Report for Marketing Customer Analysis

Author: Alperen Demirci - Mail: alperendemirci65@gmail.com

Problem Definition

Introduction

In the world of marketing, understanding customers is key to success. With so much data available, businesses can use techniques like customer clustering to group customers based on similarities. This helps in creating targeted marketing strategies that resonate with different customer groups.

What is our problem ?

The problem we're tackling involves using customer data from various marketing channels to group customers into meaningful clusters. The main aim is to understand what makes each group unique and how we can tailor our marketing efforts to meet their specific needs.

Our goals are:

1. **Clustering:** Divide customers into groups based on things like age, buying habits, and preferences.
2. **Strategy:** Develop marketing strategies tailored to each group to improve engagement and sales.
3. **Evaluation:** Keep track of how well our strategies are working and adjust them as needed to keep customers happy and coming back for more.

Why are we doing this ?

We're sorting our customers into **groups** to make our marketing better. By knowing what different types of customers like, we can make ads and products they'll **really want**. This helps us sell more and stay ahead of other businesses. Sorting customers also helps us see trends and make smart choices using data. In short, it's about giving customers what they want and making our business grow. Here are some positive sides:

1. **Personalized Marketing:** Customer clustering allows us to tailor our marketing messages and products to match the specific preferences of different customer groups.

2. **Increased Sales:** By delivering targeted offerings, we can increase the likelihood of customer engagement and purchase, ultimately driving higher sales volumes.

3. **Data-Driven Decision Making:** Customer clustering helps us identify patterns and trends in our data, empowering us to make informed decisions that align with customer preferences and market dynamics.

4. **Enhanced Customer Relationships:** By catering to individual customer preferences, we can build stronger relationships and foster customer loyalty over time.

In summary, customer clustering supports personalized marketing efforts, drives sales growth, fosters competitive advantage, facilitates data-driven decision-making, enhances customer relationships, and promotes adaptability in the marketplace.

Ok, so have anybody tried this ?

Sure, vast majority of big marketing companies have implemented data driven clustering analysis. The well known example for this is made by Netflix which got its' strength from data driven recommendation system. Let's talk about the companies which have implemented this technology and empowered their systems.These examples show how companies use customer clustering to make customers more interested, keep them coming back, and make more money by giving them personalized experiences that match what they like.

Here are a few examples with numeric data on their result:

1. AMAZON

- Amazon extensively uses customer clustering to personalize recommendations and target marketing campaigns.
- They analyze customer purchase history, browsing behavior, and demographic data to segment customers.
- For instance, Amazon reported that their recommendation engine drives **35% of their sales.**

2. NETFLIX

- Netflix employs sophisticated customer clustering algorithms to personalize content recommendations for its subscribers.
- They analyze viewing history, genre preferences, and user ratings to cluster users into different groups.
- Netflix estimated that their recommendation system saves the company **$1 billion per year** by retaining subscribers.

3. STARBUCKS

- Starbucks uses customer clustering to tailor its loyalty program offerings and promotional campaigns.
- They analyze transaction data, purchase frequency, and customer preferences to segment their loyalty program members.
- Starbucks reported that their personalized marketing efforts through the loyalty program contribute to a significant portion of their sales, with **41% of transactions coming from the loyalty** program members.

4. SPOTIFY

- Spotify leverages customer clustering to create personalized playlists and recommend new music to its users.
- They analyze listening history, genre preferences, and user interactions to cluster listeners into different music taste profiles.
- Spotify reported that their personalized playlists and recommendations contribute to increased user engagement, with over **30% of listening time** driven by personalized content.

Now you may think that your company is not big as these examples. However it doesn't matter since this investment on your business won't cost as much as their did. Furthermore, this is a low-budget technique which will help you competing with your competitors . All of these companies were start up once and now look where they are by using 'Customer Analysis with Data Driven Clustering'.

Data Handling

Before we dive into the clustering project, first we need to handle the data. In simpler terms this task is done because of the real world effect. For instance if we are asking a persons' marital status, they can give answers such as **'Married', 'Together' or 'Has Significant Other'** etc. which are all the same. For these kind of situations; before diving into algorithms that groups people, we need to make some Data Science. By science, we mean that we need to clean the data from unreasonable values such as '**Age' being equal to 346**. We need to check the balance between categories in the categorical features because we want our model to be **fair on its grouping**. There are more operations that we can and should do on data, we'll explain them basically in this chapter called 'Data Handling'.

