A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from Nova School of Business and Economics.

FIELD LAB JERÓNIMO MARTINS – OPTIMIZATION OF RETAIL OPERATIONS

FROM ENTRY TO THE EXIT:

A PINGO DOCE & GO NOVA CUSTOMER JOURNEY ANALYSIS

JOÃO ANTÓNIO QUINTA RAIMUNDO

Work project carried out under the supervision of:

Professor Qiwei Han

Mr. Rui Tomás

17-12-2021

Acknowledgments: To Professor Qiwei Han, for such an insightful feedback throughout the making of this thesis and for all the time dispensed. To Mr. Rui Tomás and all the collaborators from Jerónimo Martins, for the challenging topic proposed, the provided data, and the many guidelines that helped outlining this project.

Abstract: Recognizing the recent efforts and uses of process mining, this paper propositions the use of this analytical discipline to delineate the customer journeys at Pingo Doce & Go Nova. With a goal to detect improvement guidelines, and increase sales and profits, a curated event log on Celonis Execution Management System, a process mining software, was used to detect operational inefficiencies and outline the customers behaviors and experiences within the store, offering result-based management strategies and recommendations of future works.

Keywords: Process Mining; Retail Analytics; Event Log; Customer Journey.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Index

Introduction

The retail industry, one of the biggest sectors in the world, has faced unprecedented transformations over the past two decades with the growth of e-commerce and the increasing entries of new competitors into the market. As a consequence, merchants have suffered from the ever more market rivalry, feeling obliged to adjust their prices in an attempt to maintain their "low-price position" in the market, and exposing themselves to low-profit margins (Carter 2019). Therefore, retailers are confronted with an urgent need for continuous improvements, fostering disruptions in products, services, and operations. Among the most recent innovations are cashier-less supermarkets, leveraging on the latest technologies including IoT devices, numerous sensors, cameras, and machine learning techniques (Schogel and Lienhard 2020).

One of the most famous examples of this new-generation retailer is Amazon Go. Using several technologies, the store was able to automate a typical supermarket operation, eliminating pain points, namely long queues and long-lasting checkouts, commonly found in a traditional grocery store (Nadar, V. et al. 2018). Moreover, and despite requiring higher fixed costs at launch, being a nearly staff-free store, it significantly reduces employment expenditures, not to mention the explosion of customer data generated, gathered, and subjected to further analysis (Schogel and Lienhard 2020). Amazon Go pioneered a revolutionary grocery shopping experience, where the customer moves into and within the store, grabs the desired products, and leaves, being the shoppers' actions and payment assured on the Amazon Go app (Nadar, V. et al. 2018).

Like Amazon, the Portuguese retail giant Jerónimo Martins brought to life the Pingo Doce & Go Nova Lab, taking advantage of these new technological evolutions, and started exploiting the inherent benefits, specifically the generated new information. Initially inspired by the requests and needs of students for an extended range of cheap and convenient meal offerings at the old campus at Nova School of Business and Economics (Nova SBE), JM introduced its vision of the future retail to the market in October 2019, a project whose planning started before the new campus of the faculty was inaugurated. Fully packed with the latest technologies, a completely new shopping experience is provided on over 250 sqm in the brick-and-mortar store, designed to serve the customer needs for convenience and freshness in a matter of minutes. Besides traditional grocery articles for everyday use, it offers a wide variety of both on-demand and takeaway freshly prepared food items at low prices, directly competing with other on- and off-campus food options.

The typical shopping process starts and ends within the app, a key component throughout the purchase. After a self-check-in with their mobile devices, the customers use the app to self-reliantly add their desired articles to their virtual baskets through scanning or NFC, creating a quick and convenient customer experience. Having finished their item selection, the customer does not finish his purchase through a physical checkout but rather by in-app payment or at the payment stand. Throughout this process and beyond, data is created at various points. Apart from gathering data about customers at the point of registration, every action done through the app is saved. Moreover, the store generates production data at its IoT devices, ranging from ovens to kitchen devices that are connected to the internet. As every purchase can be assigned to a specific user, one has access to a detailed transactional purchase database that can investigate customer habits in an unprecedented way in the current retail context. Furthermore, the system stores data about article details and inventory movements. This never-seen-before variety and volume of data empowers a profound dive into various analyses about operational pain points and customer behavior.

Having predominantly young, tech-savvy students as shoppers, PD&Go Nova presents the ideal opportunity to experiment with innovative technologies and business models. In fact, solutions have proven themselves to be demanded and accepted in this lab-style store, qualifying for being potentially rolled out to other stores of the group. Thereby, more assessments of the lab's performance are required, for JM to continuously take the its potentialities to the fullest extent.

Problem Statement and Motivation

After launching the experimentation store at Nova SBE, JM identified two major areas for progress: gathering knowledge about customer behavior and the shopping process and identifying issues that hamper the retailer's operations from running smoothly. In other words, the retailer aims to detect, address, and solve problems and inefficiencies, whilst exploring areas with further advancement opportunities, seizing the support of data-driven decision systems that directly translate into bottom-line profitability progresses.

Thus, and in detail, this Field Lab covers four areas of interest. Firstly, process mining techniques are applied to get a deeper understanding of the purchase process of in-store shopping and the overall customer experience within. Secondly, market basket analysis is performed on shopping mission clusters to identify substitute and complementary products with a specific focus on freshly prepared food items. Thirdly, a demand planning tool is developed to support the day-to-day production planning for ultra-perishable items. Lastly, a semi-supervised learning anomaly detection model is developed to ensure regulatory conformity in the food preparation process.

The motivation for the store to focus on the abovementioned aspects are manifold. Indeed, identifying and modeling a standard shopping process and gaining insights into the shopping habits of customers enables the business to respond more appropriately to customer needs, behaviors, and preferences, ultimately enhancing the in-store experience for increased customer satisfaction. Furthermore, by revealing relationships between products, marketing actions can be taken to steer demand and provide recommendations for the store's assortment and replenishment strategy. Additionally, introducing a data-driven demand planning tool, comprising demand forecasting and operational planning, allows optimizing the trade-off between product availability and food waste. Lastly, guaranteeing conformity with the food regulation standards is fundamental for the store's operation ability and reduces inefficiencies regarding energy consumption and food waste.

All proposed measures are likely to contribute to sales or costs favorably, implying significant upside potentials for the store's profitability. The scope of this thesis, however, will focus on the customer journey analysis.

Objectives of a Solution

With the ability to track the customers' movements in the store, the ultimate goal of the analysis is to delineate the general steps a customer goes through in-store by using process mining, understanding the customer behavior inside the store, and the respective customer experience that is being provided. By doing so, it is expected that through observation of shoppers' flows JM can retrieve insights on how to better display the offered items, increasing the average shoppers' basket size and, thus, maximizing sales and increasing profits.

Moreover, with process mining, it is anticipated that one is going to be able to identify challenges occurring within the systems behind the storage of the data, which may harden the analysis, offering a starting point for improving possible inefficiencies found, as well as guidelines for future, more complete analyses on the topic.

In the next section, an extensive literature review on process mining and customer journeys is provided, justifying the use of this analytical tool to answer the addressed problem.

Literature Review

The emerging and incessant advancements in Information Technology have made it possible for companies to incorporate end-on processes, that is to say, a structured chain of ordered activities, with well-established starting and ending points and records, associated with a given case or process identifier (Davenport 1993). Over the past decades, corporations benefited from the digitization of said activities, and from the consequential rise of the amount of available data – the so-called event data, or event logs –, as it provides a massive opportunity to analyze and retrieve insights on the actions of people and systems. And here is where process mining plays its part (Janssenswillen 2019).

