NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Business Cases with Data Science

Case 3: Recommender System

David Psiuk	number: 20230818
Peter Falterbaum	number: 20230956
Luis Penteado	number: 20230441
Noah Campana	number: 20230996

Group A

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

April, 2024

Table of Contents

1. Executive Summary

This report outlines the development and implementation of a smart recommender system, aimed at increasing the proportion of purchases customers make with Recheio.

Recheio is a leader in the Portuguese wholesale market. The company performed very well in the last year and now want to invest into a more digitized infrastructure. Therefore, the primary objective is to increase customer retention and sales through targeted recommendations to already existing customers, utilizing historical purchase data to enhance the customer experience and influence future buying behaviors.

Several crucial stages were identified during the customer journey on Recheio´s website, where recommendation systems can be strategically deployed on several subpages such as Product Search, Product View and Basket View. We created customized recommendation systems for each of these important touchpoints.

For the Product Search we suggest employing a recency and frequency ranking system that not only highlights relevant products but also adapts to the different client types. In the Basket View, the integration of two distinct recommendation features will be suggested: “Did You Forget?” reminds customers of items they often purchased but haven´t added, while “Try Something New” suggests new products based on similar customers. For the Product View, the suggested recommender system displays similar products.

An evaluation process was designed to assess the performance of different similarity measures used in the “Did You Forget?” recommendations, identifying the most effective metric for real-life application. Additionally, the recommendation systems were optimized for computational efficiency by pre-calculating general similarity matrices and storing them in files, we ensure that recommendations are generated rapidly, enhancing the user experience without compromising response times and reducing needed computing capacity.

To accurately evaluate the impact of these recommendations, it is essential to conduct a real-world assessment of consumer behavior. By conducting A/B testing with different recommendation combinations, we can improve placements, enabling us to identify and deploy the most effective sales strategies by evaluating performance metrics.

Our results indicate a significant potential to improve customer satisfaction by effectively aligning product suggestions with consumer needs. This alignment is expected to significantly boost Recheio's sales, demonstrating the value of the proposed data-driven methods.

2. Business Needs and Required Outcome

The retail sector faces an exceptionally high elasticity of consumer demand. (Casado & Ferrer, 2016) Therefore, Recheio faces great pressure to offer competitive prices to its customers while maintaining the company's profitability. This situation requires an appropriate market positioning, combined with a profound pricing strategy.

According to Porter's four generic strategies we can divide a strategy into two dimensions: competitive advantage and market scope. Competitive advantage can either be achieved through lower costs than its competitors or differentiation. The market scope can be either broad or narrow. An overview of the strategies can be found in Figure 1. (Porter, 1980) In the case of Recheio we can observe a strong tendency to a broad market scope with a cost focus, resulting in a cost leadership strategy. This strategy allows the company to position itself as the price leader in the market, which is in line with customer elasticity.

2.1. Business Objectives

According to Harvard business review it can be anywhere from 5 to 25 more expensive to acquire a new customer than retaining an existing one (Gallo, 2014). To keep costs down and still increase sales, the company decided to target the behavior of existing customers rather than marketing to new ones. To keep costs down and still increase sales, the company targets the behavior of existing customers rather than marketing to new ones, i.e. increase the sales of already existing customers. This requires the implementation of an in-depth recommendation system that could adapt to each individual customer and their specific needs. By using these algorithms, the company can achieve the result of personalized product recommendations. By doing so, Recheio can improve customer experience and satisfaction, and ultimately drive sales growth. Thus, we can build loyalty and create space for market expansion while focusing on the already existing customer base.

2.2. Business Success Criteria

As stated before, the main goal of the recommendation system is to increase Recheio's weight in every customer's purchase. Therefore, the purchases made of an individual customer with other food retailers need to decrease, while the share of purchases made with Recheio need to increase. This might seem straightforward, but the growth of the market needs to be taken into account. According to Report Linker's overview of the Portuguese restaurant market, the expected growth rate is 5.2% per year (Report Linker, 2023). This implies that in order to increase Recheio's share of total sales, sales would have to grow by more than 5.2%, otherwise the increase in sales would only reflect the growth of the individual businesses.

2.3. Situation assessment

Recheio Cash and Carry is part of a Jerónimo Martins Group that holds assets in the food sector, mostly in retail, with market leadership positions in Poland and Portugal. This parent holding allows Recheio to do larger investments and represents a more comfortable situation in terms of company liquidity.

Recheio cash and carry, with a network of over 600 stores and ongoing expansion efforts, navigated a challenging market in 2023, marked by decelerating food inflation and weak consumption in Portugal. However, despite these challenges, the company capitalized on increased tourism, particularly in the first half of the year, driving notable growth in the cash & carry segment. This growth was underscored by a successful marketing campaign focusing on traditional local food, resulting in an impressive 18% increase in sales within the HoReCa segment, leveraging regional product offerings to meet local customer preferences. (Jerónimo Martins, SGPS, S.A., 2023)

In 2024, Recheio plans to continue refurbishing stores to elevate the customer experience, ensuring a modern and inviting shopping environment. The company remains committed to prioritizing food services as a key growth area, catering to the evolving needs of its customers. Additionally, Recheio will continue its digital transformation journey, leveraging technology to enhance operational efficiency and customer engagement using a recommender system (Jerónimo Martins, SGPS, S.A., 2023). Currently the recommendations are provided by Einstein from Salesforce. The company and its customers are not quite satisfied with the recommendation system as for example pet food is being shown to non-pet owners, which does not meet the company's requirements of recommending relevant products.