Understanding the data

First we need to **understand our data** which requires us to inspect on attributes of people. When we check the data, one can obtain **4 different type of feature types**. These feature types makes it easier to interpret and understand results. We will try grouping on different combined features with Demographic Features like Demo-Spending, Demo-Campaign etc.

1. Demographic Features

- These features represent the personal information about people such as age, education, marital status etc.

2. Spending Features

- These features explain the shopping habits of the entry. For example amount of money sent on wine, meat, fish, sweet etc.

3. Purchase Features

- You can access the purchasing habits of the customer such as number of web, catalog, shopping purchases.

4. Campaign Features

- From these features, we can extract information about if the marketing on a customer is successful or not. For instance acceptance of marketing 1, marketing 2 etc. .

Cleaning the data

What do we mean by cleaning? Of course not with a vacuum cleaner or something :). By cleaning, we find **missing values** in entries or illogical values in features. Example for this would be if something urgent interrupted the retrieval of the data (like a kid crying while his dad filling out the data), then the entry is not healthy. Another example is that a customer entering his date of birth as 1000 BC.

As you may understand, we are doing this because of **real data errors** due to retrieval of the data. We handle the **extreme or unnecessary values** by dropping or transforming them to a useful values using special functions. For missing values, sometimes we drop them but it loses us information so we usually impute them with synthetic data using prediction techniques like regression.

Playing with attributes

Sometimes we can mine more information using more than one variable. For example if we have the values for Kids at home and Teens at home, then we can create a new feature which is called 'total_children' which could represent the information of both features depending on the situation. This helps us to infer new attributes and lower the complexity of the model.

Since we will use mathematical algorithm for grouping, we need to transform our textual - categorical data into numerical data. In order to achieve this, we check the type of the categorical data. It can be ordered(ordinal) or non-ordered(nominal) like education(ordinal) and marital(nominal). If we do not transform these features we can not use them in our algorithm and that loses information and chance to infer new knowledge.

Standardization

Since our algorithm works with numbers, the quantity of those numbers are really important. We don't want our algorithm to think that the greater the number, greater the effect is. Also we need to **visualize the groups** by plotting them on 2D. For this task we will apply another algorithm called **PCA** which will help us to visualize the groups better. Math behind the PCA forces us to standardize our data which means transforming them into a way that the quantity won't affect as much as the raw version. From these two reasons, we standardize our data using statistical techniques like Robust Scaler which uses IQR.

Exploring the data

This is the last step that we will engage with data itself rather than algorithms. In this part, we check the relations between every feature pair. This helps us to reduce number of features used in algorithm because if we select two features that carry the same information, then we may use only one of them and our algorithm will work more efficient. This step is done by checking the correlation between features which tells us how dependent are these variables.

This is a crucial step for our grouping because we may see some patterns even before the algorithm. These patterns can help us test our groups check if they are partitioned well or not.

Visualizing and simplifying the data

In this step we are using PCA to reduce the number of features that will be used in grouping. That is because we can not visualize more than 3 dimensions and in this data we have over 20 features which is impossible to visualize. PCA calculates planes to visualize data using the combination of these features, that's why we can explain a part of variance using PCA. After we select the first two PCA's ,which will express all of our data, we now visualize the PCA's to see if we have any visual groups.

Clustering - Grouping

This is the final step which was the easiest in my opinion. That's because we have prepared our data to fit well to our algorithm and now the algorithm needs to do its' job - clustering. First we find how many groups can we separate the data best. We find this using the most intuitive algorithm KMeans which uses center approach to group data. If we use Elbow Method while running the KMeans algorithm, we can find the best number of clusters we can separate the data best.

After finding the best number of clusters, we proceed with trying with different algorithms such as Agglomerative Clustering and Gaussian Mixture. These other algorithms work differently due to their mathematical background.

We find the best clustering algorithm using silhouette_score(a metric which express the quality of grouping) on 3 different datasets as Demographic - Campaign, Demographic - Spending, Demographic - Purchase. After that now the time has become for evaluation of the groups by checking their mean values on most correlated features.

Evaluation of groups

Finally, we have to make inference based on the algorithm-found groups. We will mention them on the next chapter more openly.

