This data analysis report explores customer segments and purchasing behaviors at Mock Company, using data from their CRM and E-commerce systems. By combining these datasets, we uncover valuable insights into customer demographics, geographic distribution, revenue share, and income analysis. With the power of Power BI, we create a comprehensive dashboard that meets the requirements of the marketing stakeholders, empowering them to make data-driven decisions for their targeted campaign.

Data Analyst Challenge

Mock Company

Ifeanyi Udeze

Table of Contents

Introduction
This data analysis exercise focuses on providing insights into different customer segments at Mock Company, utilizing data from their CRM and E-commerce systems. The goal is to create a comprehensive dashboard that meets the requirements of the marketing stakeholders and supports their targeted campaign.

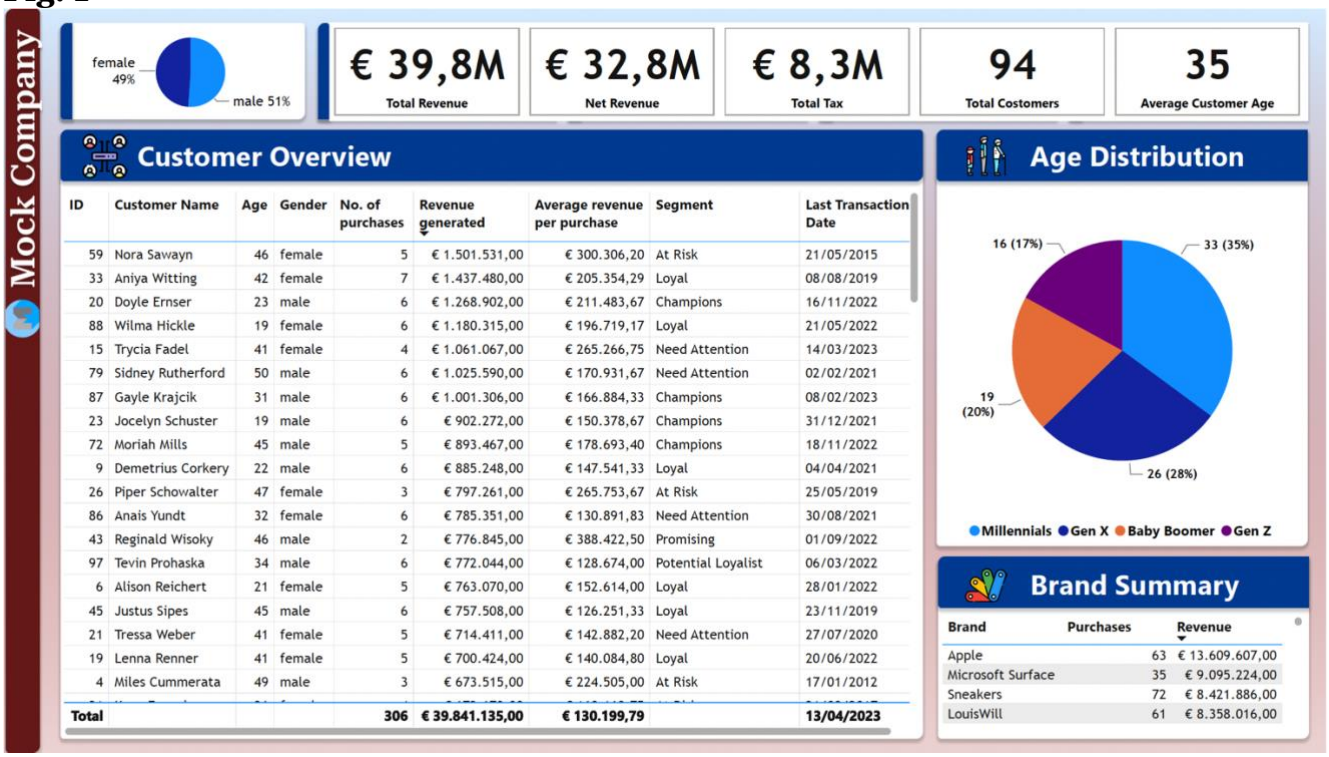
Data Sources and Methodology
Mock Company maintains customer master data in their CRM system, stored in the file customers.csv. The E-commerce system contains information about all the purchases, available in the file purchases.csv. By combining these datasets, we can gain valuable insights into customer behavior and characteristics. The data source is deemed reliable, original, comprehensive, current, and cited.
We will use Power BI for our data cleaning, data modeling, and data analysis. All observations and recommendations will be included in this report.