Van der Aalst defined process mining's as an analytical discipline whose end goal is to *discover, monitor, and improve real processes by extracting knowledge from event logs readily available in today's information systems* (Van der Aalst 2011, p. 1). It graphically portrays a real-life sequence of human and system action-generated data points – for example, setting a particular configuration within a mobile application, a customer receiving a receipt at a supermarket, or a machine detecting an error (Werner and Gehkre 2013). The uttermost process mining practices are three-fold, as described in figure 1: discovery, conformance checking, and enhancement. Process discovery makes use of the logging data to unveil a real process model and thus does not entail the exercising of any assumptions or *a priori* information. Conformance checking, on the other hand, establishes comparisons between the initial process and the true, real behavior exposed by the event log, with a view to detecting conformities – or the lack thereof. Finally, model enhancement takes diagnostic perceptions – commonly retrieved from the two previously mentioned techniques – to leverage the initial model (Wiriyarattanakul 2016).

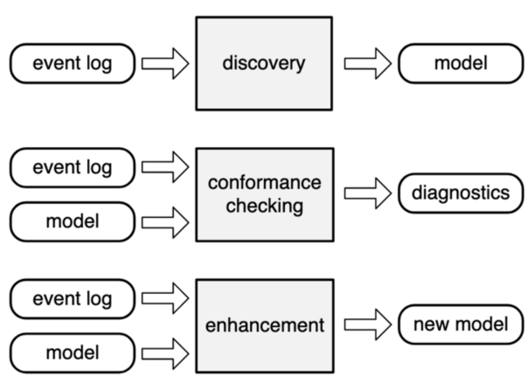

Figure 1 - The three techniques of Process Mining
(Source: Process Mining Manifesto 2011)

Each diverse technique has its applications: process discovery, say, can be used to extract process glitches in such a way that explores potential computer-engineering improvements; conformance checking for assessing the quality of a previously drawn process and therefore auditing real vs. modeled processes; and model enhancement to escalate the potential data-driven insights, for example, through the addition of more event-related attributes (Van der Aalst 2011).

The benefits and aptitudes of process mining are undoubtedly remarkable. However, such analysis requires a careful and extensive way of manipulating the data and understanding the output. On the Process Mining Manifesto, the IEEE Task Force on Process Mining enumerated several encounters[1] that come along when pursuing a process mining study.

Firstly, what is the quality level of the inputted data? Frequently, the records of an event log comprise identifiers that relate to objects across other dimensions, which need to be combined. For example, if the event log holds information regarding the customer's usage of a particular software, and any customer-related information is stored on a different database, the two need to be merged. Moreover, as it refers to real processes, it may enclose infrequent – although correct – actions, defined as noise or outliers, which boost the complexity of the output and, therefore, extend the need to clean the data. This topic is magnified by the data's risk of incorporating erroneous entries, due to *software malfunctioning, user disruptions, hardware failures or truncation of process instances* (Werner and Gehkre 2013, p. 10). Finally, the singularities described by the logging data occur within a framework – say, a period of the day – and that context should be reflected on the analysis, implying, once again, merging of data, or feature-engineering, and consequential feature explosion.

Another denoted problem concerns the multiplicity of individualities of the event log. In some cases, the extent of low-representative events is such that they are of little to no interest for the analysis, as it noticeably heightens the complexity of the visualizations and hardens its

[1] More challenges are presented on the Process Mining Manifesto. However, only the most relevant for the scope of the thesis were selected.

comprehension, and thus requiring some event aggregation.

Thirdly, a question of assessing the output's quality emerges, namely the balancing between the four quality benchmarks: fitness, simplicity, precision, and generalization. Say, a model should be capable of representing most of the paths in the logging data, whilst explaining the process in a simple way. Simultaneously, it should be precise on the displayed behavior, evading less-probable paths, while maintaining capacity to generalize over broaden routes beside the observed.

Lastly, the PMM points out the importance of ensuring the correct interpretation of the process outcomes. Every detail aiding the user's understanding of the output, such as the *trustworthiness of the results* (Process Mining Manifesto 2011, p. 17), ought to be plainly specified.

One last indication regarding process mining concerns the features needed to perform such analysis. These include the case identifier pinpointing the different cases recorded and the respective events, the timestamps of such traces, and the event itself. Naturally, more attributes can be included on the log, as mentioned before (van Zelst, et al. 2020).

Having said that, how does this tool apply for the scope of the thesis, that is to say, for the analysis of the customer journey at Pingo Doce & Go Nova?

A customer journey is the process that a customer goes through to reach a specific goal, involving one or more service providers […], *detailed as a series of touchpoints or interactions between the customer and the service providers.* (Føstad et al. 2014, p. 2). It maps the footsteps of the customer's contacts with a business, for example, in a retail store (Richardson 2010), evidenced by reference points or events that allow for this mapping to be possible (Føstad et al. 2013).

A first approach to the consideration of process mining for customer journey analysis was conducted by Bernard and Andritsos. The two authors defended the application of logging data on the process discovery technique to expose the true customer's path – a "*de facto" model* –, along with the conformance checking procedure to establish comparisons – a *"de jure" model* –, bringing the gap between expected customer journeys and actual ones (Bernard and Andritsos 2017). In fact, companies are starting to perceive the benefits of mapping customer journeys with process mining. Prolifics, for instance, offers process mining solutions for processing and analyzing logging data, emphasizing how a company can know the true interactions of the customers – plus their unbiased preferences and experiences – with the corporation, by simply diagramming their factual movements (Prolifics 2020).

Taking into consideration everything referred above, process mining, with its values and challenges, is undeniably the tool needed to scrutinize the customer journey at Pingo Doce & Go Nova, from the entry to the exit of the store.

Design and Development

Data Collection

The core data – the logging data – was provided by JM, and ranges from October 1st to October 31st, 2021. Furthermore, the company supplied two look-up tables: one containing product-related data – from the location of the product at the store to its category and subcategory – and the other holding anonymized customer-related data. In appendix IV, one can observe a snapshot of the abovementioned tables.

As pointed out before, the quality of the output of the process mining highly depends on the quality of its input. Consequently, a careful data pre-processing was put in place in the next section.

Data Curation and Understanding

By having a first perception of the event data, one can easily denote that it comprises all events generated either by the users' interactions with the store via the PD&GO Nova App or by the system itself. Case in point: a customer introducing a new payment method on the app, entering the store, and scanning a product, the system performing basket valuations to infer the risk of theft, the opening of the Go 24/7 cabinet door, the coffee machine storing in queue demanded products, every little operation is recorded on the event log. As so, the first step was to build an understanding of every single event observed and, subsequently, organize the different events into groups.

The next stage was outlined by investigating missing values, which exposed approximately 3% of the registrations without an associated customer. Being this a customer-centric analysis, all the correspondent rows were deleted from the dataset. In appendix II, table N provides an overview of the abovementioned categories, together with a brief explanation of the client-linked events registered on the log. Here, the customer look-up table also played an important part. This dataset had all customer-related information, including its status in JM systems and an indicator portraying whether the client eliminated his or her account. Once and again, a customer-centric analysis should not reflect the paths of blocked or inactive customers – perceived here as buyers who will not return to the store –, and thereby, with these attributes, one was able to depict active shoppers, both on backend and frontend. In light of this, only events linked to these customers were preserved.

Thenceforward, a careful examination of the descriptions of the events – the attribute EVENT_DESC – was conducted. These column values covered large amounts of information, frequently already present on other features, thus, only the products' description was kept.

