DECEMBER 13, 2022

MARKET ANALYTICS

REVENUE ANALYSIS AND CUTOMER SEGMENTATION

HADIQA MALIK

IEC CAPSTONE PROJECT

Business Problems & Project Plan

The following table outlines all the objectives of this Capstone in a broader term. The plan of this project is to utilize my skills in the Marketing domain of a business to come up with the basic insights that a marketing team needs to make future strategies.

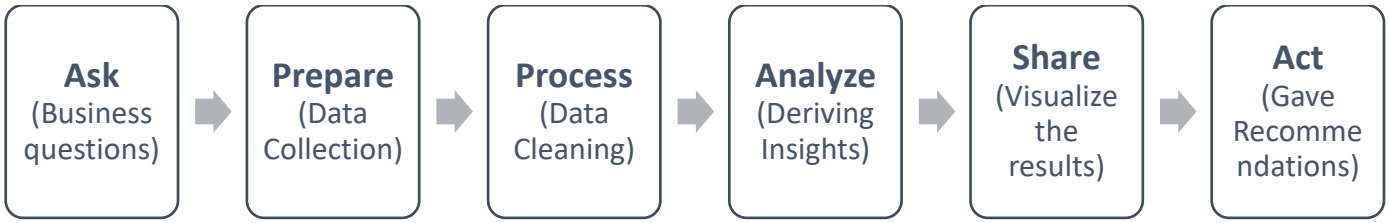
This is a descriptive analysis that may include some feature engineering (extracting new columns from the existing ones) and derived insights about the existing target market and its revenue with Customer Segmentation using RFM Model. These results assist the company to match its results with its objectives and analyze the gap. Also, it will lead the marketing team to design future marketing decision by keeping its target market and potential customers in mind and obtain increasing results than before.

Title	Marketing Analytics
Industry Focus	Online Retail
Problem Statement	Study Customers' data to design to: • Analyze potential customers and increase revenue • Allocate marketing budget according to customer type to attain optimum results
Business Use Case (What am I solving?)	• Market Analytics helps businesses understand the performance and impact of their marketing investments. • Businesses use these tools to facilitate the collection, modeling, analysis, and visualization of marketing data and make data-driven decisions.
Goals/Matrices	• Top Products and Location Analysis • Revenue Analysis • Customer Segmentation using RFM Modeling • Most Profitable and Churn Customers Rate
Deliverables	• A document outlining the detailed data Analysis process • Python for Revenue Analysis • Dashboard for RFM Modeling • A presentation that includes findings and recommendations •
Dataset Link	Marketing Dataset Link
Dataset Attributes	• Unique Invoice, Stock and Customer ID • Product Description with Quantity and Unit Price • Location

Table of Contents

Data Analysis Process

This project is based on following steps:



1. Ask Phase

Key tasks

1. Identify the business questions
2. Consider key stakeholders

a. Business Questions

The business questions revolve around **Customer Segmentation**, and **Revenue Analysis** among different customers' attributes.

1. What are the monthly orders and revenue patterns of the company?
2. What are the weekly orders and revenue pattenrs of the company?
3. What are the hourly orders and revenue patterns of the company?
4. Which product is performing well in terms of revenue?
5. What is the ratio of customers (total no. of customers) in each country along with the revenue generated?
6. Who are the top and low customers from customer segments?
7. Who has the potential to become valuable customers?
8. Which of your customers are most likely to respond to engagement campaigns?
9. From which segment, the company is generating most of its revenue?
10. What is the churn rate of the company?

b. Stakeholders

This case study could have multiple stakeholders from different departments within a company

- **Marketing Head:** Learn about Customer Segments using RFM model to optimize marketing strategies and design customer–centric marketing ad campaigns to increase the retention rate and generate more revenue.
- **Customers:** As the marketing activities will have an impacted on the response of customer

2. Prepare and Process Phase

Data Analysis

The analysis will be done in Python and divided into two parts

The data is based on two parts:

1. Revenue Analysis (Python)
2. RFM Analysis (PowerBI)

Data Collection and Source

The data is a secondary data imported from third party source. It is structured in attributes and records. Their attributes are qualitative and quantitative in nature.