Data Exploration and Descriptive Analysis
In the data cleaning and modeling phase of the project we analyzed each column for completeness and errors: duplicates, null and incorrect data types.
The customers and purchases table can be joined on the **purchases.customer_username** and **customers.username** columns as the primary and foreign keys. This is done by creating a relationship in the model view of Power BI.
The **customers.username** column contains duplicate entries for “null” and “witting”. We have deleted both “null” values as the primary key column should be unique. We have deleted on instance of “witting” as we observed all other attributes of the row contained the exact values except the gender column.
The **purchases.cart** column is a JSON array column for all the products within a specific purchase. We extracted the array into a new table called cart and created a relationship with the purchases table with purchase.id column. The cart table contains details about each purchase instance. Each cart can contain multiple products within one purchase.

Findings and Insights
Customer Overview
Requirements 1 – A tabular overview of the customers with basic information about them along with the number of purchases they have placed, the revenue they have generated, and the average revenue per purchase
Output 1 – The dashboard contains all basic information about customers. We have segmented the customers into generations to get any insights from their spending habits.

Generation	Start Date	End Date
Baby Boomer Generation	1946-01-01	1964-12-31
Generation X (Baby Bust)	1965-01-01	1979-12-31
Generation Y (Millennials)	1980-01-01	1994-12-31
Generation Z	1995-01-01	2010-12-31
Generation Alpha	2011-01-01	Present

Fig. 1



The dashboard (**Fig. 1**) includes other information such as:
Total Revenue, Net Revenue, Total Tax, Total number of customers, Average Customer Age, and a Brand Summary which shows summarized information about all brands.

Geographic Distribution

Requirement 2 – We would like to know which departments our customers work in, and how they are distributed geographically within the US.

a. Create a stacked bar chart showing which departments customers are working in, and color coded by their location.

b. For more details on the above visualization, please build a matrix that shows the monthly income we make from different departments with the possibility to expand on the locations, and the individual customers.

Output 2a – The dashboard contains a stacked bar chart titled "Customer Dept. and Geographic Distribution" (**Fig 2**). From the visual, we can see Business Development has the most customers, followed by Legal and Support. The dashboard also contains "Purchases by Dept." which lists the number of purchases by department.

Output 2b – The "Monthly Income Breakdown by Dept., Location, & Customer" visual is a matrix showing the monthly income by department. This can be drilled down by location then to individual customers.

These visuals can be sliced to see the performance by Quarter.

Revenue Share

Requirement 3 - We would like to see a pie chart of the income per customer city showing the percentage each city is responsible for.

Output 3 – **Fig 3** "Revenue Share by Customer City" shows income per customer city via a pie chart with the data labels showing the city, percentage share, and total revenue for that city.

Customers from Nashville contribute 10.8% of the total revenue with the top 3 cities contributing 24.4% of the total revenue for the period.

A customer state visual is also included to show revenue by state and can be sliced by month of the year.