Within this context there are several ways to apply more personalized recommendations. By mapping the customer journey in the e-commerce tool, it is possible to find some opportunities to meet the customer’s needs. Potential customers need to register on the website to make their first order and see product prices as well as personalized recommendations. In the registration users provide some characteristics such as specifying a client type and their address. Several sections were identified on the website.

1. Landing page: This is the first page a customer sees when they visit the website. It displays various product categories for customers to browse and select products.
2. Basket view: This checkout page shows a list of products the customer has selected. Customers can return to the product list or search to add more items.
3. Product search: Here, customers can type the name of a product they're looking for into a search box, which then shows a list of similar products.
4. Product view: On the product details page, customers can view essential information like the brand and type and select the quantity they want to add to their basket.

2.4. Determine Data Mining Goals

In this project the main goal is to find similarities between customer, purchases and products to give better product recommendations for (potential) customers visiting Recheio's website. By analyzing historical purchase data, we aim to find patterns to build customer-based and product-based systems. This analysis involves comparing several similarity measures, well known market concepts, such as "Did You Forget" and "Smart Basket Analysis".

However, the success criteria extend beyond finding the patterns, but find relevant information that can give us better recommendations which will in fact change the customer behavior by purchasing more within the company. The data mining goals can be categorized in two approaches which shall lead to increasing Recheio's ratio in the total of the customers' purchases:

Increase the average number of items per purchase: an increase in the basket size is an indicator that the customer tends to be more satisfied with the available products and with the recommendations. This is an opportunity to meet customer needs without requiring the purchase of additional products from competitors.

Increase the frequency of purchases: Holding the basket size fixed, increasing the frequency of purchases allows us to meet customers' needs more frequently, thereby reducing the likelihood of them purchasing from competitors, while increasing the products they buy with Recheio.

In conclusion, by leveraging advanced data mining techniques and targeted customer engagement strategies, this project aims to enhance product recommendations on Recheio's website. The ultimate goal is to increase Recheio's share of total purchases for each client, thereby encouraging more frequent and larger transactions exclusively within the company.

3. Methodology

3.1. Data understanding

We received four different datasets, namely 'customer types', 'customers', 'products' and 'transactions'. The customer types dataset simply assigns a unique identifier to each customer type. The clients dataset describes the postcode of the client and assigns the client type identifier to each client. The dataset contains transactions of 930 customers.

The products dataset contains a detailed description of each product, which includes commonly the name of the brand, the name of the product family and sometimes its volume. It also contains distinct product category, separating products from another. Nearly 2,500 products are listed in it.

Finally, the transactions dataset combines the customer ID with the product ID and the date on which the transaction took place, naturally this is the largest dataset as it exhibits every single transaction between 01.03.2019 to 31.05.2019, containing more than 230,000 transactions. A transaction captures one sold product per day and customer, quantities are not given. The transactions have been fulfilled in a total of 17,396 purchases. A short overview of the dataset and the proportion of products can be seen in Table 1. Despite most dimensions being numerical most of them are categorical, except for the date column which is continuous.

3.1.1. Product Categories

A first glance at the product categories being sold, illustrates that “Alimentação Corrente” and “Frutas e Vegetais” are the most sold in the respective order. These two categories are each twice as popular as the third. Figure 2 shows that, there are approximately ten categories which are sold very rarely. In addition, there are a significant number of categories, some of which could be combined into a single, more generalized category. For example, “Vinho”, “Agua”, “Refrigerantes” and “Cerveja” could be combined into a single category, “Bebidas”. In order to facilitate more insightful analysis and straightforward interpretation, we consolidated some existing categories, resulting in 13 distinct product categories, 20 fewer than the original categorization. The final categorization can be seen in Figure 3. The initial top two categories remain the most sold, while the lower end is showing fewer small values, making the graphic easier to comprehend and quicker to analyze.

3.1.2. Client Analysis

This report presents an analysis of a dataset comprising 28 client types, a total of 930 customers, providing insights into customer behavior and preferences.

The five predominant client types by frequency are “Cozinha Portuguesa”, “Cozinha Japonesa”, “Cozinha Italiana”, “Rodizio de Carne” and “Bares/Snacks/Cafés/Gelaterias/Brunch”. Among these, 'Cozinha Portuguesa' emerges as the most frequent, with approximately 70,000 purchases, indicating a strong preference among the clientele.

When analyzing the intersection of client types with product categories, two categories stand out across the board: “Alimentação Corrente” and “Frutas e Vegetais”. These categories align with the overall purchase distribution, suggesting that these are universally preferred among various client types. A unique observation is that for “Cozinha Japonesa” client type “Frutas e Vegetais” are the most purchased category, almost doubling the sales of “Alimentação Corrente”. That shows a divergent consumer preference from the other client types.

When considering the broader customer base, which includes 930 customers not segmented by client type, a diverse transaction becomes evident. A set of clients has a higher number of transactions. This discrepancy between the mean (around 250 purchases) and median (156 purchases) suggests a skewed distribution, where a smaller number of customers have a very high number of transactions, potentially highlighting them as potential strategy to the business.

3.1.3. Product-Client Match

It can be observed that the preferences and most popular items vary across different client types. As previously stated, the three most purchasing client types are Portuguese, Japanese and Italian cuisine, respectively. Traditional Portuguese clients tend to purchase more basic groceries, such as milk and oil. In contrast, Japanese clients tend to purchase highly specific products, such as mango, cucumber or cream cheese. Italian customers' most popular products are surprisingly tuna, pineapple and cream.

Although the different client types show differentiation in their purchasing patterns, it is evident that the main ingredients of different cuisines are likely to be purchased elsewhere. For instance, salmon is not a top product for Japanese clients, while pasta and flour are not included in the top products for Italian clients. This indicates that there is significant potential for Recheio to gain a larger share of client purchases.