The previous step gains importance when bearing in mind the need to combine the products look-up data with the event log, the next stage on the data curation corridor. Although the key element for this merging would be the item code, such information was not available for some records, and therefore the union of the two tables was performed separately: when the product code was available, the datasets were merged on said identifier; otherwise, the join occurred on the products description. Here, a problem arose: not only 62 product codes did not have a match on the articles' look-up data, but also 302 item descriptions failed to meet their pairs as well, mostly due to spelling differences, meaning that multiple rows had missing information. To solve this, an extensive manual process was put into practice.

The following phase was about keeping the relevant data points to depict the customer journey. In short, all events not directly related to the clients' path through the store were removed, such that only three categories were kept: Check In/Out, Add/Remove Products, and Payment – as described in appendix II. In addition, it was of importance to ensure that, for each visit, the first and final events resembled the customers' entrance and exit from the store. When beholding the true movements of the clients on the beginning and end of their paths, the data evidenced, for example, customers with two entry-related events, so these flawed records were deleted.

Thereafter, a breakdown of the information described on the article look-up table was performed. Only one problem was found: for the equivalent description, a product could have two different codes, and thereupon, when merging the log and product data on the article descriptive, one would face rows duplication. The solution passed through observing which products comprised this poser, tracking down the correspondent duplicated rows, and removing them.

Finally, an intermediate table was created to support the preliminary analysis and extend the data cleaning. In short, said dataset holds in each row a visit, namely the related client, and entry and exit datetimes, hence allowing to compute the duration of each visit. Hereby, it became feasible to explore outliers – visits lasting for more than 30 minutes – and erroneous entries – visits enduring for hours or even days –, all of which were spotted and excluded from the log data by keeping all events occurred between entry and exit touchpoints distanced by less than 30 minutes.

As a final, though crucial step, a case identifier was built to run the data through a process mining software, and a new event column – CATEGORY_AGG – was fashioned based on the products category to reduce the number of unique or low-representative paths. For a stronger and deeper understanding of the process undertaken in this section, consider appendix III.

Descriptive Statistics and Exploratory Data Analysis

After passing the raw event log through the data cleaning tunnel, one was left with a total of 271 742 records in the month of October, corresponding to 55 415 visits within the criteria established in the previous section, an average of 2 217 visits per day – considering the lack of visits on Sundays and national holidays –, and an approximate average of 4.39 events by visit.

The first analysis performed took into consideration the thought that the consumers' desires when going to the Pingo Doce & Go Nova are likely to vary with the time of the day. By looking at the top products added to shoppers' baskets, one can easily detect differences in the customers' missions[2]. For example, a client entering the store on a weekday early morning typically goes for coffee machine products, pastries, and bakery products, whilst on lunchtime is looking for pasta, pizza, or hamburger, and on a Saturday afternoon, he or she is longing for a beer.

The aforementioned reasoning was bear in mind for further analyses. When examining the distribution of the visits across the different days of the week, and throughout the time of the day – figure 2 –, one can observe two undeniable facts: during the weekdays, the customer presence within the store tends to follow the same behavior, which in turn differentiates itself from the visits on the Saturdays – no visits are recorded on Sundays; the peaks occur at break times between classes, which was expected.

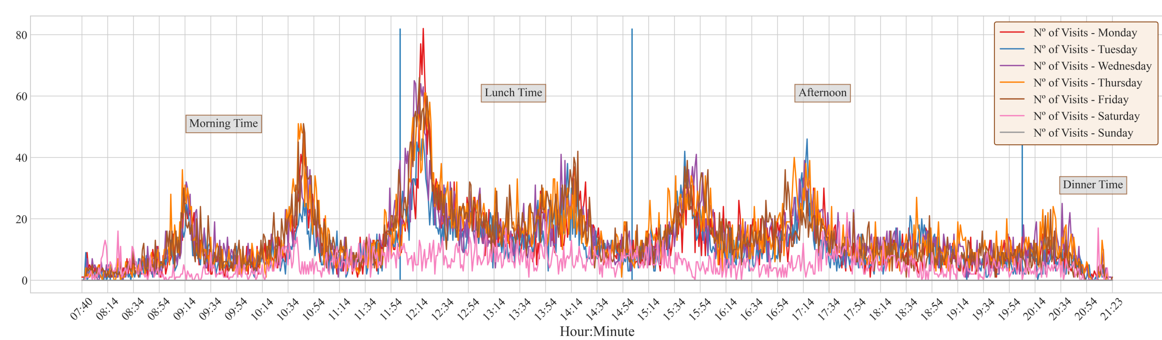

Figure 2 - Visits throughout the time of day, by day of week

[2] JM refers to the clients' goals when going to the store as missions, that is, as the goal they have in mind in each visit.

Regarding the duration of the visits, the data indicates that the common customer spends an average of 4 minutes and 39 seconds within the store and that more than 50% of the visits last between 1 to 4 minutes, as described by figure 3. Moreover, table 1 indicates that the average time inside the Pingo Doce & Go Nova varies throughout the time of the day and whether the visit occurs on weekdays or on the weekend.

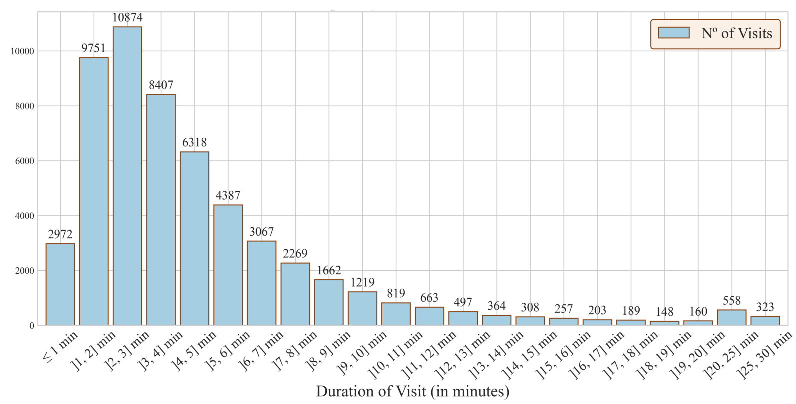

Figure 3 - Frequency of Visits' Duration, by time intervals

	Weekdays	Weekends
Morning	3 minutes and 41 seconds	3 minutes and 42 seconds
Lunch Time	5 minutes and 22 seconds	6 minutes and 4 seconds
Afternoon	3 minutes and 57 seconds	4 minutes and 26 seconds
Dinner Time	6 minutes and 16 seconds	5 minutes and 3 seconds

Table 1 - Average duration of visits, throughout the time of the day and days of the week

All the same, it is strongly believed that the shoppers have different behaviors according to the different times of the day and different days of the week, and thus the customer journey analysis will be segmented as follows: mornings, as the products demanded and average durations are equivalent; lunchtime on weekdays *versus* lunchtime on weekends and afternoon on weekdays *versus* afternoon on weekends, given that durations and most demanded products vary; and dinner time throughout a whole week – see appendix V.

Another aspect worth mentioning is the number of visits where customers enter and leave the store without recording any other event. According to the log, these correspond to 21 366 visits, a considerable percentage (38.6%) of the total visits, and a number that could have been enlarged by the cleaning process undertaken. An explanation of the cause of this singularity is later on provided in chapter "Challenges and Limitations", after providing a broad explanation of the Celonis software used – the next section – and presenting the process findings.

Celonis Execution Management System

Celonis, a pioneer in process mining services, is one of the biggest companies offering data handling solutions for firms to run their business processes. In 2020, the Bavarian company launched its own process mining software – the Celonis Execution Management System (EMS) – , built upon technology that combines multiple areas of interest, from analytics to management, making it the perfect system for the customer journey analysis. With case identifiers, activities, and timestamps built upon the provided log data, the software aggregates it to uncover the true flows inside the PD&Go store, providing ample visualizations, and therefore fulfilling the scope of this thesis. However, some precautions were to be considered before applying the data to the system.

