

# RFM Analysis

**Sameen Fatima**

### 13/12/2022

### —

### Portfolio Project 02

### —

### Data Analytics

Cohort 04

**About Dataset:**

A Data Set of Supply Chains by the company Data Co Global was used for the analysis. The data contains information about customers, orders and shipment details.

<https://data.mendeley.com/datasets/8gx2fvg2k6/5>

**Stakeholders:**

* Industrialists
* Suppliers
* Operational Managers
* Consumers
* Government (imposing international and national regulations)

**Objective:**

The goal is to perform RFM analysis in order to discover the answers to following questions:

* Who are our best customers?
* Which of our customers could contribute to our churn rate?
* Who has the potential to become valuable customers?
* Which of our customers can be retained?
* Which of our customers are most likely to respond to engagement campaigns?

**Instrumentation:**

The tools used in this process are:

* Python for Analysis
* Power BI for visualizations

**Methodology:**

**Loading and Cleaning Data in Python:**

* Corrected the data type of few cols.

df['order date (DateOrders)']=pd.to\_datetime(df['order date (DateOrders)'])

df["shipping date (DateOrders)"]=pd.to\_datetime(df["shipping date (DateOrders)"])

etc

* Deleted the duplicate columns

df.drop(["Benefit per order","Sales per customer","Order Item Cardprod Id","Order Item Product Price","Product Category Id","Order Customer Id"],axis=1,inplace=True)

We find something odd in this excel file. No customer having Customer Id less than 12436 has purchased in 2018. Also, after customer id 12436, all the customers have made purchase exactly once in the current year at that point of time.

Let’s dive further into data to figure out the issue.

This particular customer has purchased at 2017-10-02 12:46:00. Let’s see the data around this timestamp to see if we can find something unusual.

After looking at the data in excel we can say that even when customers were having same name and city, the different data points such as Customer Street and order region were different. From this we can infer that they are different individuals having same name and same city.

In this scenario we will remove the data after the above date time value and perform or RFM analysis on the remaining data because even if an old customer purchases from our store after above date time, his transaction wouldn’t be recorded in the system and his frequency and monetary value would not increase.

So, final data frame to perform RFM would be:

df\_clean=df\_clean[df\_clean["order date (DateOrders)"]<datetime\_val].copy()

There are 7754 orders where shipping has been canceled. Lets try to find why it may be happening.

df[df["Delivery Status"]=="Shipping canceled"].groupby("Order Status").agg(tc=("Order Id","count"))

We can see shipping has been canceled because customer canceled his order or order was marked as suspected fraud. We will have to remove these orders when we do RFM analysis as these were transaction that didn't add any value for store or for customer.

df\_clean=df[df["Delivery Status"]!="Shipping canceled"].copy()

So, this is the final dataset I applied RFM analysis on.

**Other Col measures:**

* Sales = order item product price \* order item Quantity
* Order Item total (sales per customer) = sales – (sales \* order item discount rate)
* Order Profit Per Order = Order Item total \* Order Item Profit Ratio
* Recency = DAYS ("1/31/2018", Y4) (excel)

**Quantiles:**

The data was then divided in 3 quantile range using formula and then assigned scores of recency, frequency, monetary accordingly.

**R- Score:**

Here we will find last purchase date and subtract it from the last recorded date in the database. This is done because dataset is not recent and last date in our cleaned data is of 2017.

df\_recency["recency"] = max\_date-df\_recency["last\_purchase\_date"]

We assigned a score of 3 to the first 33% of customers(0-33), score of 2 the next 33 % of customers(33-66 percentile), and score of 1 to the last percentile

Because the lesser the recency, more responsive is customer to promotional offers.

df\_recency['R'],Intervals\_Recency=pd.qcut(df\_recency["recency"],q=3,labels=[3,2,1],retbins=True)

**F- Score:**

df\_frequency["F"], Intervals\_Frequency = pd.qcut(df\_frequency["Frequency"], q=3,labels=[1,2,3],retbins=bool)

Here the opposite will be true. The higher the frequency, the more is value generated for the store.

So, (0-33 percentile) => 1, (33-66 percentile ) => 2, (66-100 percentile ) => 3

**M- Score:**

Monetary value represents total amount of all orders of a customer. Here more is better, so higher score will be assigned to higher monetary value.

df\_monetory["M"],Intervals\_Monetory=pd.qcut(df\_monetory["Sum\_of\_Sales"],q=3,labels=[1,2,3],retbins=True)

**RFM score:**

Now we will join all three values of recency, frequency, and monetary on Customer Id. We will find a combine score of RFM for each Customer Id. Then we will segment the customers based on that score.