3.2. Data preparation

The analysis of the provided data showed no missing values or inconsistencies which require modification before being used for the further modelling and analysis. Customers which do not belong to the area of Lisbon have been identified as outliers, leading to a removal of 26 customers out of total 930. Originally four datasets have been provided to the team. Primary keys of the datasets were used to combine all of the information and were stored for further use in this project.

3.3. Modeling

For recommender systems in environments such as websites, speed is a crucial limitation as the user experience shall be unhindered by recommendation rankings done in the background. Therefore, this project uses pre-calculation steps to prepare the live recommendations and only simple calculations in real time. As such, mainly similarity measures are implemented to get recommendations based on fitness to either customer or product attributes.

The recommendations proposed in the following are designed to be implemented in specific stages of the customer journey. This is labeled as the *Purpose* of each recommender system. The area the recommender is pointing to is described with the *Data Basis* and its algorithm is laid out in the *Functionality* part. With the composition of all recommenders the customer journey will be enhanced by customized recommendations of different scopes.

3.3.1. Recency and Frequency (RF) – Recommendations

Purpose: When users enter the landing page, they need to select a product domain. Within this product domain the products shall be listed according to their best match. The best match is assessed in this project by the recency and frequency of sold products, ranking frequent and recent bought products higher. If the user is logged in, the client type shall also be considered for the ranking.

Data Basis: The recommender is based on all purchases stored in the records. Because its purpose is addressing the recency as a time critical variable, the dataset should be maintained updated to accurately contain recent purchases and trends. Additionally, the frequencies are updated with this – although considering it less time critical.

Functionality: This recommender is in line with an industry-wide clustering technique called RFM, which stands for Recency, Frequency and Monetary (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011). As we do not have access to the monetary value of transactions, we focused on recency and frequency. Recency is calculated for each product and indicates when it was last sold to a customer. The recency flag divides the products into four equal-sized bins. Frequency counts the number of times each product appears in all transactions. These frequencies are then stored and indexed, with the least bought product having an index of 0 and the most sold product having the highest index.

To generate a ranking, we decided against the traditional concatenation of recency and frequency values, e.g. recency = 4 and frequency = 4 → RF score =44. Instead, we multiplied the recency value by the rank of the frequency, e.g. recency = 4 and frequency = 4 → RF score = 16. Having the ranking for recency and frequency separately ordered the product of both inherits the same order combined. This ensures a continuous ranking rather than clusters that are not ranked internally, and products that are just on the edge of a recency cluster can be ranked higher because they may be bought more often.

The described steps deliver the recommendations for users that are not logged in. For customers that are logged in, the client type is additionally taken into account as a factor. In this case, the RF ranking is calculated again for all products in the domain bought by the client type. This results in a subset of the product domain, having the recency, frequency and RF score calculated on that basis. Both rankings are then weighted and summed up. In our implementation we set the weight of the client type ranking to 20% and the ranking of all products to 80%. Further analysis and assessment of customer expectations should tune this weighting to create a suitable customization of the recommendations.

3.3.2. Did You Forget? – Recommendations

Purpose: The recommendations for the "Did You Forget?" problem are designed to be displayed in the customer´s basket view and are based on the client's historical data. We hereby want to offer personalized recommendations for the customer and increase sales such as customer satisfaction. By displaying commonly historical purchased products we also want to facilitate the client's decision-making process. The main criterion for the recommendation ranking is the fit of previous purchased products to items of the current basket.

Data Basis: The analysis is solely based on the historical purchase data of a single customer. To minimize the computational demand, the dimensional space of the dataset is limited from the beginning on to only contain products bought in the past – reducing the number of potential products from the initial volume of nearly 2,500 products.

Functionality:
The recommendations are generated by calculating distances between products in the customer´s historical transactional data. Several distance metrics have been tested for this approach leading to the selection of "dice" as the best performing metric. This choice will be further elaborated in the Evaluation chapter. The Dice algorithm calculates the similarity by taking twice the number of elements common to both sets, divided by the sum of the number of elements in each set, with the Dice distance being 1 minus this similarity score. The calculated distances will be transformed and ordered into a similarity matrix which represent the ranked similarity between products bought by the selected customer, considering each purchase of the history. In other words, previous purchases are compared to each to find products commonly bought together.

For the products in the current basket, the most similar products get identified by examining the similarity matrix. Products that are already in the basket are excluded from their own similarity calculations to prevent redundancy in the product selection.

We suggest recommending the best 5 products, ranked by the similarity score for each basket item. These recommendations are tailored for potentially forgotten items that complement the customer´s current basket selection.

3.3.3. Try Something New – Recommendations

Purpose: This recommender system for the "Try Something New" problem aims at recommending products that the customer might like but doesn´t usually buy. Therefore, similar customers and their purchase behavior are in scope of the analysis. This will specifically be implemented in the basket view.

Data Basis: The system is based on customer similarity and not the client's historical data. The similarity of clients is expressed through the comparison of their previous purchases. With this the similarity of clients is derived by the similarity of their purchases. Products bought by similar clients build the range of possible recommendations.

Functionality:
Firstly, the similarity of all customers to each other are calculated and represented in a similarity matrix. Calculating the similarities of customers with a cosine distance matrix allows to identify and categorize customers based on their level of similar behavior. Using cosine allows to get a scoring of similar behavior of customers rather than a ranking of distances between the individual purchases. This matrix is then transformed and ordered into a ranked similarity matrix to reflect similar customers more accurately. We advise on doing this step separately from the individual recommendation and store or better cache it for use in live environments. For example, it can be calculated once a day and stored in a file which will be easily accessed when calculating the "Try Something New" recommendations.