Data Source Link: https://archive.ics.uci.edu/ml/datasets/online+retail#

Process of Data Analysis in Python

Data Cleaning

1. Libraries Importing

2. Data Cleaning
   - Drop Duplicates (total 5268 duplicates have been removed)
   - Check Null Values (rows with null values has been removed)
   - Check Data types (Recheck all data types to make sure that the data is consistent)
   - Removing Extra Spaces

3. Data Transformation
   - Extracting New Column "Revenue" (Price * Revenue)
   - Divide a data frame into three parts based on
     - Fulfilled Orders (Orders started with **5** are *fulfilled orders*)
     - Cancelled Orders (Orders started with **c** are *Cancelled orders*)
     - Adjusted bad debt orders (Orders started with **a** are *Adjusted orders*)

4. Data Exploration and Analysis
   - Top Customers Month and Year Wise, Week and Hour wise
   - Top Selling Products
   - Top Countries

5. Data Visualization

Visualize the following aspects of the dataset:

1. What is the total number of customers in each month?
2. What is the Total Revenue Month and Year Wise?
3. What are the total orders week and hour wise?
4. What is the top number of orders fulfilled in each country?
5. What is the total revenue generated in each country?
6. What are the top selling products (quantity wise)?
7. What are the top selling products (revenue wise)?

3. Analyze and Share Phase

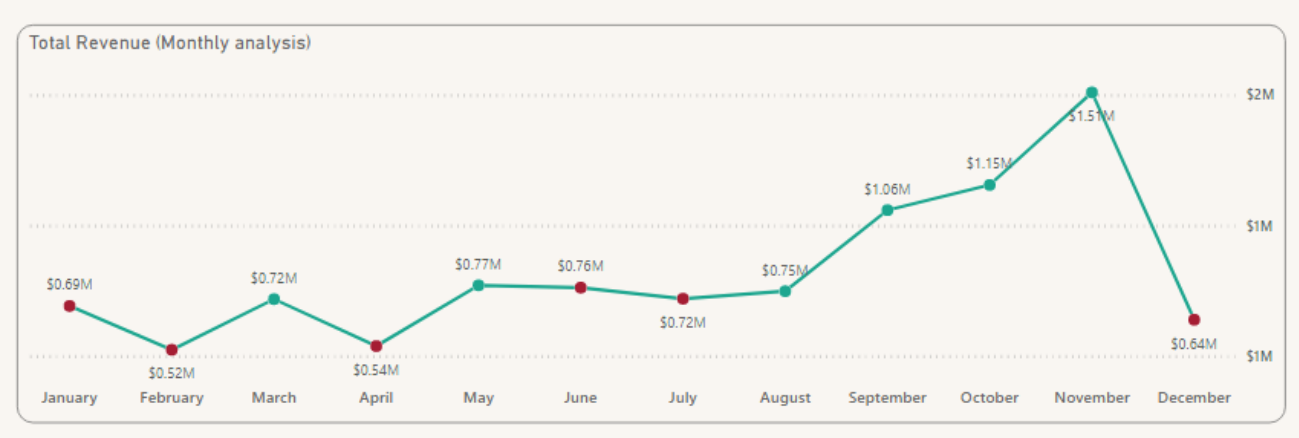
Analysis and Visualizations

At this stage, my goal was to analyze the Revenue of the business across certain attributes such as Country, Products, and quantity wise.

Along with that, I will analyze Customer Segmentation to help the marketing team to curate campaign that fulfill the needs of their customers based on their purchasing history.

Monthly Analysis

Year	Month	Total Revenue	Revenue MoM%
2010	December	$823,746	
2011	January	$691,365	-16.07%
2011	February	$523,632	-24.26%
2011	March	$717,639	37.05%
2011	April	$537,809	-25.06%
2011	May	$770,536	43.27%
2011	June	$761,740	-1.14%
2011	July	$719,221	-5.58%
2011	August	$748,076	4.01%
2011	September	$1,058,590	41.51%
2011	October	$1,154,979	9.11%
2011	November	$1,509,496	30.69%
2011	December	$638,793	-57.68%
Total		**$10,655,622**	**6.38%**