Results of Clustering

Demographic - Campaign

There exist 3 types of customer groups on when we look from the Campaign point of view. We will explain them all under here. You can see the means of features on Image 3.A.

Cluster 0

- **Middle-Income Segment with Older Customers:** Cluster 0 represents customers with moderate incomes who are older compared to the other clusters.
- **Moderate Campaign Acceptance:** They are receptive to campaigns, falling between Cluster 1 and Cluster 2 in terms of acceptance rates.
- **Highest Number of Children:** This cluster has the highest number of children at home, including both teens and younger children, compared to the other clusters.

Cluster 1

- **High-Income Customer Base:** Cluster 1 consists of customers with higher incomes compared to the other clusters.
- **High Campaign Acceptance:** They tend to accept campaigns more frequently than other clusters, indicating receptiveness to marketing efforts.
- **Middle-Aged Demographic:** Customers in this cluster are typically middle-aged, reflecting a stage of life where they may have higher purchasing power and different consumption priorities.
- **Smallest Family Size:** With the lowest number of children at home compared to other clusters, marketing strategies could focus on luxury or premium products special to smaller households.

Cluster 2

- **Low-Income Segment with Younger Customers:** Cluster 2 comprises customers with lower incomes who are younger compared to other clusters.
- **Low Campaign Acceptance:** Customers in this cluster tend to accept campaigns less frequently than those in other clusters, indicating potential challenges in engaging this demographic.
- **Youngest Age Profile:** They are the youngest cluster, suggesting a focus on products and services appealing to younger demographics.
- **Moderate Family Size:** While not having the highest number of children, this cluster maintains a moderate count, indicating potential family-oriented marketing opportunities.

Sights for Purchasing Team

- **Campaigns 1,4 and 5** are all different campaigns which **target to different groups** are customers. These campaigns can be inspected for creating new campaigns for different kind of groups in the feature.
- Since Low income, young customers have a **lower education**; campaigns can be **simplified** for them to understand better and to make them feel more comfortable.
- Campaign team can focus on creating **kid themed campaigns**. This would result **cluster 0** customers to take their interest since they have the most amount of children among clusters.
- Given their age and family size of **cluster 0**, marketing strategies could emphasize **family-oriented products and services** tailored to their life stage and preferences.

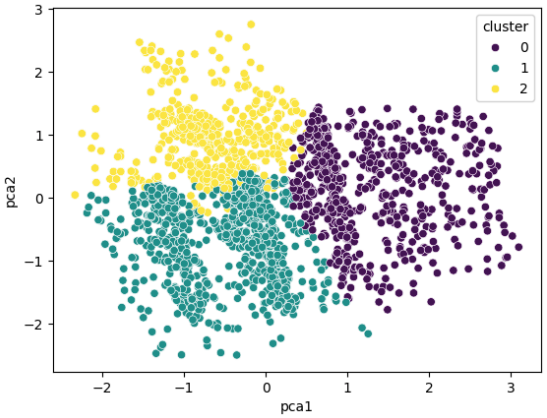
• Here you can see the mean of metrics for every cluster:

Image 3.A: Mean of every feature of entries grouped by their clusters. (KMeans)

cluster	Education	Income	Total_Children	Teenhome	Age	single	Total_Campaign_Acceptance
0	0.875123	-0.054758	0.450836	0.961652	0.295040	0.163225	0.151646
1	0.832248	0.461076	-0.809446	0.141694	0.179967	0.245928	0.991606
2	0.147783	-0.838345	-0.162562	0.113300	-0.439610	0.279146	0.131253

Image 3.B: Mean of every feature (including campaign numbers) of entries grouped by their clusters. (KMeans)

- You can see the customers separated into **3** groups. Each color represents a group (cluster). This graph is obtained from **KMeans** clustering since it has the best silhouette score among other algorithms on this data.



Demographic - Purchase

There exist 3 types of customer groups on when we look from the Purchase point of view. We will explain them all under here. You can see the means of features on Image 4.A.

Cluster 0

- **Lowest Income, Highest Website Visits:** Cluster 0 represents customers with the lowest income levels who visit the website frequently. This suggests that despite limited purchasing power, they actively engage with the brand online, possibly seeking deals or information.
- **Lowest Purchases, Highest Number of Kids:** Their tendency to make fewer purchases despite frequent website visits may indicate budget constraints. The high number of kids suggests family-oriented behavior, potentially influencing their spending priorities and patterns.
- **Youngest Cluster:** Being the youngest cluster, they may represent a younger demographic segment with evolving needs and preferences.