Fig 2

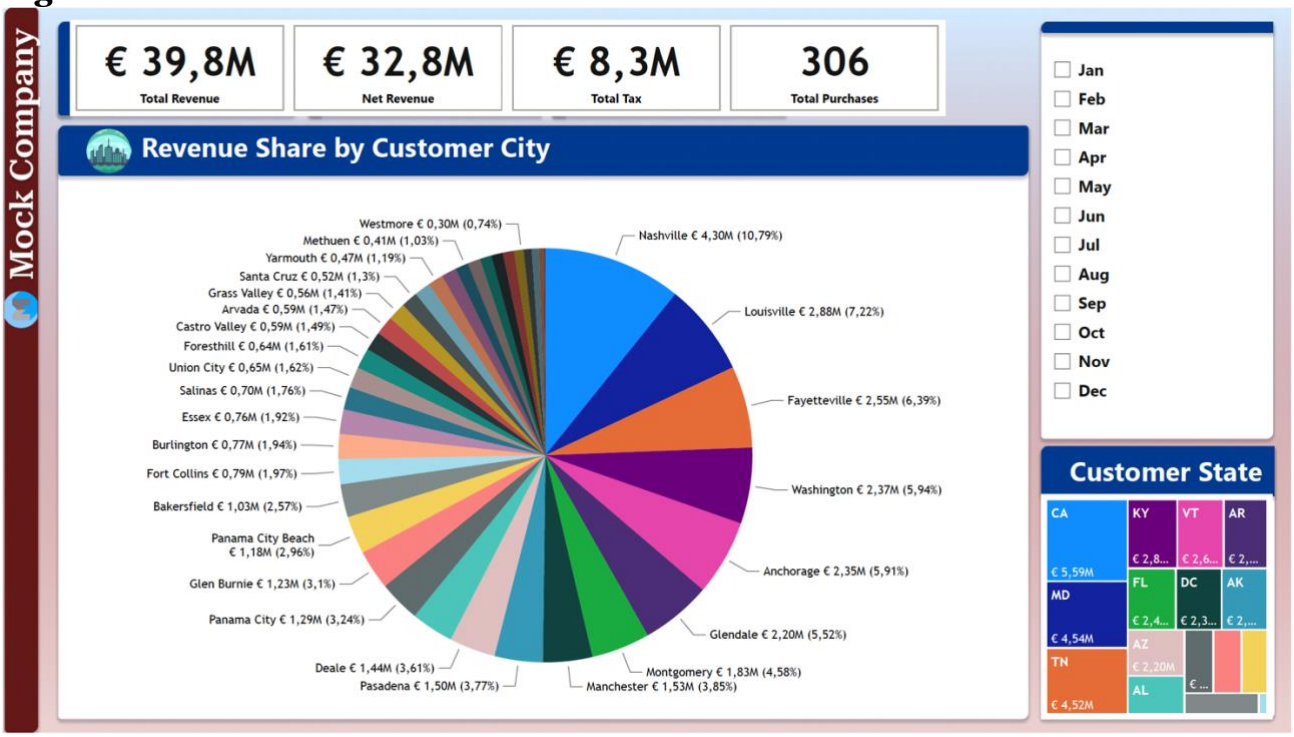
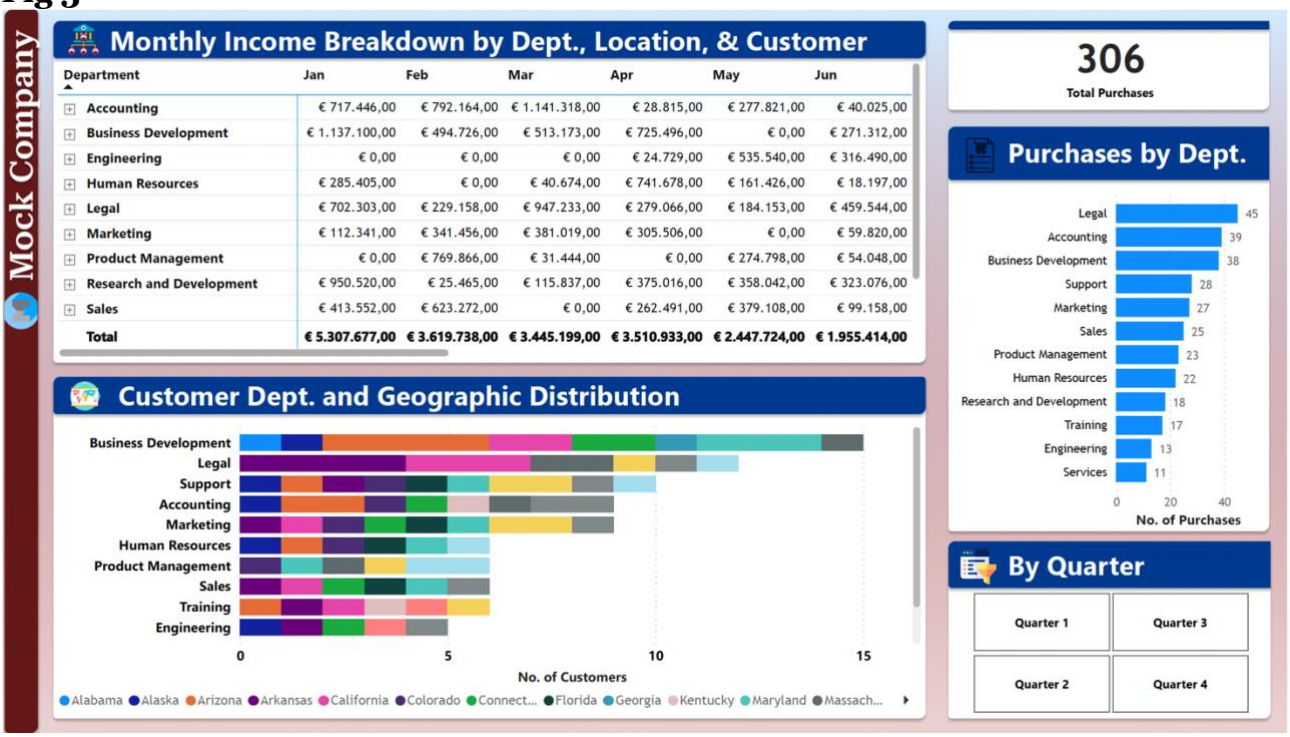