By utilizing this approach, the company saves significant computing time and avoids the need to calculate the similarity of customers on an ongoing basis. This is crucial, as the customer base and actual transaction history for Recheio is even bigger compared to the used database for this report. In addition, only a small ratio of the data will be updated – implying a regular, well thought calculation update.

The next step is calculated customer specific and therefore required to be done live. For the selected customer, the most similar clients are identified using the given customer similarity matrix. The number of similar considered clients can be dynamically adapted in the implementation – a number of 7 clients was used. All purchases of these close clients are listed together, building the base for the recommendations.

In a next stage the existing system developed for the "Did You Forget" problem is employed, as it uses the same purchase-based approach, required in this phase. The functionality is used this time to get recommendations matching the closest customers and not the current customer´s history. Subsequently the products that are most similar to the current basket of the selected customer will be returned in terms of the purchase history of the closest clients.

Furthermore, only products that have not yet been purchased by the selected customer will be recommended to ensure the introduction of new products.

Several fallback methods are included, for instance, in the event of the customer's first purchase, the recommender system will return 'None' and indicate that the recommendation of new products is not possible. Additionally, deploying this recommender system in praxis allows for the recommendation of alternative products in case the initially selected items are not currently available. This can be done using filters and will be elaborated in more detail in the deployment chapter.

3.3.4. Similar Products – Recommendations

Purpose: This recommender system should present a list of products for a given product. This shall be used when a customer is looking at a single product (product view). In this use case other products which are commonly bought together with the product in scope shall be presented to the customer.

Data Basis: The system is based on product similarity. The similarity of products is measured through the comparison of all purchases made, not filtered by client type nor product domain. The purchases are leveraged to rank products which are often bought together.

Functionality: At the beginning, the similarity among all products is calculated based on all historical purchases and stored in a similarity matrix, which maps the similarity of each product to each other. The similarity matrix can then be used to get an ordered list of similar products for a given product. The top of this list can then be used as recommendations on the product view. For potential future use the functionality of this recommender is extended to receive a list of products. If more than a single product is given the recommender retrieves the similar product ranking for each of the given product and combines all lists together keeping the given ranking by similarity. The performance of this recommender is ensured by pre-calculating the product similarity matrix and retrieving solely the ranking just in time. This separates the computational effort from the consumer experience.

3.4. Evaluation

This section describes the internal evaluation of the prior described recommendation systems. It is not advisable to use an internal evaluation process for the “Try Something New” problem. This is since in this case, we are aiming at recommending products that the customer has not purchased before, which makes metrics such as accuracy or precision misleading. This recommender should be evaluated in practice, e.g. through A/B testing which will be further explained in the next chapter. The additional recommender systems show similar products in the product view and make a RF(M) scoring for the product search and can therefore also be hardly tested internally on their performance. Accordingly, the internal evaluation will be utilized for the “Did You Forget?” recommender system.

Typically, in data science use cases, machine learning models are trained and assessed on a separate test data set. However, in this case, training a machine learning model is not necessary. An alternative approach is to utilize distance measures and assess the performance on the entire dataset. The objective is to evaluate the recommendation system for different distance measures, such as cosine, jaccard and dice, and select the best performing one to make the final recommendation.

The evaluation employs a Monte Carlo cross-validation, which operates as follows: for a number of iterations, a set of unique customers is randomly selected, and for each customer, a random order is selected. This order is split into a partial basket (items used as historical data) and hold items (used for testing performance). Subsequently, the "Did You Forget?" recommender system is called to generate recommendations based on the partial basket. The top 15 recommendations are selected and the precision score is calculated by dividing the number of true recommendations by the number of recommendations displayed. This process is repeated according to the number of iterations, and each iteration is repeated for the number of customers defined.

Thus, this evaluation measures if the recommended products accurately reflect items that customers frequently purchase but might have omitted from their current shopping basket. Finally, the overall average precision will be displayed to enable an effective comparison of the different distance metrics used. The results from this internal evaluation are relative and primarily serve to compare the performance of different metrics, they should not be interpreted as absolute measures of effectiveness, as scores vary on factors such as basket size and composition. The best performance was achieved with the dice coefficient which will further be used in the “Did You Forget?” recommender.

4. Results Evaluation

Evaluating the performance of recommendation systems statically with a small subset of the historical data is not optimal. Instead, we advise to assess the performance of the different recommendation techniques in practice. This can be done by comparing the recommendation systems with different parameters in varying environments.

A/B testing can be applied to evaluate the performance of differently combined recommendations. For instance, the basket view can include types of recommendations differently displayed for the user. A first option (A) could be represented by displaying “Did You Forget?” and “Try Something New” recommendations in the same section of the website, maybe combining both recommenders into a single recommendation ranking. The alternative option (B) could be represented by displaying the two types of recommendations in different sections in the basket view. As shown in Figure 4, these options can be tested separately for different sets of customers. Also, simultaneously both groups should be compared with a control group, with comparable customers. Subsequently, the better performing option can be chosen as the final basket view. The performance can be measured accordingly to specific needs. A performance measure could for instance be represented by the number of clicks or generated sales through the recommendations.

Another important measure to assess the performance of the implemented recommendation systems is the computing time. As customer satisfaction decreases for delayed loading on websites, the duration of the computation for different types of recommendations should be measured and optimized. As priorly recommended and implemented, similarity matrices that are not individualized should be computed in advance and stored accordingly. With this method we efficiently distribute the necessary computation and enable significantly faster on-demand calculation of the recommendations. All implemented recommendation systems run locally in less than 0.5 seconds using the database provided for this use case, making them suitable for real-world use. Furthermore, in a competitive setting the infrastructure shall be more advanced, optimizing for computation and generate recommendations at an even faster rate. Specifically, it could be beneficial to cache the pre-calculations – techniques like Redis could support that.