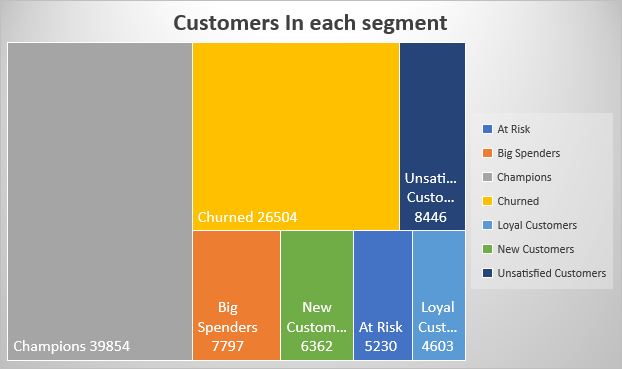
df\_rfm=pd.merge(df\_recency[["Customer Id","R"]],df\_frequency[["Customer Id","F"]],on="Customer Id",how="inner")

df\_rfm=pd.merge(df\_rfm,df\_monetory[["Customer Id","M"]],on="Customer Id",how="inner")

df\_rfm["RFM"]=(df\_rfm["R"]).astype(str)+(df\_rfm["F"]).astype(str)+(df\_rfm["M"]).astype(str)

RFM scores are now calculated to know the purchase behavior of any customer.

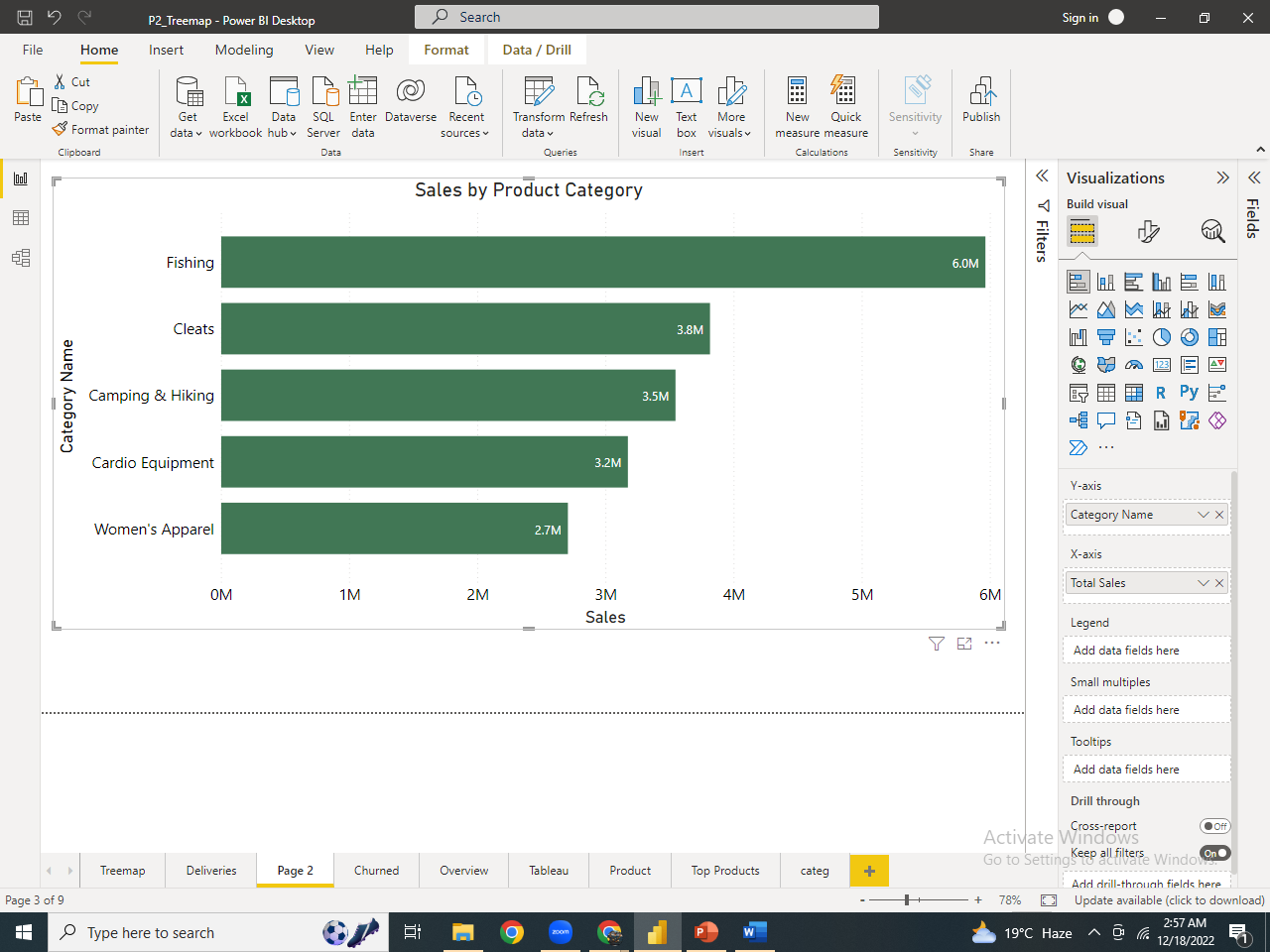
For example, a customer with R = 3, F = 3, and M = 3 is most profitable and loyal customer, while a customer with R = 1, F = 1, and M = 1 is the almost churned customer.