Fig 3



Income Analysis

Requirement 4 - We would like to see the cumulative income for the last 3 years to determine if there is a month where customers purchase more/less than other months. We would like to be able to slice this by department and location.

Requirement 6 - We are also considering if we should rather be segmenting by what the customers buy. Could you provide the top 3 product categories based on income?
Output 4: **Fig. 4a** "Monthly Cumulative Income Analysis" is a bar chart showing the cumulative income per month for the period. We have included slicers to show the result for different company states, company departments, and time periods (Last 1 year, Last 2 years, and Last 3 years). The dashboard also contains the distribution of purchases by day in the month to see the relationship between number of purchases and time of the month.

Fig. 4a – All visuals filtered to "Last 3 years"

Output 6: **Fig. 4b** "Top 3 Product Categories" shows the top 3 product categories and the Net Revenue. This visual is responsive according to the slicer option selected. The top 3 product categories for the period are laptops, motorcycle and smartphones.

Fig. 4b

Top 3 Product Categories

Category	Net Revenue
laptops	€ 14,533,260.00
motorcycle	€ 6,905,009.00
smartphones	€ 3,987,845.00
Total	**€ 25,426,114.00**

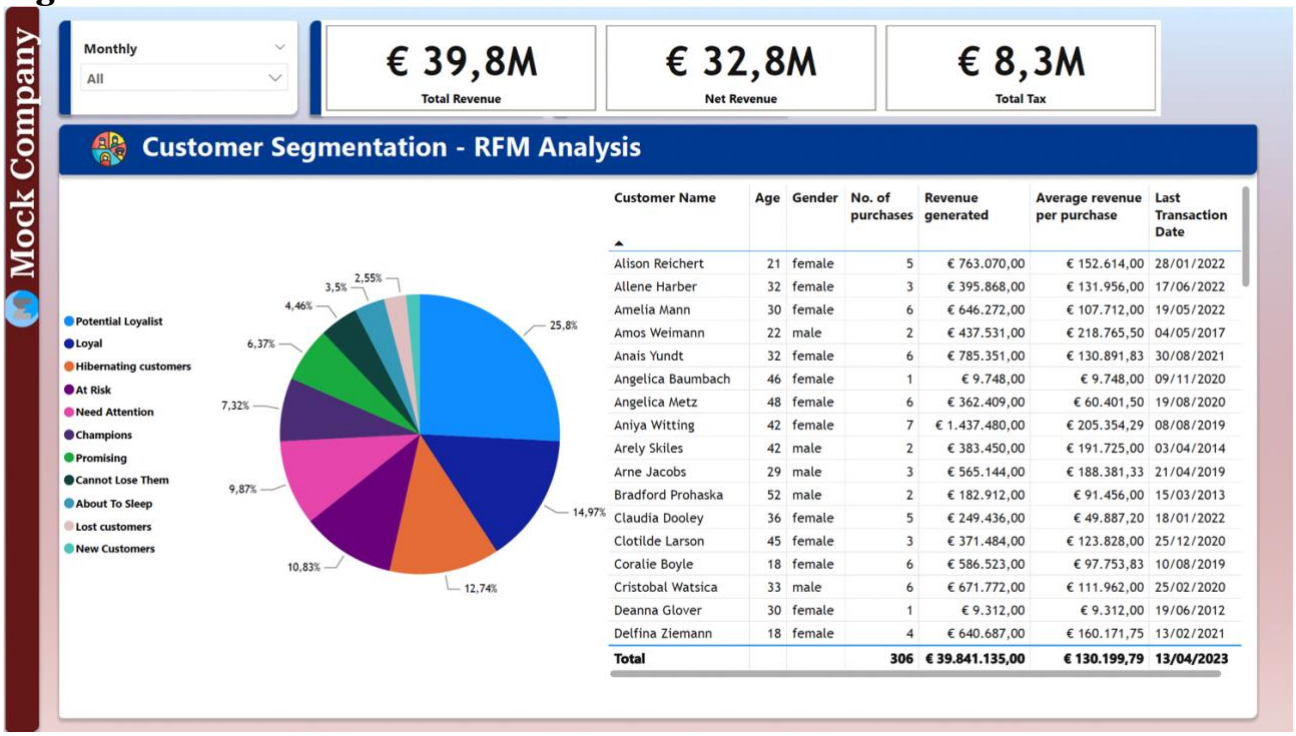
Customer Segmentation – RFM Analysis (Recency, Frequency, Monetary)
Requirement 6: Are there any other customer characteristics in the data that could help us segment the customers?