5. Deployment

Customer centricity is vital for businesses to stay competitive and innovative. Therefore, the customer journey as depicted in Figure 5 shows where recommendations to the customers take place and which data the individual algorithms use for a better understanding how the system will be implemented and behave. The recommendation systems address the website of Recheio, for which we identified four critical views, namely the landing page, basket view, product search and product view which are shown in Figure 6.

When a customer enters the Recheio website, it checks if the customer is logged in or not, which will be triggering a specific recommender. For customers who are not logged in, the website forwards the customer to the **Landing Page**, which is described later. Since users that are not registered cannot purchase from Recheio their only option is to browse and eventually leave the website.

For logged-in users, our proposal is to check if this is their first purchase. If not, the customer is directed to the **Basket View** which is pre-filled with the same products which were selected in their last purchase. This not only speeds up the shopping process but also makes it more convenient by reminding customers of their usual needs. The client has the option to select the items he wants to re-purchase from his last basket or simply select all of them at once. The customer shall also have the possibility to create or save previous purchases as templates which can be reused and selected during this phase. The customer has the choice to take the selection to the Basket View or search for additional products, which leads to the landing page.

Additionally, individualized recommendations are displayed in the basket view. Firstly, the "Did You Forget?" section suggests products based on the client's historical data which is not only limited to the last purchase. Within this recommender system the current basket also influences the selection of recommended products, enabling an individualized recommendation of products for the customer which differs depending on the selected products. If the customer is satisfied with their selections, they may proceed directly to checkout from the basket view. Secondly, the "Want To Try Something New" section is displayed to the customer. These recommendations are based on customer similarity and show products that would get recommended to similar customers. Hereby, products that were never purchased by the customer will be recommended and possibly increase the number of purchased products from Recheio and the company's revenue.

Otherwise, if it is indeed the clients first purchase, or once they've reviewed their pre-filled basket, customers have the option to continue shopping. If they choose to do so, they are directed back to the Landing Page. This page illustrates the different product domains which Recheio offers to their customers. As the assignment of products to these domains was not given in the provided dataset, a proxy of this was implemented. For the deployment of the implemented approach therefore a mapping of the valid product domains to the products needs to replace the proxy.

When customers visit one of the product domains they arrive at the **Product Search** where the selection of products is sorted via the RF(M) recommendation system, meaning products are sorted after recency and frequency. If the customer is logged in, products bought by the specific client type are combined by a weight into the ranking of all product of the domain.

After selecting the desired product, the **Product View** will be displayed to the customer. Besides the description of the product and the selection of the desired quantity, the client will receive an individualized recommendation of similar products. The selected product will therefore be analyzed for combinations to all other products in the recorded purchases and their similarity in the selection of the customers is assessed and ranked. The number of recommendations should be limited to support an easy choice for the customer.

Subsequently, the shopper has the option to examine the updated Basket View and continue the purchasing process or ultimately finish the shopping journey by checking out.

This customer journey emphasizes a personalized shopping experience that leverages data-driven insights to enhance user satisfaction and convenience. By integrating intelligent recommendations throughout the process, Recheio not only facilitates a smoother shopping experience but also increases the likelihood of customer retention and higher sales volumes.

5.1. Next Steps

To further enhance the integrity and effectiveness of the RF scoring system, an important next step would be the integration of the monetary value into the existing scoring to evolve into a full RFM scoring system. Incorporating monetary data will allow us to assess not just the frequency and recency of purchases but also their value, providing a more comprehensive view of customer behavior and potentially increase the performance of recommendations.

Additionally, it is crucial to implement back-testing protocols for the recommendation systems to ensure that recommended products are available in stock (also called filters). This measure will prevent customer frustration and improve the efficiency of inventory management by aligning product availability with promotional activities. Ensuring that the system dynamically adjusts recommendations based on stock levels will further refine the shopping experience and operational responsiveness.

To optimize the effectiveness of the recommendation systems it is essential to apply A/B testing strategies. Avoiding or enforcing the practice of making the same product recommendations across various touchpoints, like the basket view and product view could be an additional goal of this testing. For instance, if a product recommended in the product view is added to the basket, it should not reappear in subsequent recommendations in the basket view. This dynamic adjustment can be tested by creating two variants: one where recommendations are static and another where they adapt based on the customer´s actions. Evaluating this will provide insights into customer preferences and help determine the most effective strategy for customer satisfaction.

The implementation of the RF scoring within the Product Search includes an internal weight system – 80% were applied to the RF score while 20% were considered for the individualization based on the customer´s client type. To further refine the effectiveness, conducting A/B testing with different weight configurations can devise valuable insights. By experimenting with these weights, we aim to determine if recommendations for this section should be kept more general or rather individualized.

6. Conclusions

In conclusion, the implementation of a smart recommender system at Recheio has the potential to revolutionize the shopping experience for its customers by providing personalized and strategic product suggestions. By leveraging historical purchase data, the system can enhance customer engagement across multiple touchpoints on Recheio´s website, including the Product Search, Product View and Basket View. The introduction of tailored recommendations such as “Did You Forget?” and “Try Something New” not only personalizes the shopping journey but also encourages larger and more frequent purchases, thereby increasing customer retention and boosting sales.

A robust internal evaluation process was designed to test the efficiency and metrics used in the “Did You Forget?” recommendations to optimize the system`s precision. Furthermore, the system has been optimized for enhanced computational efficiency by pre-calculating and storing similarity matrices. Thus, allowing for rapid recommendation generation in under 0,5 seconds to enhance the user experience without compromising response times.