Cluster 1

- **Middle Income, Moderate Behavior:** Cluster 1 comprises customers with moderate income levels and purchasing behavior. They make average purchases and visit the website moderately, indicating moderate engagement with the brand.
- **Middle Number of Kids, Oldest Cluster:** Their average number of kids and older age compared to other clusters suggests a stage in life where family responsibilities may be significant, impacting their spending habits and priorities.

Cluster 2

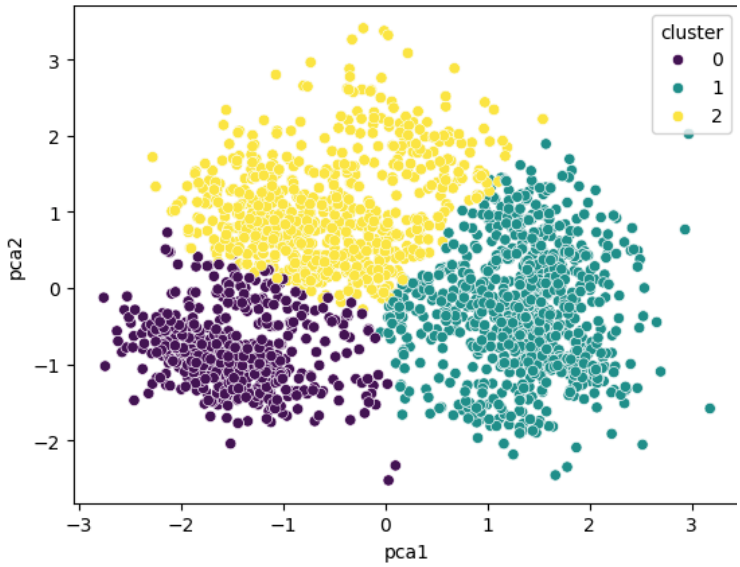
- **High Income, Low Website Visits:** Cluster 2 represents customers with high income levels who make the most purchases but visit the website the least. This suggests a preference for direct purchasing channels or a higher level of brand loyalty.

- **Lowest Number of Kids, Middle-Aged:** Their lower number of kids and middle-aged status may indicate a more financially stable and established segment with different spending priorities, possibly focusing on quality or luxury items.

Image 4.A: Mean of every feature of entries grouped by their clusters. (KMeans)

cluster	Income	Total_Children	Kidhome	Age	NumWebVisitsMonth	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	Total_Purchases
0	-0.764799	0.211974	0.815534	-0.130529	0.131607	-0.668739	-0.473907	-0.398274	-0.587379
1	0.158226	0.260274	0.312329	0.281202	-0.058562	0.367194	0.154381	0.477808	0.420232
2	0.532286	-0.898799	0.018868	0.100438	-0.809177	0.135181	0.517561	0.642196	0.366209

Image 4.B: Mean of every feature (including campaign numbers) of entries grouped by their clusters. (KMeans)



- You can see the customers separated into **3** groups. Each color represents a group (cluster). This graph is obtained from **KMeans** clustering since it has the best silhouette score among other algorithms on this data. Silhouette score for this clustering is **0.47** (range: -1,1)**.**

Comments of the Purchase Features

1. Marketing Focus Areas:

- Marketing efforts should prioritize customers in **Cluster 0 and Cluster 2** due to their potential for increased engagement and purchases.
- **Cluster 2** stands out as a crucial target for marketing initiatives as they represent middle-income customers who exhibit purchasing behaviors similar to those with higher incomes.

2. Cluster-Specific Strategies:

- While **Cluster 1** shows **satisfactory purchasing behavior**, there's still room for improvement. Targeted marketing strategies could further enhance their engagement and loyalty.
- Despite **Cluster 0's frequent website visits**, their low purchase rate suggests **untapped potential**. Implementing special ads or offers tailored to their demographic attributes could help convert visits into purchases effectively.

3. Shopping Preferences:

- Cluster 1 demonstrates a preference for catalog and in-store purchases over web-based shopping. Marketing strategies could focus on **encouraging web purchases** by highlighting **product visibility and convenience**.
- Cluster 2 engages in both web and store purchases, indicating **versatility** in shopping preferences. Marketing efforts could emphasize catalog purchases and offer specialized deals to encourage increased engagement across all platforms.