Output 6: RFM segmentation (Fig. 6)categorizes customers into different segments, according to their product purchases over a period which will allow us to subsequently approach these groups in the most effective way.

Segment	**Activity – 1 year period**
Champions	Bought recently, order often and spend the most.
Loyal	Orders regularly. Responsive to promotions.
Potential Loyalists	Recent customers who spent good amounts.
New Customers	Bought most recently.
Promising	Potential loyalist a few months ago. Spends frequently and a good amount. But the last purchase was several weeks ago.
Need attention	Core customers whose last purchase happened more than one month ago.
About to sleep	Made their last purchase a long time ago but in the last 4 weeks either visited the site or opened an email.
Cannot Lose Them	Made the largest orders, and often. But haven't returned for a long time.
At Risk	Similar to 'Cannot Lose Them' but with smaller monetary and frequency value.
Hibernating customers	Customers who made smaller and infrequent purchases before but haven't purchased anything in a long time.
Lost	Made last purchase long time ago and didn't engage at all in the last 4 weeks.

Other ways which we can segment customers which have been included in the visual and report:
Age Distribution - by Generation (**Fig. 1**), **Brand Summary** - by Brand Purchases (fig. 1), **gender** - by gender (**Fig. 1**), **Purchases by Department** (**Fig. 3**),

Fig. 6



Customer Name	Age	Gender	No. of purchases	Revenue generated	Average revenue per purchase	Last Transaction Date
Alison Reichert	21	female	5	€ 763.070,00	€ 152.614,00	28/01/2022
Allene Harber	32	female	3	€ 395.868,00	€ 131.956,00	17/06/2022
Amelia Mann	30	female	6	€ 646.272,00	€ 107.712,00	19/05/2022
Amos Weimann	22	male	2	€ 437.531,00	€ 218.765,50	04/05/2017
Anais Yundt	32	female	6	€ 785.351,00	€ 130.891,83	30/08/2021
Angelica Baumbach	46	female	1	€ 9.748,00	€ 9.748,00	09/11/2020
Angelica Metz	48	female	6	€ 362.409,00	€ 60.401,50	19/08/2020
Aniya Witting	42	female	7	€ 1.437.480,00	€ 205.354,29	08/08/2019
Arely Skiles	42	male	2	€ 383.450,00	€ 191.725,00	03/04/2014
Arne Jacobs	29	male	3	€ 565.144,00	€ 188.381,33	21/04/2019
Bradford Prohaska	52	male	2	€ 182.912,00	€ 91.456,00	15/03/2013
Claudia Dooley	36	female	5	€ 249.436,00	€ 49.887,20	18/01/2022
Clotilde Larson	45	female	3	€ 371.484,00	€ 123.828,00	25/12/2020
Coralie Boyle	18	female	6	€ 586.523,00	€ 97.753,83	10/08/2019
Cristobal Watsica	33	male	6	€ 671.772,00	€ 111.962,00	25/02/2020
Deanna Glover	30	female	1	€ 9.312,00	€ 9.312,00	19/06/2012
Delfina Ziemann	18	female	4	€ 640.687,00	€ 160.171,75	13/02/2021
Total			**306**	**€ 39.841.135,00**	**€ 130.199,79**	**13/04/2023**