The deployment of this system aligns with Recheio´s strategic goal of maintaining competitive advantage and market leadership in the Portuguese wholesale market. The proposed A/B testing, and continuous monitoring will ensure that the recommender system evolves in response to consumer feedback and market trends, maintaining its effectiveness and relevance.

7. Bibliography

Jerónimo Martins, SGPS, S.A. (2023). *2023 Annual Report - Non-ESEF Compliant Version.* Lisbon: Jerónimo Martins.

Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science, 3*, 57-63.

Casado, E., & Ferrer, J.-C. (2016, July 16). Consumer price sensitivity in the retail industry: Latitude of acceptance with heterogeneous demand. *European Journal of Operational Research, 228*(2), 418-426.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* New York: Free Press.

Gallo, A. (2014, October 29). *Harvard Business Review*. Retrieved May 6, 2024, from The Value of Keeping the Right Customers: https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

Report Linker. (2023, August). *Reportlinker.com.* Retrieved May 7, 2024, from Restaurant Market Overview in Portugal 2023-2027: https://www.reportlinker.com/market-report/Food-Services/9279/Restaurant?term=restaurant%20research&matchtype=b&loc_interest=&loc_physical=1011742&utm_group=standard&utm_term=restaurant%20research&utm_campaign=ppc&utm_source=google_ads&utm_medium=paid_ads

Appendix

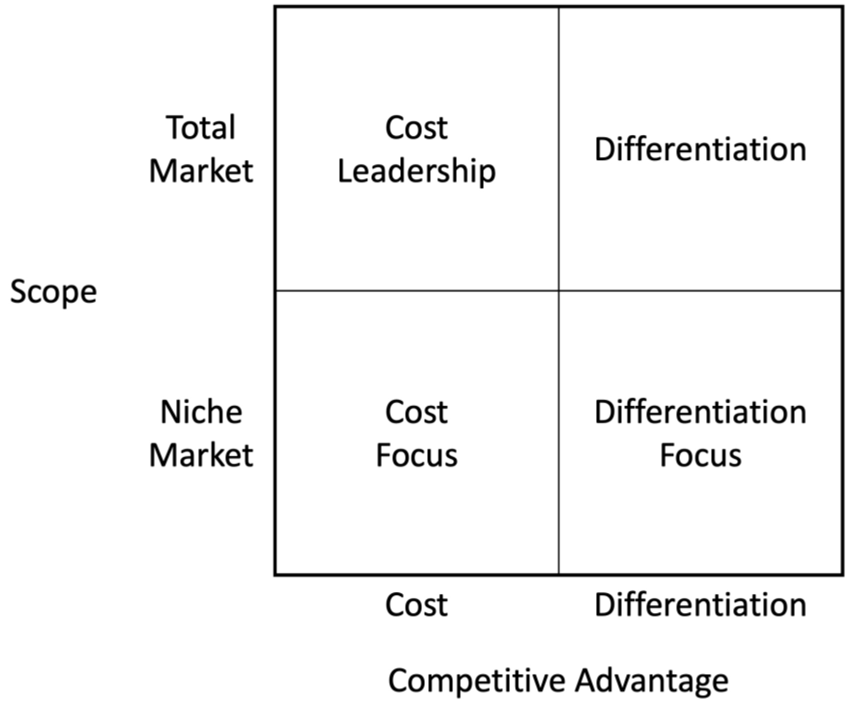

Figure 1: Porter's Generic Strategies

Table 1: Recheio's Dataset Overview

	Amount rows	Data type	Description
customer type			
ID Client Type	28	Int64	Unique identifier of a client type
Cliente Type Description	28	Object	Actual description of the client type identifiers
clients			
client_id	930	Int64	Unique identifier of a client
zip_code	930	Int64	Postal code where client is located
id_client_type	930	Int64	Foreign key of client type
products			
id_product	2498	Int64	Unique identifier of a product
product_description	2498	Object	Actual description of the indiv. product
id_product_category	2498	Object	Allocation of product category to indiv. products
transactions			
date	234224	Object	Recorded daily date for each transaction
client_id	234224	Int64	Foreign key of client
id_product	234224	Int64	Foreign key of product