As stated before, a common problem of discovering paths regards the large number of possible journeys, as more and more unique paths highly increase the complexity of the visualizations, hardening its comprehension. Ergo, instead of specifically presenting the product that was added to or removed from the basket, the analysis mainly focuses on the category the item is included in, allowing for more paths to be aggregated into one. Moreover, the software allows controlling the number of activities presented in the visualization, as well as the number of connections, evidencing a trade-off to be considered between including more information, at a cost of increasing the results' complexity. With that in mind, for all analyses, groups of products were used/created to improve the aggregation of flows and increase the amount of information displayed.

Process Mining Findings

Before entering through the findings on the customer journeys, an explanation on how to interpret the outputted visualizations is in order. To that extend, consider as an example figures 22 and 23, in appendix VI. Each event is represented in containers, and within that container, one can find the description of the activity/event and its frequency in the event log. For example, a product included in the category DRINKS was added to a customer's basket in 7 843 visits. In the same way, each arrow represents a direct connection between two events, symbolizing a link between two events that are sequenced in time, and the frequency of that connection is also displayed. For instance, in 5 964 visits, the customers successfully pay the basket at the payment stand and directly receives the respective receipt. Moreover, one can express the average time passed between the occurrence of the two events. For instance, on average, approximately one minute goes between scanning the product in category DRINKS and paying the basket. Lastly, it is noteworthy to mention that in every visualization, not all activities and sequences are displayed within each context – for example, paths at lunchtime –, as doing so would explode the paths' conception complexity and significantly harden its comprehension. Here, each image indicates the percentage of activities and connections conveyed, solving PMM's understandability problem.

A Broad View of the Customer Journeys

Starting with the shoppers' paths in the morning – figure 4 –, throughout the whole week, the most observed event is "Add BEBIDAS", occurring in 3 304 visits, meaning that the customers' mission when early arriving at Nova is all about drinking coffee machine beverages. However, other items' categories are searched for, and commonly found on the same basket, such as pastries, bakery products, and fruits – "Add PASTEIS, BOLOS, MERENDAS, …" and "Add FRUTA", respectively. Additionally, the average times between events are small – around 1 minute –, conveying the idea that a customer desires a fast though fulfilling visit to the store.

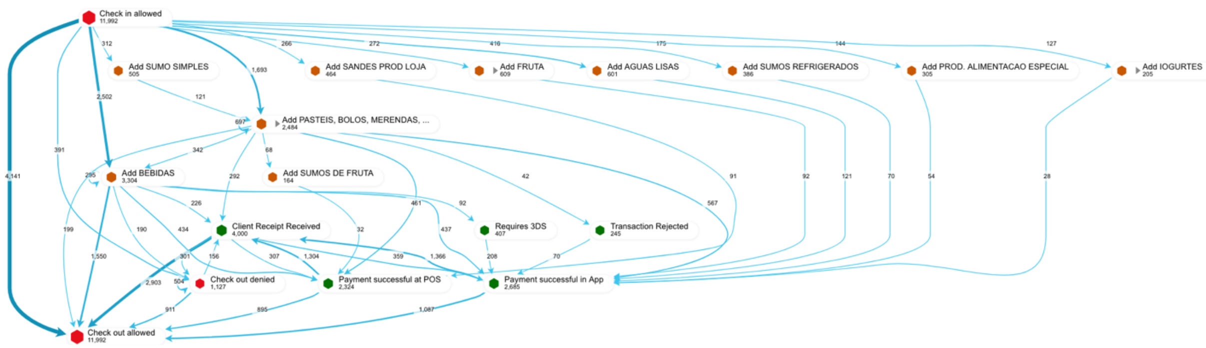

Figure 4 – Customer Journeys in the Morning (Visit Frequency View)

[93.4% activities;73.7% connections]

Regarding the lunchtime trails throughout class days, there is still prevalence of coffee machine products. Nonetheless, other expected products' categories emerge, such as fresh pastas and salads – "Add MASSA E SALADAS PRODUÇÃO" –, as well as pizzas, sandwiches, and sodas – "Add PRONTO A COMER", "Add PIZZA", "Add SANDES", and "Add BEBIDA COLA". On Saturdays, the customers' desires resemble the one described, though evidencing a clear change of preference for Beer. Furthermore, the time passed through each sequence of events varies. For example, entering the store and scanning a pasta or a salad takes, on average, one minute less on Saturdays than on class days – figures 5 and 6.

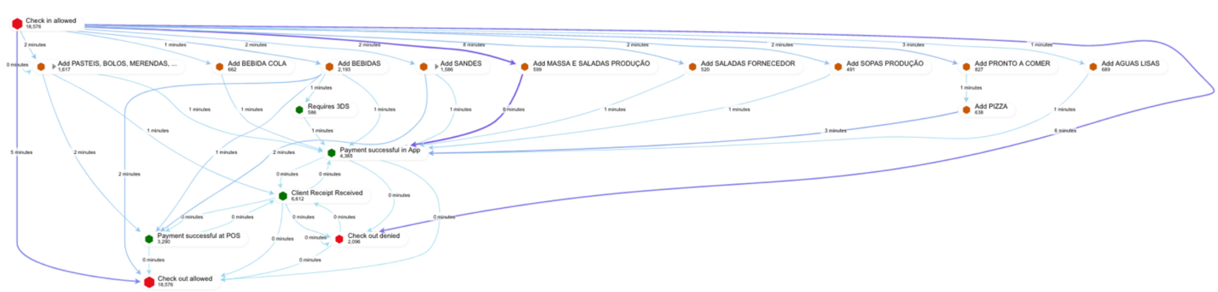

Figure 5 - Customer Journeys at Lunchtime, on Class Days (Throughout Time Average View)

[88.2% activities; 65.5% connections]

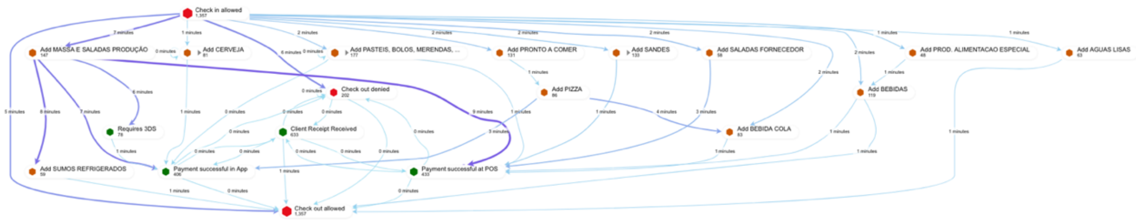

Figure 6 - Customer Journeys at Lunchtime, on Saturdays (Throughout Average Time View)

[86.4% activities; 59.6% connections]

Proceeding with the daily analysis, the buyer's behavior after lunchtime on class days reminds the one portrayed in the mornings when the products' desired are concerned, although there is a higher preference for pastries and bakery items, instead of coffee machine brews. On Saturdays, still, the most demanded drink is beer, a phenomenon also observed on Saturday lunchtime. In time terms, the paths undertaken are equivalent – see appendix VI.

Finally, at dinnertime – figure 7 –, and with respect to the desired products and respective flows, the customer journeys bear a resemblance to the lunchtime traces. Notwithstanding, some throughout time averages between flows are smaller – for example, the average waiting period to order a pasta since the moment of check in is 2 minutes shorter.