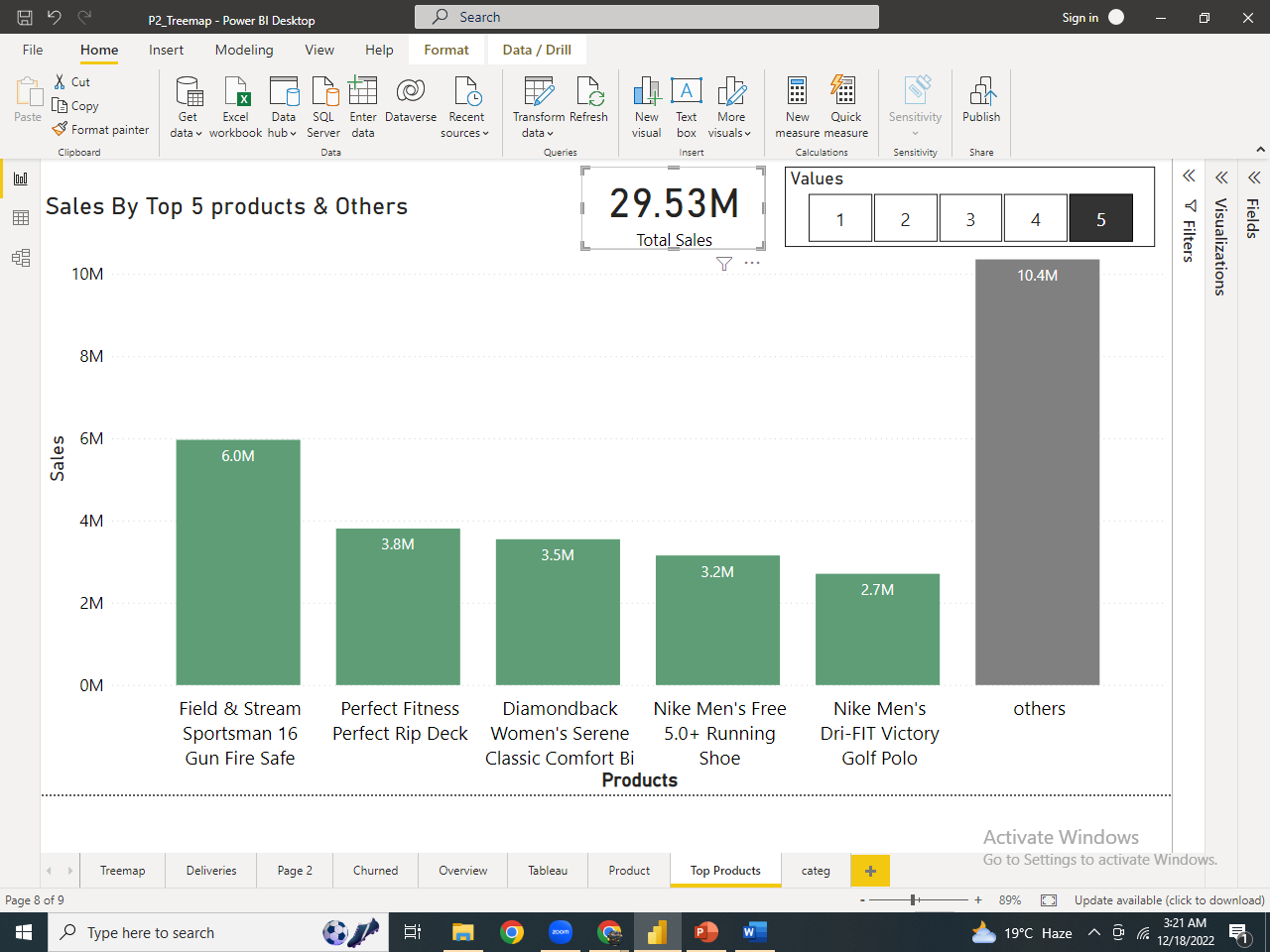


**Visualizations In Power BI:**

**Insights and Recommendations:**

Let’s first discover the products and their categories which are most popular among our ‘Champions’ and ‘Big spenders’.