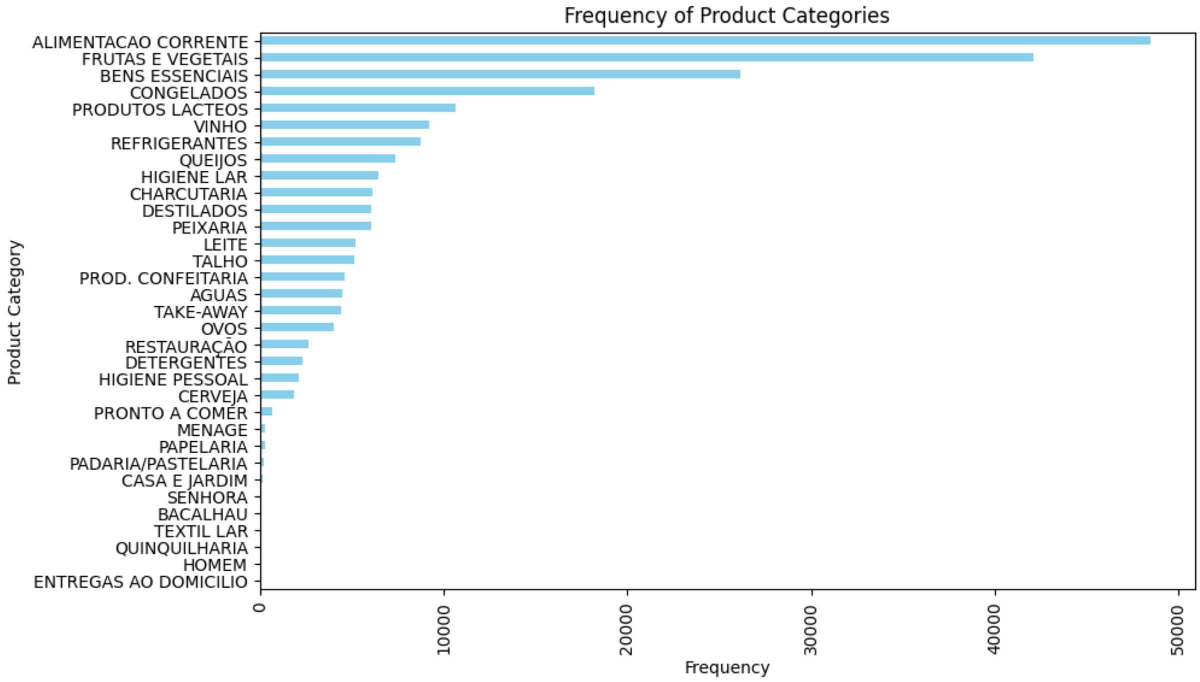

Figure 2: Distribution of Initial Product Categories

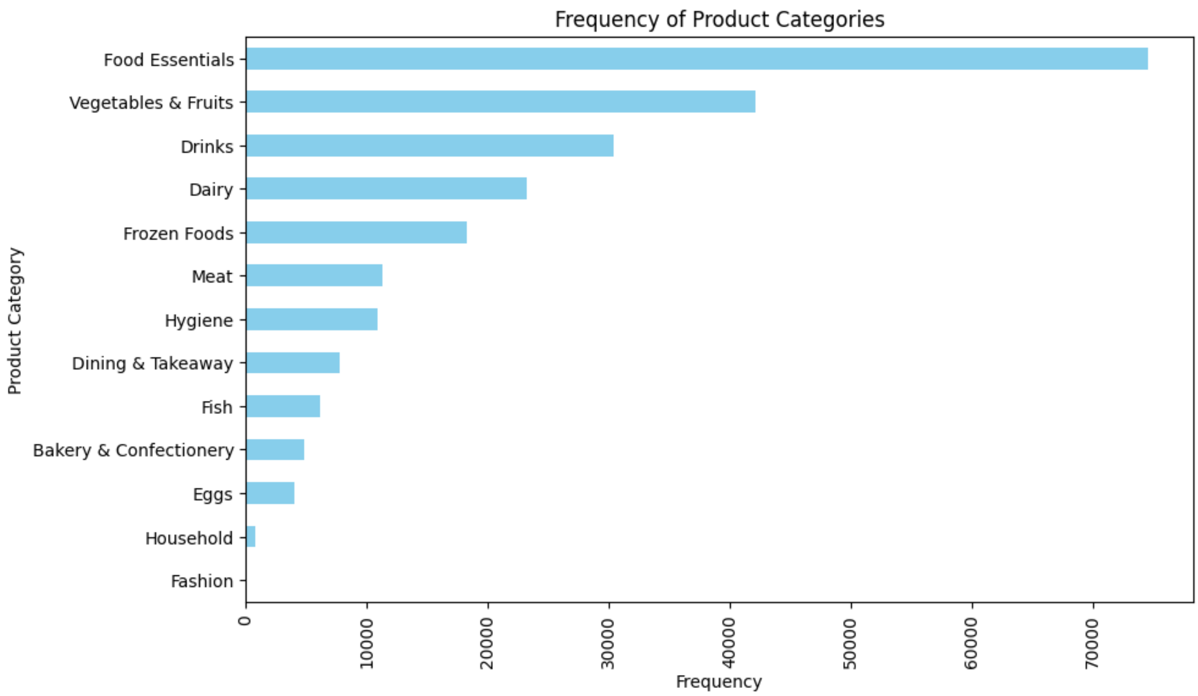

Figure 3: Distribution of Consolidated Product Categories

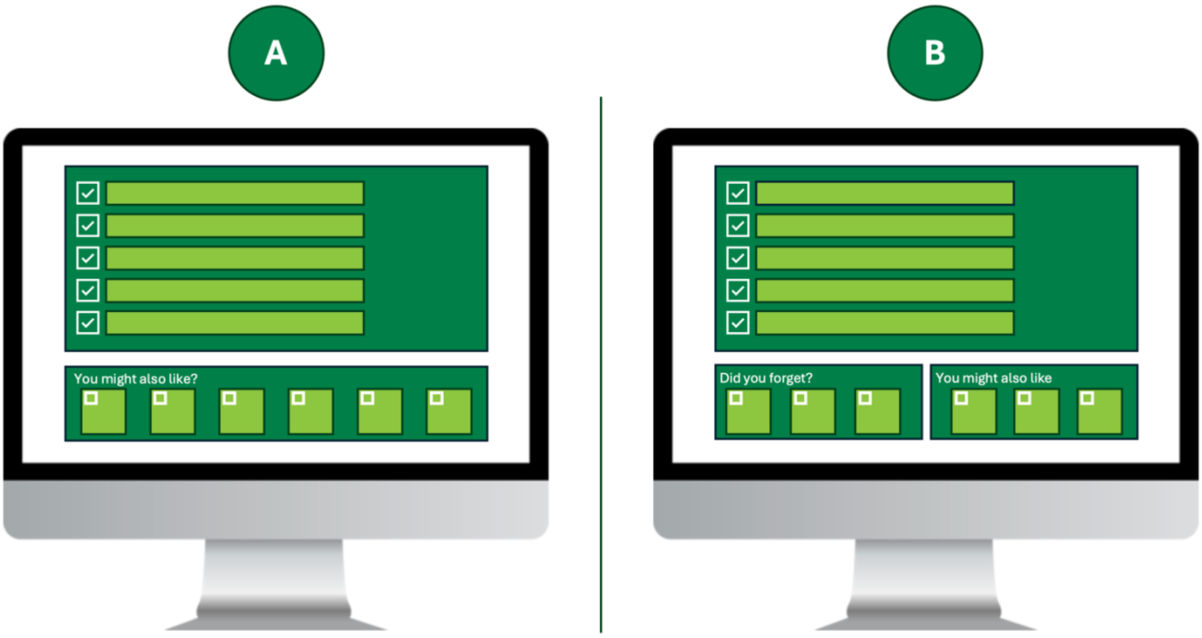

Figure 4: A/B – Testing Schema