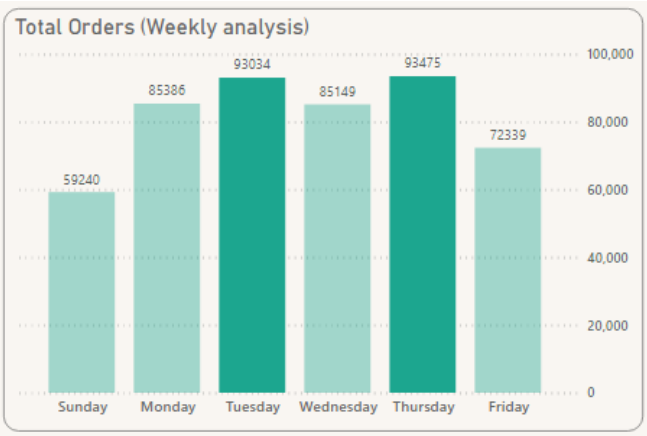
Based on the above graph, November is the month with the highest sales and February is the month with the lowest sales. The overall month over month growth is 6.38%. Half of the months shows a decrease in sales as compared to previous months and 06 months shows an increase in sales as compared to previous sales.

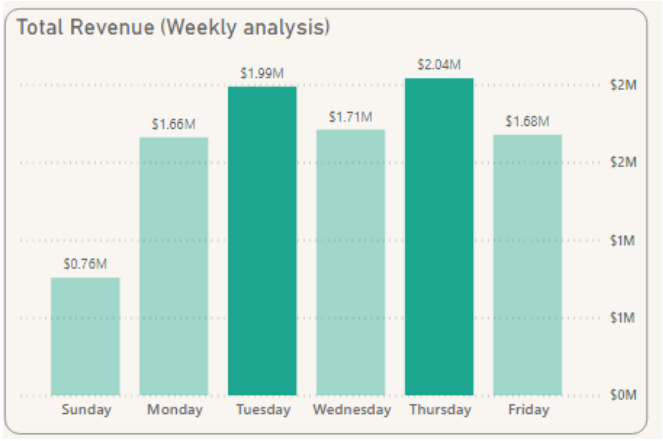
The overall Revenue is **$ 10 Million.**

Month	Total Orders	Orders MoM%
January	34306	
February	27105	-20.99%
March	35803	32.09%
April	29096	-18.73%
May	36164	24.29%
June	35977	-0.52%
July	38645	7.42%
August	34482	-10.77%
September	49261	42.86%
October	59304	20.39%
November	83369	40.58%
December	25111	-69.88%
Total	**488623**	**5.42%**

- Based on the above graph, November is the month with the highest orders and February is the month with the lowest sold quantity. 05 months faced decrease in sold quantity and rest of the month showed positive results. Overall annual increase in orders is 6%.

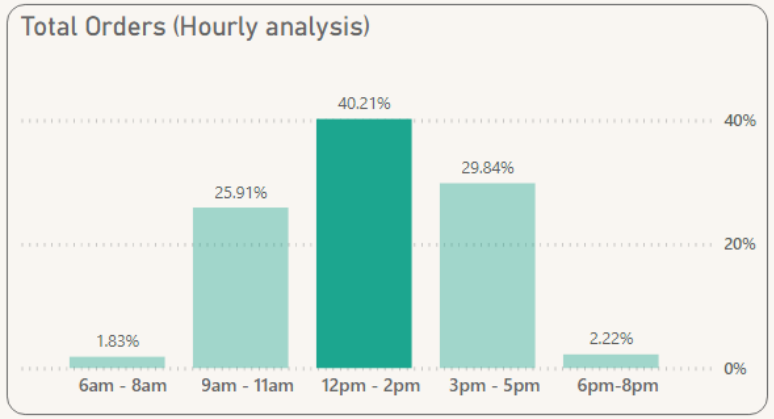
Weekly Analysis





- Mostly orders have received on Thursday followed by Tuesday and revenue generation which means that Customers are mostly active in these two days. This helps to analyze the peak time of orders receiving.

Hourly Analysis



- 40.21% orders have received from 12pm-2pm. This is the peak time of orders receiving. That's why, this graph helps to analyze the most responsive time of customers.

Customer Segmentation:

Customer segmentation is a marketing process in which the customers are divided into different groups based on their demographics, purchasing habits and interests. The purpose of this process is to analyze which customers are profitable for the business and which are not. In which way, marketing department of an organization can make customized campaigns to retain these customers and maximize their ROI (Return on investment). RFM Model is one of the ways for customer segmentation.