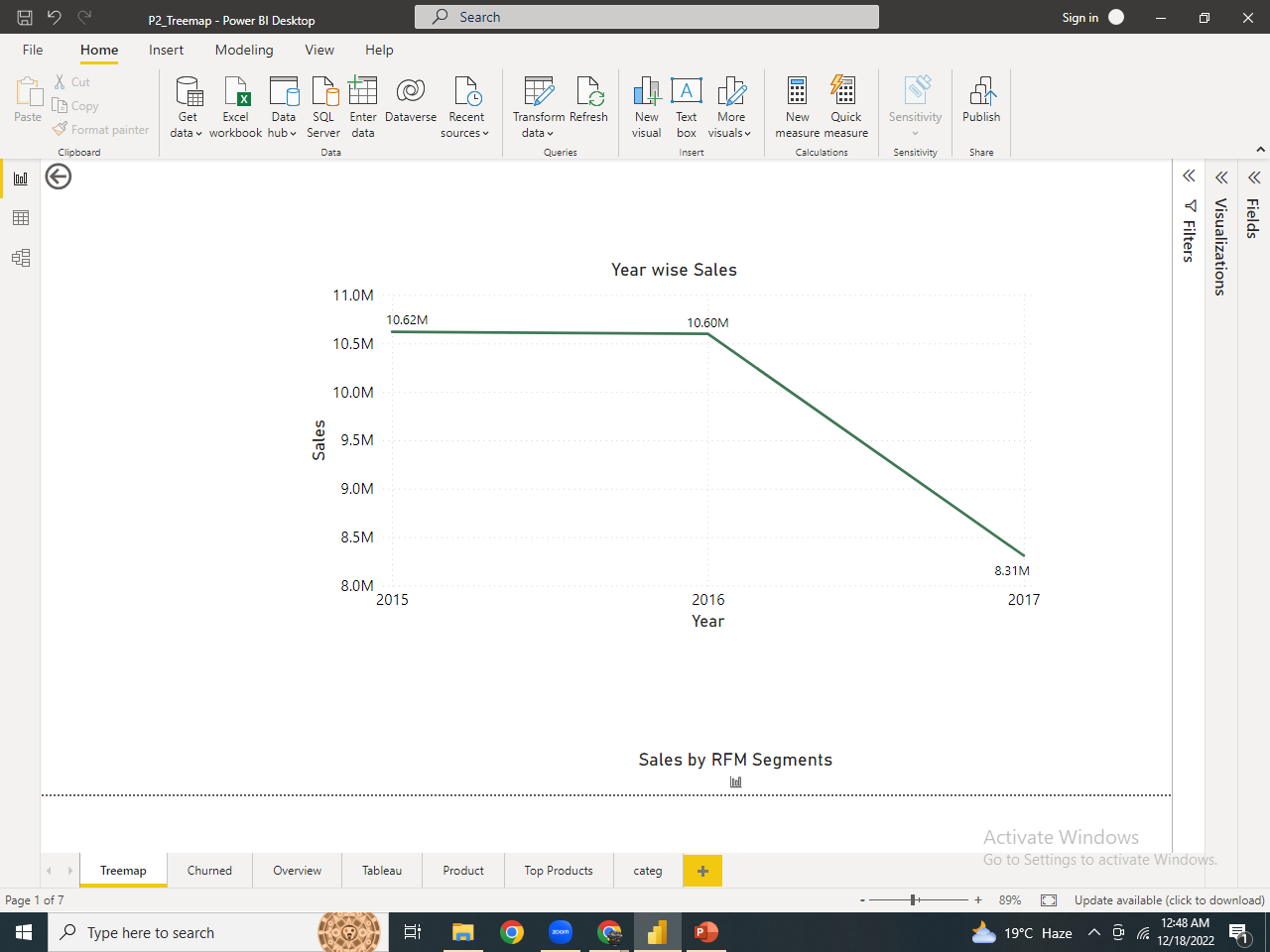


We can see that Out of 29.35M of sales, these 5 products are producing 19.2M and rest of the products are generating only 10M of sales. So, these 5 are the most important Products of our company.