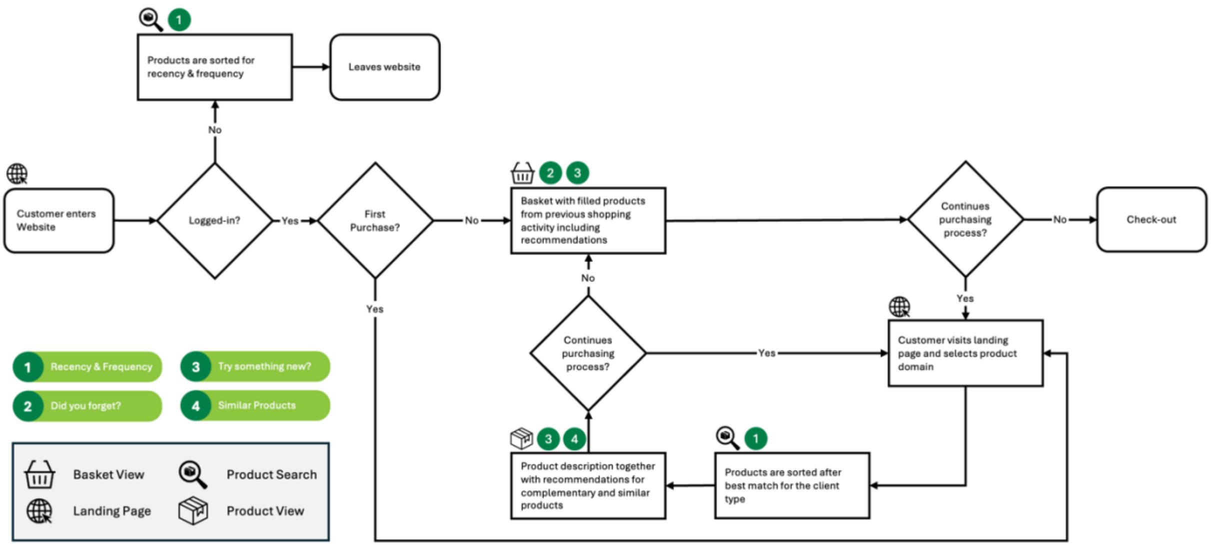

Figure 5: Customer Journey of Recheio's Website

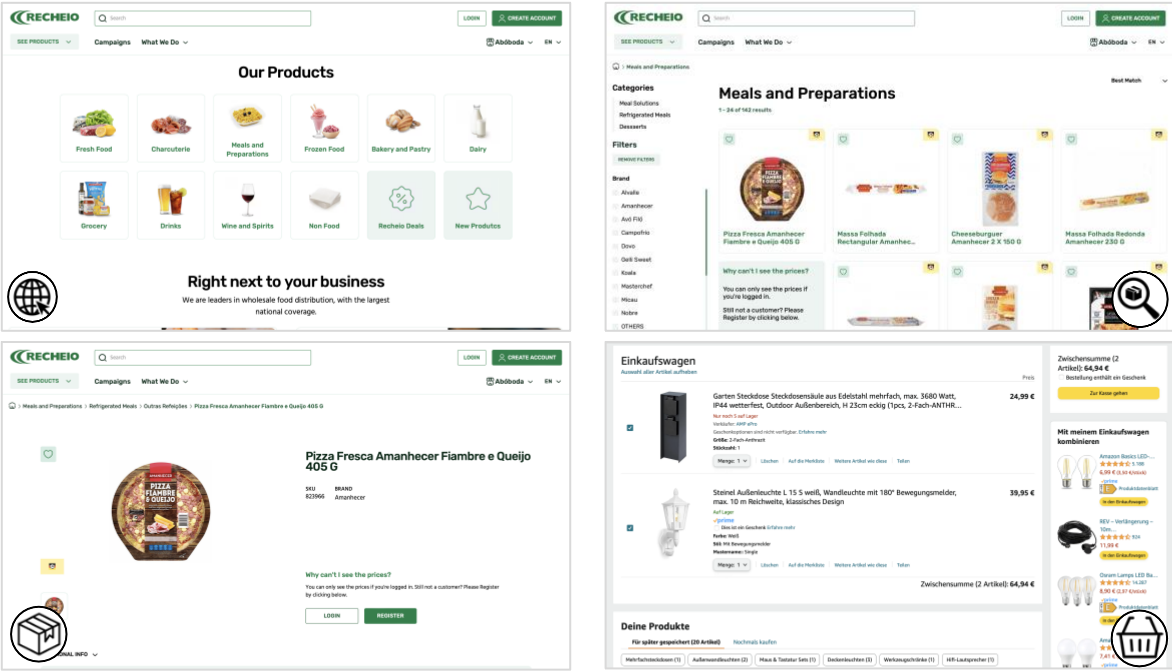

Figure 6: Landing Page (upper-left), Product Search (upper-right), Product View (lower-left), Basket View (lower-right, from another company as an example)