businesses better tailor their offerings to meet their needs.

| CUST_ID | THRESHOLD_OF_250 |
|---|---|
| 151293 | 10 |
| 259866 | 22 |
| 664981 | 12 |
| 811892 | 52 |
| 1331618 | 18 |
| 1505992 | 44 |
| 2146867 | 60 |
| 2208179 | 23 |
| 2284100 | 61 |
| 2792126 | 34 |
| 2842713 | 59 |
| 3130002 | 47 |
| 3259850 | 6 |
| 3394425 | 61 |
| 3399415 | 16 |




The chart shows the number of days that customers take to reach a spent threshold of 250 L.E starting from 0 to the maximum number of days for all customers. It also counts the customers that have a similar number of days. It shows that the biggest percentage of customers which is about 1200 customers took more than 60 days to reach the threshold of 250 LE.

## CONCULUSION:

After exploring and analyzing the dataset, we have gained valuable insights which help with identifying some keys about customer behavior.

Using the RFM model, we were able to identify the customers who were most likely to make purchases and those who were at risk of churning. This information can be used to create targeted marketing campaigns and retention strategies to increase sales and customer loyalty.

In addition, we analyzed the maximum number of consecutive days that customers made purchases, and we found that a majority of customers make purchases every day, but there are also a significant number of customers who make purchases less frequently, And this can be used to identify loyal customers who may benefit from loyalty rewards programs or other incentives.

Finally, we found the average number of days or transactions it takes for customers to reach a spent threshold of 250 L.E, which can be used to identify potential high-value customers who may benefit from targeted marketing campaigns or loyalty programs.

Overall, the insights gained from analyzing the dataset can be used to inform strategic business decisions and drive growth in revenue and customer loyalty.

- All queries are explained with clean descriptions and comments.
- All output data are attached.