*Recommendation:*

*Because 65% of the revenue was generated by these 5 products so we must never compromise on the Quality of these product in order to gain customer’s satisfaction & loyalty.*

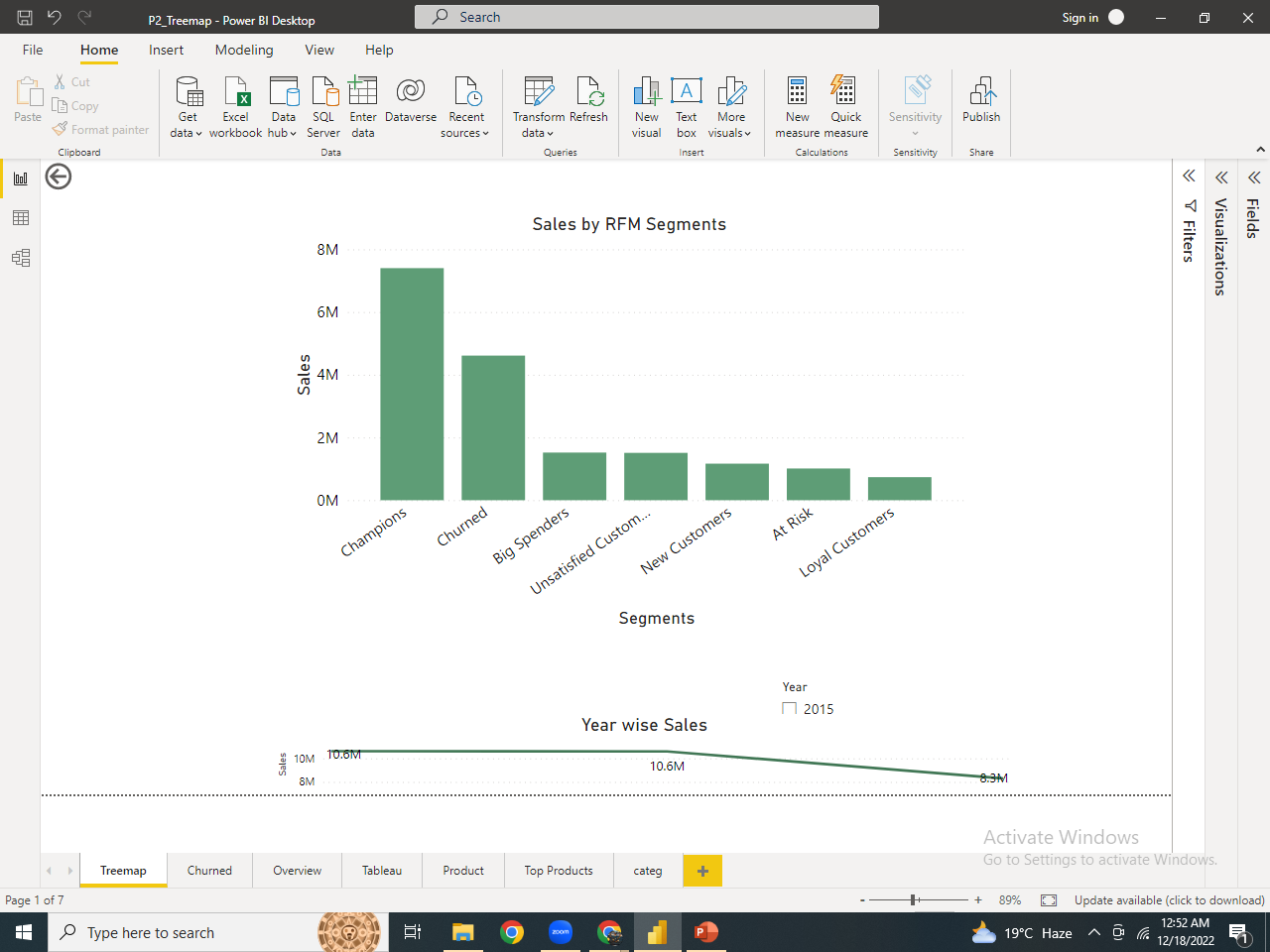
**Year-wise Sales:**



Then we analyzed sales and discovers that it was critically declining in 2017.