Recommendations

Based on the analysis and insights:
1. Data validation rules should be implemented for each unique identifier (username, id) to check for duplicate entries before new data is added to the database.
2. Millennials and Gen X should be targeted with marketing campaigns and special offers as they contribute 63% of their Total Revenue.
3. Customer Segmentation visual should be utilized in targeting customers over short periods to avoid losing them.
4. Targeted marketing campaigns, local advertising and expansion opportunities should be explored in Nashville, Louisville, and Fayetteville as these locations make up 25% of the company's Total Revenue
5. Stakeholders can use the purchases by day visual to inform sales analysis, campaign timing, inventory management and customer offer personalization to best improve sales.
6. Using the top 3 categories should be used to consider:
Collaborations and Partnerships: Knowing the top categories of products can open up opportunities for collaborations and partnerships with relevant brands or suppliers. The marketing department can explore partnerships to offer exclusive deals, bundled promotions, or joint marketing campaigns that align with these popular categories. This can expand the customer base, increase brand visibility, and drive additional sales.
Targeted Marketing Campaigns: Knowing the top categories of products that customers purchase allows the marketing department to create targeted marketing campaigns. They can develop specific promotions, discounts, or content that highlight products from these categories to attract existing customers and encourage repeat purchases. By focusing on these popular categories, they can effectively engage customers and increase sales.
Cross-selling and Upselling Opportunities: Understanding the top categories of products purchased by customers enables the marketing department to identify cross-selling and upselling opportunities. They can recommend related or complementary products from these categories to customers, either through personalized recommendations or targeted promotional campaigns. This strategy can help increase the average order value and maximize revenue per customer.

Conclusion

In conclusion, the data analysis exercise at Mock Company provided valuable insights into different customer segments and their purchasing behaviors. By combining data from the CRM and E-commerce systems, we were able to create a comprehensive dashboard that addressed the requirements of the marketing stakeholders.

The key findings and insights from the analysis include:

1. Customer Overview: We gained a comprehensive understanding of the customers by analyzing their basic information, number of purchases, revenue generated, and average revenue per purchase. We also segmented customers into different generations to identify any spending patterns or preferences.
2. Geographic Distribution: The analysis revealed the departments in which customers work and their distribution across different locations in the US. Visualizations like the stacked bar chart and the matrix provided detailed insights into the monthly income generated from different departments and their corresponding locations.
3. Revenue Share: The pie chart showcasing the income per customer city highlighted the percentage contribution of each city to the overall revenue. This information can be utilized to prioritize marketing efforts and allocate resources to cities with a significant revenue share.

4. Income Analysis: The analysis of cumulative income over the last three years helped identify months where customers made more or fewer purchases. This information, when sliced by department and location, provides deeper insights into purchasing trends and can inform marketing strategies.
5. Customer Segmentation: The RFM analysis allowed for the segmentation of customers based on their recency, frequency, and monetary value. This segmentation enables targeted marketing campaigns and personalized approaches to different customer groups, improving customer engagement and loyalty.

Based on the findings and insights, we recommend the following actions:

1. Implement data validation rules to ensure data accuracy and integrity, particularly for unique identifiers.
2. Target marketing campaigns and special offers towards Millennials and Generation X, as they contribute significantly to the company's revenue.
3. Utilize the RFM segmentation to target customers over shorter periods and prevent customer churn.
4. Focus on targeted marketing campaigns, local advertising, and expansion opportunities in cities such as Nashville, Louisville, and Fayetteville, which contribute a significant portion of the company's revenue.
5. Leverage the insights from the purchases by day analysis to optimize sales analysis, campaign timing, inventory management, and personalized customer offers.
6. Explore collaborations and partnerships with relevant brands or suppliers based on the top three product categories, enabling exclusive deals, bundled promotions, and joint marketing campaigns.

Overall, this data analysis exercise has provided valuable insights into customer segments, purchasing patterns, and revenue distribution. By leveraging these insights, the marketing department can make data-driven decisions, optimize their campaigns, and drive business growth.

Resources

RFM Segmentation
US States