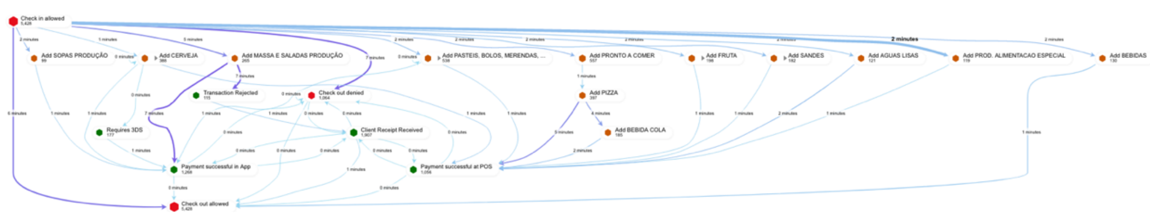

Figure 7 - Customer Journeys at Dinnertime (Throughout Time Average View)

[88.1% activities; 64.5% connections]

Being the customer paths presented, one should submerge into some aspects that were unexpected. It was known that many visits followed a particular path: check in, and direct check out, with no middle events being recorded. However, a noteworthy phenomenon occurring is the amount of time by-passed in between, which averages 5 minutes, and, in some cases, the visit lasts for more than 10 minutes, an interval where the paths of these customers are completely unknown, and therefore this reality uncovered.

Moreover, the log unveiled questionable flows of some events. A general optimal path of a customer would be represented by an entry in the store, followed by addiction or removal of products from the related basket, the payment and reception of the receipt, and, finally, a check out. However, in some cases, this trace is not strictly tracked. For example, the event data registers visits where the customer firstly receives the purchase receipt, and proceeds to payment. Another common case is visits where customers add coffee machine brews and directly leave, without paying in-between. The data seems to have ordering failures, possibly compromising the analysis.

Deep Diving into Specific Journeys

Grounded on the previously retrieved conclusions and considering that a full analysis of the customer journeys incorporates a deconstruction of their experience, this section covers specific studies on individual items demanded by the customers. For these analyses, the products themselves were considered, and manually bundled together into groups in the EMS software.

The first exploration regards all visits where coffee machine brews are added to the basket, as these are frequently demanded. As one can notice, figure 8 suggests customers going for coffee machine beverages often added other items to their basket first, which, considering the respective products closeness to the coffee machine in the store, conveys the success of the strategy followed regarding the articles display. Effectively, most of these extra products are on the same store area and division – “Specialized Perishables” and “Bakery/Pastry”, respectively – as the coffee brews.

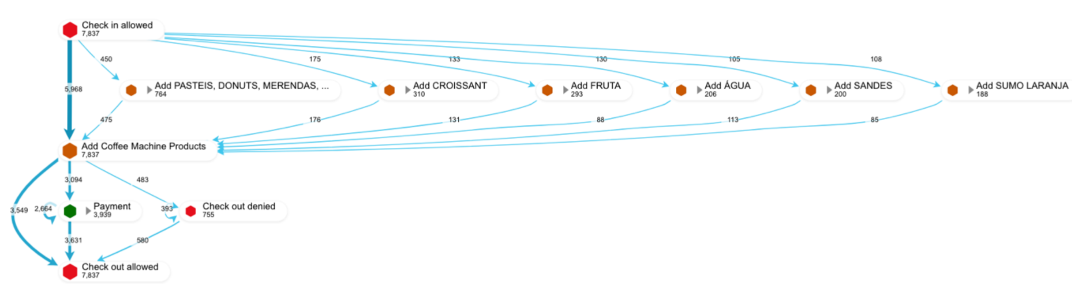

Figure 8 - Customer Journeys including Coffee Machine Products (Visit Frequency View)

[95.1% activities; 78.2% connections]

JM can benefit from these insights, promoting, for instance, bundling strategies aiming to increase the average size of a customer basket. However, is there an opportunity of improvement?

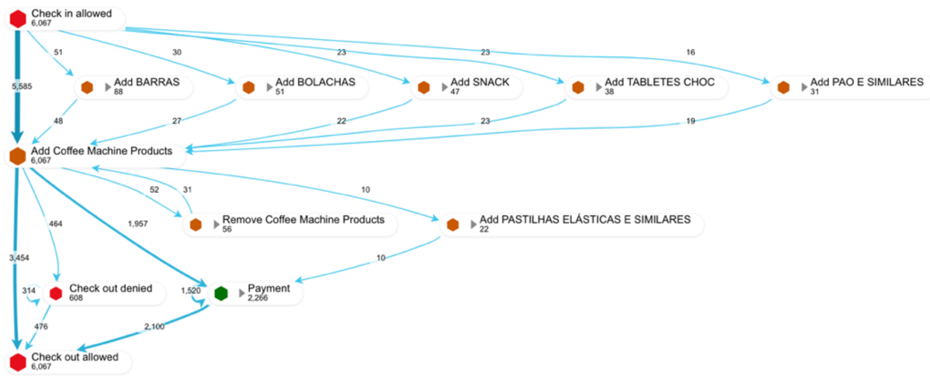

Figure 9 - Less Frequent Customer Journeys including Coffee Machine Products (Visit Frequency View) [98% activities; 88% connections]

Figure 9 presents less common customer paths, delineating other products groups added to customers' baskets, and thus exploring new possible shelves' organizations. In fact, these items are displayed on a different area and division – in turn, "Grocery + Pet Food" and "Everyday Food" – and thereby displaying these products close to the coffee machine could lead to an increase on the average basket size, and in turn, on sales[3].

Finally, it is acknowledged that going for a pasta or a salad produced within the PD&Go store often comprises waiting a considerable amount of time to order the desired meal. Therefore, to better assess the customers' experience, an analysis of these hanging around periods is in order.

Figure 10 outlines customers' paths where these customer's go for a personalized pasta. One can notice that between entering in the store and scanning the product, an average of 6 minutes goes by, and 7-8 minutes elapse from that moment until the customer records another event. In other words, the customer waits 7 to 8 minutes, on average, for the ordered meal.

[3] No segmentation on Class Days *versus* Saturdays, as the products less demanded are the same in both cases.

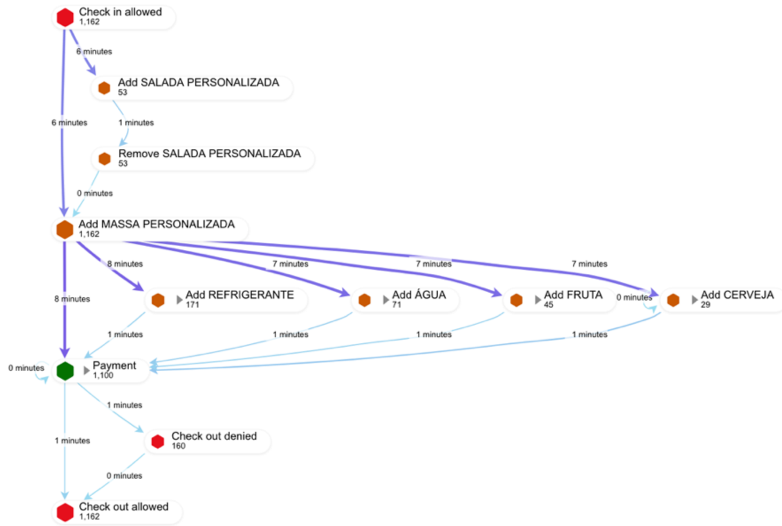

Figure 10 - Customer Journeys including Personalized Pastas (Throughout Time Average View)

[90.2% activities; 73% connections]

Deconstructing the time waited since check in until the scan of the product throughout the day, on class days and Saturdays – figures 11 and 12 –, it can be perceived that customers' awaiting time to order a pasta is exceptional at lunchtime and dinnertime, where the meal is the most demanded – peaks on class days [12:30, 13:00[and Saturdays [14:00, 14:30[minute intervals. Throughout the rest of the day, the average time per bin evidence considerably smaller hanging around times.