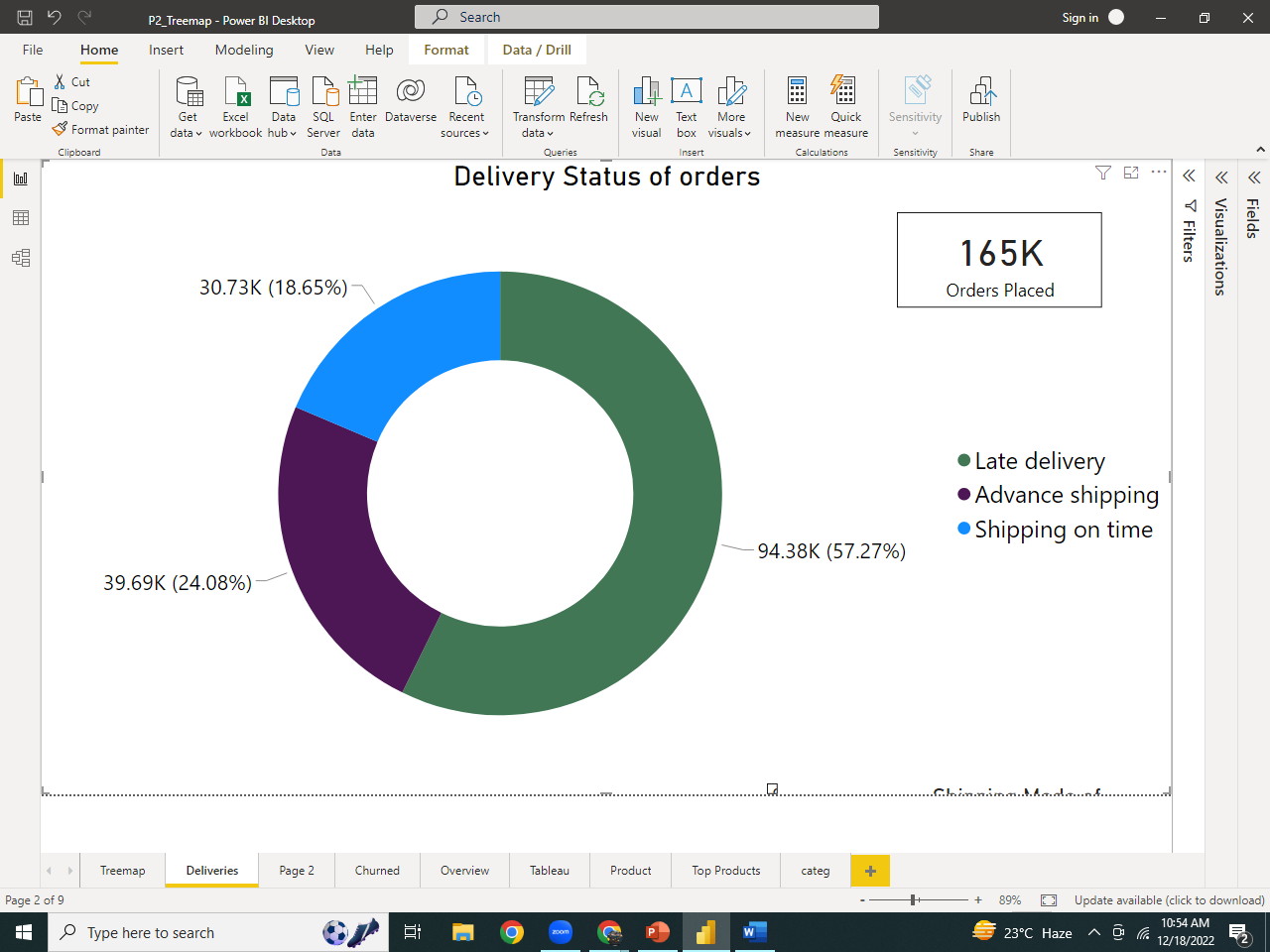
So, we went more deeper to find out the reason behind this decline.