4. Innovative Marketing Approaches:

- Introducing **unique encouragement** like exclusive web purchase codes usable in stores or online could enhance visibility and increase multi-channel engagement for **Cluster 2** customers.

Demographics - Spending

There exist 4 types of customer groups on when we look from the Spending point of view. We will explain them all under here. You can see the means of features on Image 5.A.

Cluster 0

- **Most Educated, Moderate Spenders:** Cluster 0 is the most educated group but tends to spend the least compared to others. Their spending is evenly distributed across different product categories.
- **Large Families, Less Gold Buying:** They have the highest number of children but don't purchase gold products as much as other clusters.

Cluster 1

- **Oldest with Moderate Spending:** Cluster 1 is the oldest group and spends moderately compared to others. They allocate more spending towards wines and meat products.
- **Moderate Family Size:** They have an average number of children compared to other clusters.

Cluster 2

- **Lowest Income, Minimum Spending:** Cluster 2 has the lowest income and spends the minimum compared to other clusters. They allocate more spending towards fruits and sweet products.
- **Moderate Family Size, Lower Education:** They have an average number of children and are the lowest educated group among the clusters.

Cluster 3

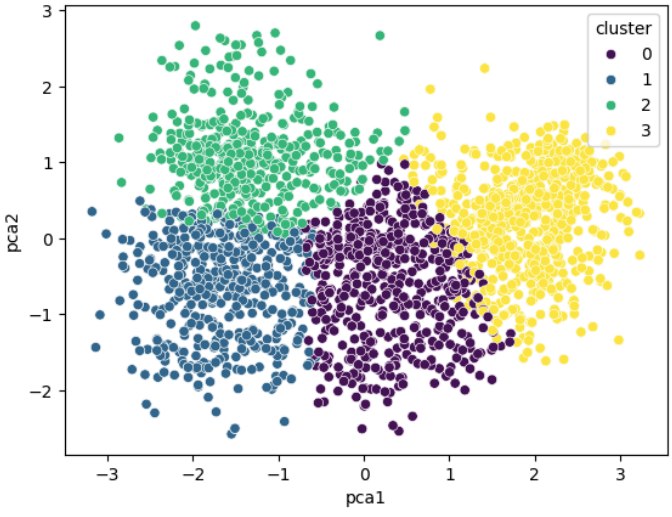
- **Highest Income, Highest Spending:** Cluster 3 is the highest income group and spends the most compared to others. They allocate more spending towards fish, meat, and gold products.
- **Small Families, Middle-Aged:** They have the lowest number of children and are in the middle age group compared to other clusters.

Image 5.A: Mean of every feature of entries grouped by their clusters. (KMeans)

cluster	Education	Age	Income	Total_Children	Kidhome	Teenhome	MntGoldProds	MntWines	MntSweetProducts	MntFishProducts	MntMeatProducts	MntFruits	Total_Spent
0	1.049485	0.212600	-0.426009	0.685567	0.874227	0.857732	-0.670532	-0.534158	-0.449832	-0.561585	-0.587898	-0.460786	-0.307751
1	0.865000	0.316852	0.142762	0.148333	0.246667	0.908333	0.191767	0.192309	0.032859	-0.001266	0.124586	0.047866	0.260334
2	0.161224	-0.435601	-1.026025	-0.159184	0.800000	0.040816	-0.378104	-0.870798	-0.179427	-0.237095	-0.543000	-0.190826	-0.309590
3	0.577444	0.094820	0.512134	-0.723308	0.046617	0.230075	0.500401	0.347378	0.650884	0.710754	0.632942	0.647821	0.941126

Image 5.B: Mean of every feature (including campaign numbers) of entries grouped by their clusters. (KMeans)

- You can see the customers separated into **4** groups. Each color represents a group (cluster). This graph is obtained from **KMeans** clustering since it has the best silhouette score among other algorithms on this data. Silhouette score for this clustering is **0.39** (range: -1,1)**.**



Advices for Business Owner

Marketing Focus Areas:

- Marketing needs to focus on customers in Cluster 0 and Cluster 2. These clusters present significant opportunities for increased engagement and spending.