RFM (Recency, Frequency and Monetary) Model:

RFM (Recency, Frequency & Monetary) analysis is a behavior-based technique used to segment customers by examining their purchasing history based on three factors:

- **Recency:** How recently a customer has purchased?
- **Frequency:** How often do they purchase?
- **Monetary:** How much the customer spends?

It is based on the marketing axiom that:

"80% of your business comes from 20% of your customers."

RFM helps to identify customers who are more likely to respond to promotions by segmenting them into various categories.

RFM Model is based on three steps:

1) Calculate RFM Metrices/Values
2) Building RFM Scores and RFM Segment
3) Grouping into Name Segments

RFM Table Building

1. Calculate RFM Metrices/ Values

Calculate the following columns according to each customer.

- **Recency:** Their most recent purchase date.
- **Frequency:** Number of purchases within a set time (i.e., one year).
- **Monetary:** Total sales from that customer

a. Recency Value

Last transaction date of each customer

```
last transaction date = MAXX(FILTER('orders','orders'[NewID]='orders'[NewID]),'orders'[InvoiceDate])
```

Extracting the difference between the last date of the period and the last data of each customer visit

```
Recency = DATEDIFF('orders'[last transaction date], orders[Last_date],DAY)
```

b. Frequency Value

Unique Counts of Invoices/Order number

```
Frequency = DISTINCTCOUNT('orders'[InvoiceNo])
```

c. Monetary Value

Total Sum of Revenue

```
Monetary = SUM('orders'[Revenue])
```

Creating Tables

Now create an RFM table using the modeling tab and DAX. Here summarize function groups the data based on customer IDs.

```
RFM table = SUMMARIZE(
    'orders','orders'[NewID],
    "R Value",[Recency],
    "F Value",[Frequency],
    "M Value",[Monetary])
```

2. Building RFM Scores and RFM Segment

Assign a number from 1 to 5 for each category, where 5 is the highest and 01 is the lowest. To calculate RFM, we´ll use the above three columns that has been created in the previous step.

Now create three measures or Columns in the same RFM Table to calculate the R, Fand M scores.

```
F Score =
SWITCH (
  TRUE (),
   [F value] <= PERCENTILE.INC ( 'RFM table'[F Value], 0.20 ), "1",
   [F value] <= PERCENTILE.INC ( 'RFM table'[F Value], 0.40 ), "2",
   [F Value] <= PERCENTILE.INC ( 'RFM table'[F Value], 0.60 ), "3",
   [F value] <= PERCENTILE.INC ( 'RFM table'[F Value], 0.80 ), "4",
   "5"
      )
```

```
R Score =
SWITCH (
  TRUE (),
   [R value] <= PERCENTILE.INC ( 'RFM table'[R Value], 0.20 ), "5",
   [R value] <= PERCENTILE.INC ( 'RFM table'[R Value], 0.40 ), "4",
   [R Value] <= PERCENTILE.INC ( 'RFM table'[R Value], 0.60 ), "3",
   [R value] <= PERCENTILE.INC ( 'RFM table'[R Value], 0.80 ), "2",
   "1"
      )
```

```
M Score =
SWITCH (
  TRUE (),
   [M Value] <= PERCENTILE.INC ( 'RFM table'[M Value], 0.20 ), "1",
   [M value] <= PERCENTILE.INC ( 'RFM table'[M Value], 0.40 ), "2",
   [M Value] <= PERCENTILE.INC ( 'RFM table'[M Value], 0.60 ), "3",
   [M value] <= PERCENTILE.INC ( 'RFM table'[M Value], 0.80 ), "4",
   "5"
      )
```

Create the new column called 'RFM' by concatenating column 'R Score', 'F Score', and 'M Score'

```
RFM = 'RFM table'[R Score]& 'RFM table'[F Score]&'RFM table'[M Score]
```

3. Grouping into Name Segments

RFM Table Model

Following is the RFM table. This simple approach of scaling customers from 1-5 will result in, at the most, 125 different RFM scores (5x5x5), ranging from 111(lowest) to 555(highest). Each RFM cell will differ in size and vary from one another, in terms of the customer's key habits, captured in the RFM score.