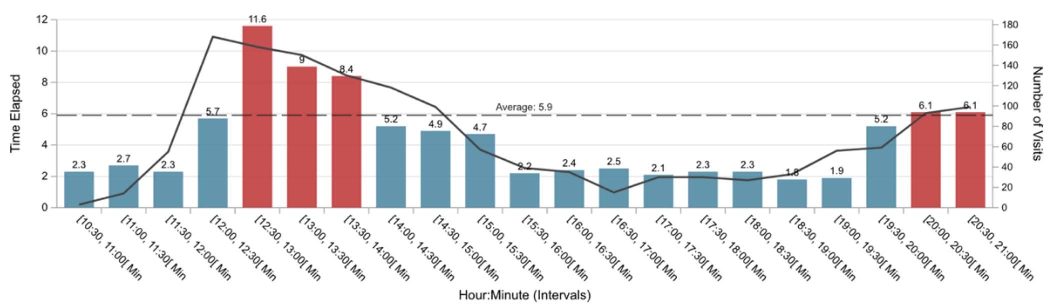

Figure 11 - Distribution of Ordering Times for Ordering Pastas, by Hour:Minute Interval (Class Days)

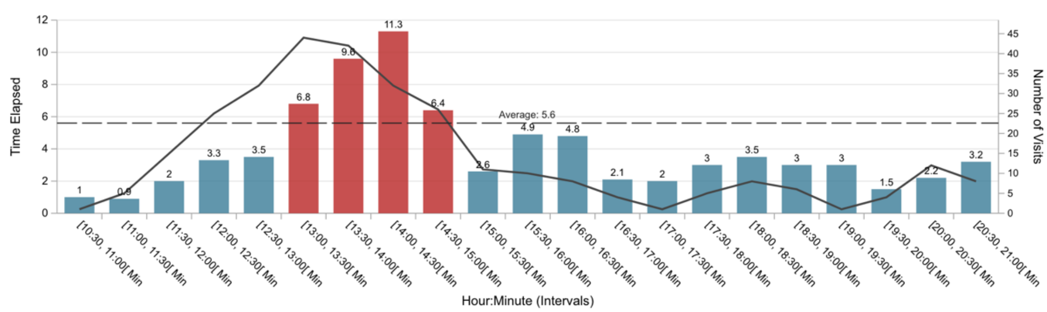

Figure 12 - Distribution of Ordering Times for Ordering Pastas, by Hour:Minute Interval (Saturdays)

Regarding the waiting times after ordering a personalized pasta – figures 13 and 14 –, it can be perceived that customers' wait for pastas beyond average at peak hours, when the demand is higher. Moreover, the proportion number of visits/waiting time is greatly elevated at dinnertime on class days, where demand is not as high as during lunchtime, and the lingering times are higher or equivalent. For the Saturdays, one can see that for more time bins the customers wait above average, perceiving the idea that the, on Saturdays, the service for this fresh meal is overall slower.

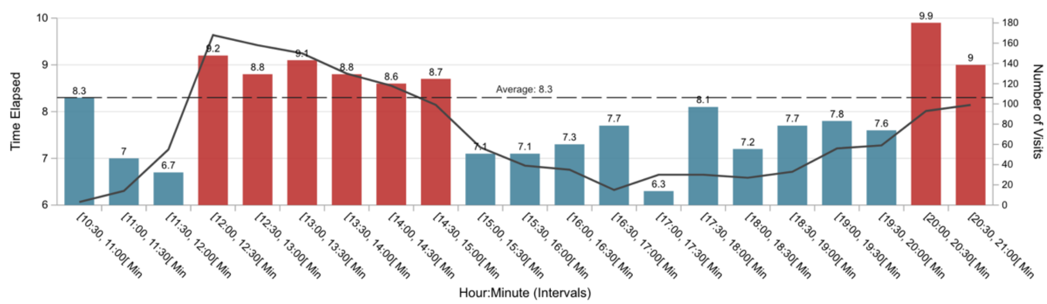

Figure 13 - Distribution of Waiting Times for Personalized Pastas, by Hour:Minute Interval (Class Days)

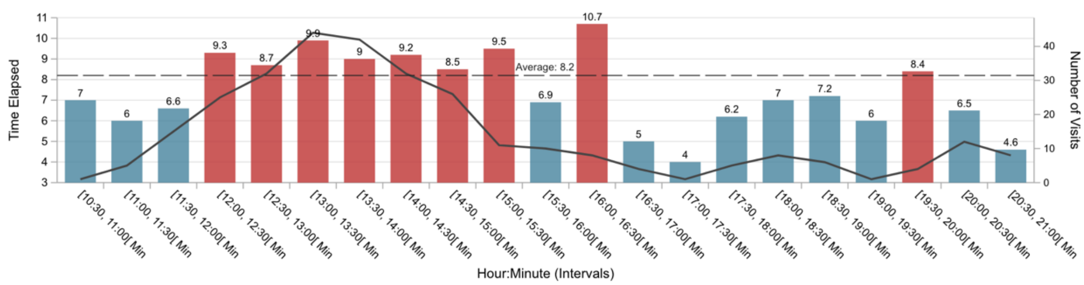

Figure 14 - Distribution of Waiting Times for Personalized Pastas, by Hour:Minute Interval (Saturdays)

Lastly, for personalized salads, the waiting periods for both ordering the meal and getting hold of it are smaller, on average. Opposable to the pasta service, the cooking times on class days seem to be higher for salads, and the overall demand lower on Saturdays, but this insight may not be stronger enough considering that, on Saturdays, less data is available. At heart, JM could benefit from reducing the cooking periods of these two meals, and thereby dropping the time customers' wait in queues, improving their response to the demand and, again, boosting sales. Moreover, tweaking these friction points would allow for a faster service, enhancing the customers' experience within the store, and thus refining customer satisfaction.

With respect to the evaluation of these results, one can observe, in the legend of each log visualization, the percentage of events and unique connections that are being observed. In all, more than 80% of the events are analyzed, and more than 50% of the connections are being displayed. Considering that a major pilar of assessing process mining results consists of their simplicity and that less-probable paths are to avoid – and thereby justifying the lower percentage of networks shown –, it is believed that the visualizations convey a robust result to the addressed problem. Nevertheless, generalizing the model with more data would convey a stronger assessment.

Challenges and Limitations

The potentialities of conducting shoppers' paths analysis are undoubtedly vast. However, some risen challenges and limitations ought to be addressed.

The first regards the merging of the event log with the products' look-up table. Adding more information to an event log is a problem already conceived in 2011's PMM and experienced on the cleaning step of the process. With missing product identifiers, combining the two datasets using as key the items description was an issue, as a noticeable number of items had different descriptions according to each set of data, due to misspelling mistakes, hardening the automation of this step.

The second concerns the temporal ordering of the events, one of the findings presented. An initial fact that must be directed is that JMs event-processing system does not entirely ensure the correct succession of records, that is, events may not be necessarily processed by the order through which they occur. Related to this challenge is the fact that no visit identifier is automatically stored in the event log, and therefore, unsystematic records are harder to associate to a particular visit. This represents a significant drawback. When cleaning the data, and to ensure each visit had one and one only entry and exit touchpoint for a proper visit identifier to be built, all events outside of these reference instances were eliminated, meaning that some visits are missing certain records.

Having a case spotter would allow to detect these incidents and thus to analyze them, gaining more insights on the context and reasons behind this problem.