Advices for Cluster 2:

- Cluster 2 is the oldest customer cluster and they tend to spend the middle amount of money. Marketing can focus on them to increase their spending by adding special ads or offers based on their demographic attributes. For example, they can utilize traditional marketing channels like TV or radio ads. Additionally, emphasizing wines and meat products could capitalize on their spending tendencies in these categories.

Advices for Cluster 1:

- Marketing is already doing well for Cluster 1 but they can still focus on them more. Since they are the highest income cluster, marketing can focus on the products which are expensive and which are not bought by them as much as the other clusters. Their education level is lower than the other clusters so marketing can make ads which will make these customer feel special and unique.

Advices for Cluster 0:

- Cluster 0 is the **most educated cluster** and they tend to spend the least amount of money. Since their education level is high, they might be more open to the ads which are more **detailed and informative**. They have the most amount of kids so marketing can focus on natural and healthy products which have detailed ads for them. By this way, these customers will be more likely to buy products which they think it's good for their kids.

Advices for Cluster 3:

- Cluster 3 is the **lowest educated** cluster and they tend to spend the minimum amount of money. Marketing can concentrate efforts on encouraging increased spending by tailoring special advertisements or offers to match their demographic attributes.For example they can **simplify the ads** for them to understand better and to **make them feel more comfortable**. Also they can add more ads for **fruits and sweet products** since they tend to spend the most amount of money on these products.

Strategies for Marketing Team

We have given some marketing techniques, region of different interests based on the characteristics of clusters. These strategies are split based on the groups they focus.

Strategies for Cluster 0:

1. **Education-Centric Marketing:** Develop marketing campaigns that resonate with educated consumers. Emphasize product features, **benefits**, and **quality** to affect them from their noticeable sides.
2. **Family-Focused Promotions:** Create **family-oriented** promotions and bundles to cater to their larger family sizes. Offer discounts on bulk purchases or family-sized packages.
3. **Gold Product Awareness:** Educate Cluster 0 about the benefits and value of **gold products** through targeted advertising and promotions. Highlight the **long-term investment potential and cultural significance** of gold items.

Strategies for Cluster 1:

1. Wine and Meat Promotions: Launch special promotions and events focused on wine and meat products to capitalize on Cluster 1's preferences. Offer wine tasting events or exclusive meat-based recipes.
2. Senior Discounts and Loyalty Programs: Implement senior discounts and loyalty programs to reward Cluster 1 for their **loyalty and encourage** repeat purchases.
3. Health and Wellness Initiatives: Introduce **health-focused initiatives** and products to cater to Cluster 1's aging demographic. Offer wellness programs, nutritional guides, and senior-friendly products.

Strategies for Cluster 2:

1. Budget-Friendly Options: Provide **budget-friendly** options and promotions to accommodate Cluster 2's lower income levels. Offer value packs, discounts, and affordable product lines.
2. Sweet and Fruit Promotions: Highlight promotions and discounts on **sweet and fruit products** to align with Cluster 2's spending preferences. Create themed promotions around seasonal fruits and sweets.
3. Education and Skill Development: Offer educational resources and workshops tailored to Cluster 2's needs and interests. Provide opportunities for skill development and personal growth at affordable prices.

Strategies for Cluster 3:

1. Premium Product Experiences: Curate **premium** product experiences and services to cater to Cluster 3's higher income levels. Offer exclusive access, personalized services, and luxury product lines.
2. Investment Opportunities: Position **gold products** as **investment opportunities** and **wealth preservation** assets for Cluster 3. Educate them about the long-term value and stability of investing in gold.
3. Convenience and Exclusivity: Provide convenient purchasing options and exclusive opportunities for Cluster 3. Offer **priority access** or maybe **VIP events** to enhance their shopping experience.

Conclusion

In summary, these strategies are like **customized plans** that help businesses connect better with **different types of customers**. By focusing on what each group of customers likes and needs, businesses can make them want to buy more. Paying attention to clusters like Cluster 0 and Cluster 2 and creating **special ads or deals** that match what they like can help businesses **grow**. When businesses understand and **meet the needs** of their customers, they're more likely to keep them happy and coming back. These strategies not only help make customers happy but also help businesses **make more money** and expand in the future. Simply put, these customized strategies are the key to business growth, making customers happy and businesses successful in the long run. Who knows, maybe you will be the next Amazon, Netflix, Starbucks or Spotify!

- Thanks for reading! If you have any advices or warnings about this report, please notify me from my mail.