Segment	Scores
Champions	555, 554, 544, 545, 454, 455, 445
Loyal	543, 444, 435, 355, 354, 345, 344, 335
Potential Loyalist	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
New Customers	512, 511, 422, 421 412, 411, 311
Promising	525, 524, 523, 522, 521, 515, 514, 513, 425,424, 413,414,415, 315, 314, 313
Need Attention	535, 534, 443, 434, 343, 334, 325, 324
About To Sleep	331, 321, 312, 221, 213, 231, 241, 251
At Risk	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
Cannot Lose Them	155, 154, 144, 214,215,115, 114, 113
Hibernating customers	332, 322, 231, 241, 251, 233, 232, 223, 222, 132, 123, 122, 212, 211
Lost customers	111, 112, 121, 131,141,151

RFM Segments Result:

RFM Segments

Segment	Total Customers	% of Customers	Total Revenue	% of Revenue
Champions	1004	17.42%	6,067,323.21	56.94%
Cannot lose them	546	9.47%	1,107,027.45	10.39%
Loyal Customers	524	9.09%	1,020,390.10	9.58%
Promising	375	6.50%	885,510.63	8.31%
At Risk	267	4.63%	415,579.87	3.90%
Need Attention	273	4.74%	409,495.72	3.84%
Potential Loyalist	504	8.74%	255,891.09	2.40%
Hibernating	739	12.82%	176,828.28	1.66%
About to Sleep	311	5.39%	136,547.24	1.28%
Lost	670	11.62%	101,772.86	0.96%
New Customer	552	9.58%	79,256.03	0.74%
Total	**5765**	**100.00%**	**10,655,622.48**	**100.00%**

This is the result of RFM segmentation for the given dataset.

- **26%** are Champions and Loyal Customer and more than **50%** revenue is generated from Champions customers followed by “Can’t lose them” customers.
- **25%** customers are Potential Loyalist, New Customers, and Promising customers and generating **10%** revenue. They can be converted to Loyal customers.
- **10%** customers need attention and about to sleep as they are in above and below RFM metrices only generating **5%** revenue.
- **14%** are “At Risk” and “Can’t lose them” and generating **14%** Revenue.
- **13%** are hibernating and Churn Rate is **12%** which means that company has lost 10% customers who used to generate less than 1% revenue.

Insights

- In the monthly analysis, we observed that most of the orders have place in the month of November. The annual growth of Revenue is 6.38%. 5.48% increase in the orders has been observed.
- In the weekly analysis of orders and revenue, pattern has been observed that mostly orders have places on Thursday followed by Tuesday and revenue generated in the same pattern.
- In the hourly analysis of orders and revenue, it has been observed a pattern that 40% orders are placed in between 12pm to 6pm followed by 3pm to 5pm and 9am to 11am.
- We have 17% Champions customers who are generating more than 50% revenue followed by **Hibernating-13%** (Almost lost customers) and **lost customers -12%** who are generating less than 1% revenue.
- After **Champions, Loyal Customer** and **Cannot Lost them** segment generating second highest revenue which is **10%** each.
- Churn Rate is 12% of the company.
- average recency of our customer is 117 days which is almost 04 months. Average customer visits our website 3 times to order and average worth of order is $1,848.

4. Act Phase

Recommendations

Based on the generated insights, following recommendations has been given:

- The business team needs to optimize sales so that business revenue is even greater in low revenue generating months. In addition, given the decline in sales that occur every year in December, January and February, the business team needs to develop ideas and offer incentives offer to customer to earn revenues.

- Company should publish their online ads in those times when customer traffic is high, which is Thursday and Wednesday and from 12 pm to 2 pm

- Company should announce incentive offers in low order receiving months (Dec, Jan, and April) to increase their revenue.

- Company should post their online ads by keeping the peak customer traffic time and days in mind to achieve optimum results.

- Top Customers need to be pampered to stay loyal and willing to spend a lot of money to shop, for example by encouraging them to spread the love with reward programs and send personalized product updates. They are early adopters and helps in promoting our brand.

- For customers having potential to become top customers and recent customers, we need to build a healthy relationship to retain them over a long period of time by offering free trials, free shipping, and complementary services.

- It is important to reconnect with the customers who need attention and are about to sleep by offering renewals and win back campaigns, give recommendations based on past purchases. We can ask for feedback to get insights about their pain points.

- For customers who are almost lost and churn, we can recreate brand value and offer relevant product suggestions or asking for feedback to make relevant improvisations in our process.