Thirdly, another limitation of this analysis relates to the vastness and complexity of customers journeys conveyed by the event log, a problem also stated on the Process Mining Manifesto. Referring to customers actions within a grocery store, there are probably thousands of different paths to be portrayed, and therefore an important step of this analysis was to aggregate the individualities as much as possible, whilst still preserving the maximum feasible amount of information. Outliers and erroneous entries were removed, and products grouped in order to ease this problem, and yet not all individualities are studied.

Finally, regarding the trustworthiness of the results, this indicator could be higher if a larger period of data was considered, allowing to convey stronger customers' paths. This limitation is enlarged considering that, as referred above, a large portion of the paths throughout the month of October are to be discovered, on which no possible insight on the shoppers' path can be retrieved.

Recommendations for Future Steps

One of the process mining most used techniques is, as aforementioned, model enhancement, which leverages on previously insights exposed by process discovery to enrich later developed analyses. Therefore, it is suggested for JM to take full advantage of the main findings of this study to improve the respective events-processing system, ensuring an accurate sequence of records in time and wholesome customer journeys. To this extent, including a visit identifier in the event log is a plus, warranting a clear relation between each individual record to each particular visit.

The following proposal regards the problem of high number of individualities, typical of a process mining analysis. To solve this problem, performing behavioral clustering on each visit of each customer is a recommended approach. Behavioral clustering is commonly used to aggregate shoppers' purchase patterns, with a view to identify consuming behaviors. Thereby, ways could be explored of applying such algorithm to customers' paths, in such a way that groups of traces can be identified, substantially easing and completing the analysis conducted on this thesis, as the most can be retrieved from all paths. To that extent, one would firstly need to collect all the events corresponding to a unique visit into one single representation, considering major attributes that identify each visit picture – such as time of visits, items added to the basket, visit time of the day, time spent within the store, frequency of that visit type, and more.

Finally, it is proposed for process discovery to be conducted on "online fashion", that is, that real time data is processed as it occurs. To do so, data must be extracted on-demand into an event log, in such a way that is perfectly structured to be ran on the process mining software and thereby meaningful insights – for instance, deviations from common paths in a particular day or week – can be retrieved faster, and adequate react-to-change strategies can be put in practice.

Conclusion

In essence, the thriving of process mining applications and the explosion of respective successful cases led to the use of this data science practice on discovering customers' journeys at the Pingo Doce & Go Nova grocery store, in attempt to delineate technical improvement guidelines and advancing new management strategies in favor of boosting store sales. Towards this end, subsequent to an extensive, though prudent treating data procedure, a process mining software was used to unveil the customers behavior and experiences throughout the different periods of a day and considering a class-day *versus* non-class-day analysis. Uncovering store technological inefficiencies, shoppers' pain points, and determining possible new products' display approaches, whilst considering the study's challenges and limitations, several recommendations for future steps were presented, with a view to make the most of the aforementioned work.

References

Anand, A., Fosso Wamba, S., Gnanzou, D. 2013. *A Literature Review on Business Process Management, Business Process Reengineering, and Business Process Innovation*. Paper presented at the Workshop on Enterprise and Organizational Modeling and Simulation.

Baier, Thomas, et al. 2018. *Matching events and activities by integrating behavioral aspects and label analysis.* Software & Systems Modelling 17, 573–598. https://link.springer.com/article/10.1007%2Fs10270-017-0603-z.

Bernard, Gael and Andritsos, Periklis. 2017. *A Process Mining Based Model for Customer Journey Mapping*. Paper presented at the International Conference on Advanced Information Systems Engineering, Essen, Germany.

Carter, Matthew. 2019. *Competition and Profit Margins in the Retail Trade Sector*. Reserve Bank of Australia.

Celonis. n. d. "Our Company. We're Celonis.". Accessed December 9th, 2021. https://www.celonis.com/company/.

Celonis. n. d. "What is the Execution Management System". Accessed December 9th, 2021. https://www.celonis.com/ems/what-is-ems/.

Davenport, Thomas H. 1993. *Process Innovation: Reengineering Work Through Information Technology*. Harvard Business School Press.

Dusanka, Dakic ® al. 2019. *Event Log Extraction for the Purpose of Process Mining: A Systematic Literature Review*. Paper presented at the 15th International Symposium in Management, Timisoara, Romania.

Følstad, Asbjørn and Kvale, Knut. 2018. "Customer journeys: a systematic literature review". *Journal of Service Theory and Practice*. https://doi.org/10.1108/JSTP-11-2014-0261.

Følstad, Asbjørn, et al. 2013. *Customer journey measures – State of the art research and best practices.* Norwegian Research Council, Research Report.

Følstad, Asbjørn, et al. 2014. *Customer Journeys: Involving Customers and Internal Resources in the Design and Management of Services.* Paper presented at the ServDes 2014.

Janssenswillen, Gert. 2019. *Unearthing the Real Process Behind the Event Data: The Case for Increased Process Realism*. Hasselt University, Faculty of Business Economics Doctoral Dissertation D/2019/2451/39.

Nadar, Vidhya. 2018. *Just Walk-Out Technology and its Challenges: A Case of Amazon Go*. Paper presented at the 2018 International Conference on Inventive Research in Computing Applications.

Process Mining Book. 2020. "What Is Process Mining?". Accessed November 14th, 2021. https://fluxicon.com/book/read/intro/.

Prolifics. 2020. "Process Mining and Customer Journey". Accessed November 15th, 2021. https://prolifics.com/prolifics-insights/process-mining-and-customer-journey/.

Richardson, A. 2010. "Using Customer Journey Maps to Improve Customer Experience". Harvard Business Review. November 15th, 2010. Accessed November 14th. https://hbr.org/2010/11/using-customer-journey-maps-to.

SIBS. 2018. "3D Secure". Accessed December 11th, 2021. https://www.sibs.com/produtos/3d-secure/.

Schogel, Marcus and Lienhard, Severin. 2020. *Cashierless Stores – the New Way to the Customer?*. University of St. Gallen, Institute of Marketing.

Terragni, Alessandro. 2018. "Analyzing Customer Journey with Process Mining: from Discovery to Recommendations." Master's thesis, Eindhoven University of Technology, Politecnico di Milano.

Van der Aalst, Wil ® al. 2011. *Process Mining Manifesto*. Paper presented at the Process Management Workshops – BPM 2011 International Workshops, Clermont-Ferrand, France.

Van der Aalst, Wil. 2011. *Process Mining: Overview and Opportunities*. ACM Transactions on Management Information Systems. Vol. 99, No 99, Article 99 (February 2012), 16 pages, Eindhoven University of Technology.

Van Zelst, S.J., Mannhardt, F., de Leoni, M. et al. 2020. *Event abstraction in process mining: literature review and taxonomy*. Granul. Comput. 6, 719–736. https://doi.org/10.1007/s41066-020-00226-2.

Werner, Michael ® Gehrke, Nick. 2013. *Process Mining*. WISU – die Zeitschrift für den Wirtschaftsstudenten 7/13, pp 1-16.

Wiriyarattanakul, Thatchanok. 2016. *Combination Usage of Process Mining and Adaptive Case Management*. Stuttgart University, Institute of Architecture of Application Systems Master's Thesis No 0838-012.