**Sales generated by each segment**



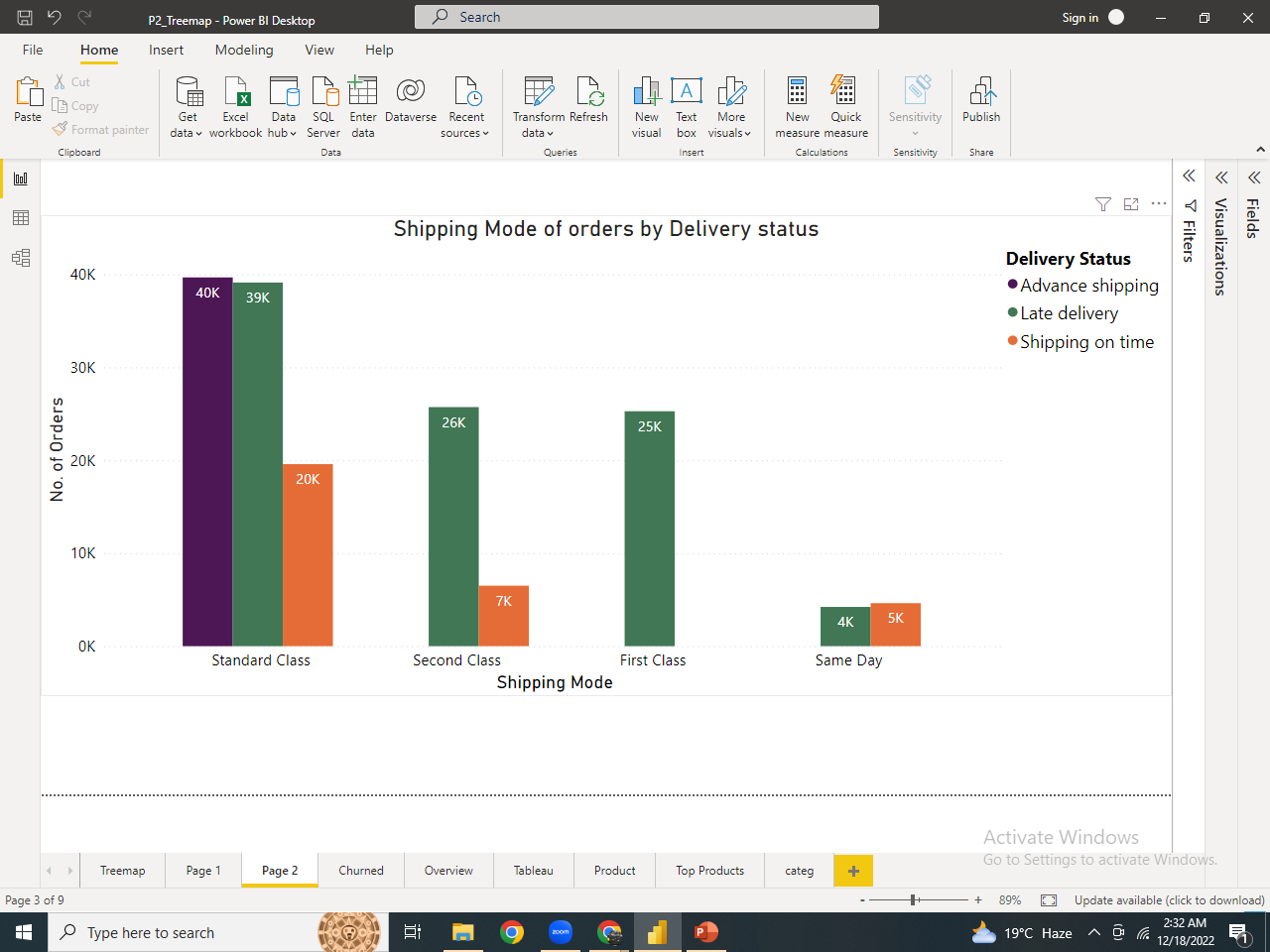
We discovered the reason why sales declines in 2017 is none but the churning of customers. Drilling further.

**Why are Customers Churning?**



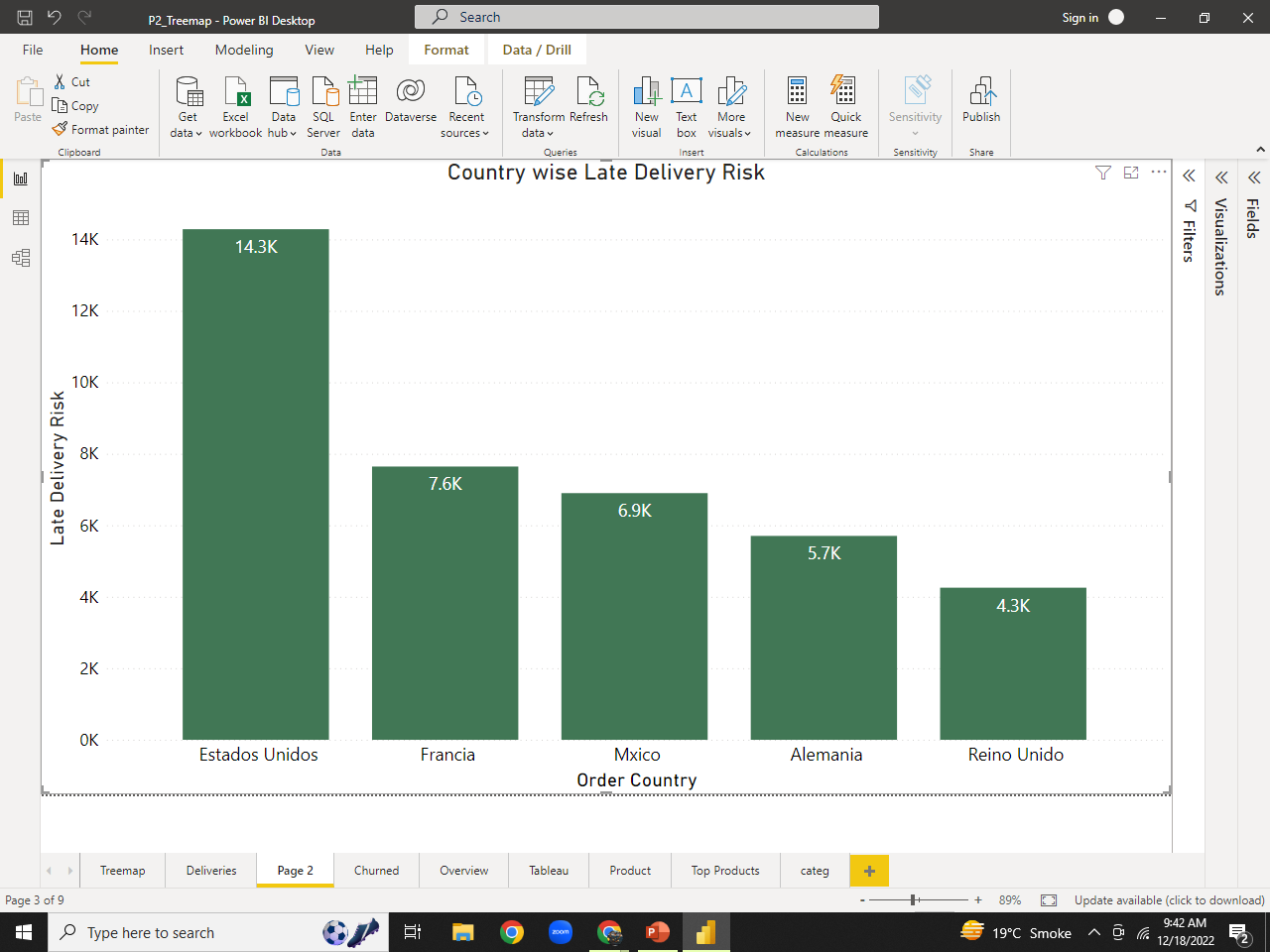
Out of 165K orders, 94.38K were delivered late. Which could be the main reason of customer’s churn.

**Shipping Mode of orders by Delivery status**

****

The visual clearly depicts the first-class deliveries were all delayed & orders were delivered in advance only in Standard class.

**Country wise Late Delivery Risk**



* Estados Unidos, A country in North America has the high risk of late Deliveries which is quite alarming and needed to be fixed soon.
* 39% of buyers are less likely to shop in our store again if their purchase is not delivered within the promised delivery days and hence contributing in churn rate.

**Overall Late Delivery Risk in Countries:**



*Recommendation:*

*We can reduce customer churn and assure them fast & smooth deliveries by:*

* *communicating & monitoring driver’s performance.*
* *Automating route planning and driver dispatch.*
* *Providing customers real-time driver tracking.*

Strategies to be adapted

|  |  |
| --- | --- |
| Segments | Actionable Tips |
| champions | Customers in this segment visits the store frequently and buys quality products even if they are expensive. The brand value and exclusiveness of the products matter for them. Send them promotional messages of exclusive products or when new stocks arrive at the store. |
| Big spenders | They look for quality and exclusivity regardless of money. We can dive into the specifics of the purchases individually for each customer, and try to offer them similar deals on their next purchase to retain them. |
| New customers | They are new customers who have recently visited our stores. Provide them the best service. Try to retain them by offering discounts on their next visit. |
| Loyal customers | They visit our store frequently and look for the quantity and quality both. Offer them discounts and promotional offers to retain them. |
| At risk | They purchase from us occasionally, less frequently but are big spenders. They are looking for quality and exclusivity and when they find it out our store, they purchase from us. we can dive into the specifics of the purchases individually for each customer and try to offer them similar deals on their next purchase to retain them |
| Unsatisfied Customers | Their low frequency indicates that they were not happy with the experience at the stores. We can also call them reluctant buyers. Our aim should be to convert them to Loyal customers or big spenders by communicating with them and providing good after sales services. |
| Churned | They were once loyal customers but now have churned. Try to win them back by offering good deals and special discounts.  Also assure them fast, smooth deliveries as late deliveries was the main reason of customer churn. |
|  |  |

**Thank You!**