Appendix

I. Attributes and Respective Descriptions

Table 2 – Final Attributes and respective Descriptions

Attribute	Description
VISIT_ID	Constitutes the case/visit identifier
EVENT_DATETIME	Date and time of the event
EVENT_WEEKDAY	The day of the week where the event took place
TIME_OF_DAY	Period of the day where the event took place
EVENT_DESC	Description of the event
EVENT_CATEGORY	Category of the event
AREA	Area of the store where the event took place
DIVISION	Division of the store where the event took place
FAMILY	Family of the product related to the event
CATEGORY_AGG	Like EVENT_DESC, but instead of having the product description, it holds the category of the product
SUBCATEGORY_AGG	Like EVENT_DESC, but instead of having the product description, it holds the subcategory of the product

II. Events Description and Explanation

Table 3 – Events Explanation by Category

Event Category	Event Description	Event Explanation	Kept?[4]
App Settings/Profile	Customer registered device for first time	Customer registered device for the first time in the App	No
	Client create profile	Customer created a profile	
	Customer entered app with new device	Customer entered the App with a new device	

[4] Indicates whether the category/event was kept for the analysis of the customer journey.

(Events related to App configurations and customer profiles)	Customer associated a poupa mais card	Customer associated a Poupa Mais (Pingo Doce's) card	
	Client Enable NFC	Customer enabled Near Filed Communication	
	Payment method removed	Customer removed a payment method	
	Client Disable NFC	Customer disabled Near Field Communication	
	Created new payment method	Customer added a new payment method	
	New client registered	New customer registered	
	Staff See Client Profile	Staff observes customer profile	
	Customer enabled PIN	Customer enabled PIN	
	Customer saved PIN	Customer saved PIN	
	Customer disabled PIN	Customer disabled PIN	
	The checkbox 'Is Greenlisted?' was changed from False to True.	Irrelevant	
<u>Add/Remove Products</u> (Events related to the addiction or removal of products to/from the Basket, either by the Customer or by the Staff)	Product	Information about the products that are added or removed from the basket	Yes
<u>Basket Audit</u>	AUDIT CREATED	Audit created	No
	AUDIT PENDING	Audit pending	

(From time to time, a customer is randomly selected, and it's made a valuation of its basket in order to determine the risk of stealing)	AUDIT SKIPPED	Audit was skipped	
	AUDIT OK	Audit passed ok	
	AUDIT DONE NOTIFY	Audit notification done	
	AUDIT NOTIFY OK	Audit notification ok	
	AUDIT NOTIFY SKIPPED	Audit notification skipped	
	AUDIT NOTIFY PENDING	Audit notification pending	
	AUDIT NOT OK	Audit not passed ok	
	AUDIT NOTIFY NOT OK	Audit notification not ok	
	RISK 0	Audit concluded risk-free basket	
	RISK 100	Audit concluded risky basket	
	CALCULATE RISK ERROR 400 Bad Request	Error in calculating risk	
	SEND 'Done TO RISK MANAGEMENT'	Send message 'Audit done' to risk management	
	SEND 'Skipped TO RISK MANAGEMENT'	Send message 'Audit skipped' to risk management	
<u>Basket States</u> (Throughout the journey of the customer in the store, his or her basket will pass through transient states in the system)	Basket in Debt	Basket is in Debt	No
	Basket no shopping	Basket didn't have any items (customer left the store without buying anything)	
	Basket Snapshot Created	A snapshot of the basket was created	
	Basket abandoned at Staff	Basket was eliminated by the staff (happens, for example, when a customer passed all his/her items to a friend or runs out of battery)	
	Basket finalize shopping with error.	Basket payment incomplete due to error	

	Basket finalize shopping with success.	Basket payment complete with success	
	Basket debt in BO	Basket is in debt (the customer left the store without paying one or more items)	
	Basket transferred from App to POS	Basket payment was firstly to be made in the App, but it was transferred to the POS	
	Basket timed ou in POS	Basket payment in POS timed out	
<u>Check In/Out</u> (Marks the entries and exits of the customers in the store)	IN	Customer entered or tried to enter the store	Yes
	OUT	Customer left or trying to leave the store	
<u>Payment</u> (All events related to the in-store payments of baskets)	PAID in App	Payment successful in App	Yes
	PAID at POS	Payment successful at POS (Point-of-Sales)	
	PAID at BO	Payment successful at BO	
	Client Receipt Received	Customer received receipt (payment successful)	
	Transaction Rejected	Payment was rejected	
	Requires 3DS	Payment requires passing through 3D Secure authentication[5]	
	Fiscalization failure	Failure at fiscalization in payment	
<u>Go 24/7</u>	BuyBye open	Go 24/7 doors opened	No

[5] 3D Secure is a security protocol which authenticates the owner of the card using for electronic payments, ensuring that the payment is indeed been made by him/her and, therefore, his/her protection.

(Events related to the transactions of the customers with the Go 24/7 closet)	BuyBye basket processed	Go 24/7 basket processed	
	BuyBye Basket Snapshot Created	A snapshot of the buybye basket was created	
	BuyBye empty basket	Go 24/7 basket has no items	
	BuyBye payment done	Payment successful at BuyBye	
	BuyBye POS Finalize Basket Failed	Payment rejected	
<u>Other</u> (All events that didn't match all the other categories)	Coffee Machine Associated To User	Coffee machine was associated to the customer	No
	CALLOUT	When a customer wants a product from the pharmacy	
	Customer Export Data	Customer exported his or her data from the app	
	Error in PingoDoceAndGoNOVA App	An error occurred on the App	
	Basket Valuation	Happens every time a product is added or removed from the basket;	
	New Complaint	Customer made a compliant	
	Call to support	A call for support was made	
	Basket set as error – Sanitization Process	At the end of the day, a process runs and every unfinished baskets	
	ADD TO QUEUE AMERICANO, ADD TO QUEUE EXPRESSO DUPLO, ADD TO QUEUE CAPUCCINO, …	These events are related to the coffee machine. The system is queuing the beverages that will serve next.	

III. Detailed Data Curation Steps

Table 4 – Data Curation Steps

Step	Description	Motive
1	Ran a function which identifies those columns with missing values, and dropped the rows where no customer was associated	Customer-centric analysis, therefore the focus is on records linked to customers
2	With the aid of the customer look-up table, kept only active and normal accounts using the following attributes and criteria: a) CUSTOMER_STATUS_ID = 1 and IS_ACTIVE = 0: customer was blocked in backoffice, and he or she deleted his/her own profile (4 cases) b) CUSTOMER_STATUS_ID = 1 and IS_ACTIVE = 0: customer was blocked in backoffice (18 cases) c) CUSTOMER_STATUS_ID = 2 and IS_ACTIVE = 0: customer had no problem in backoffice, but he or she deleted his/her own profile (169 examples) d) CUSTOMER_STATUS_ID = 2 and IS_ACTIVE = 1: normal customers (31622 cases) e) CUSTOMER_STATUS_ID = 3 and IS_ACTIVE = 0: customers' profile was deleted by the customer himself/herself and in backoffice as well (236 cases) f) It is not possible to have CUSTOMER_STATUS_ID = 3 and IS_ACTIVE = 1 customers, since when the customer deletes his/her profile, it immediately turns his/her status to 3	Being a customer-centric analysis, only customers who did not have any problem with the store and did not delete their accounts were kept for the analysis. All others were perceived as shoppers with no intentions to return or prohibited from returning to the store.
3	For example, records related to the customers' payment acts within the retailer were stored under a "Payment" category.	

4	Cleaning of the EVENT_DESC attribute as follows: remove any additional information already on other columns and preserve other information when it was not, keeping the core descriptions only. Examples: a) “Requires 3DS 2021-10-30 19:47:17” became “Requires 3DS” b) “ADD MASSA PERSONALIZADA 4 SCAN 2021-10-30 20:34:28 – SAP 000000000000894334” became “MASSA PERSONALIZADA 4”, and, on column SAP_CDE, “894334” c) “PAID in App 2021-10-30 20:55:46” became “PAID in App” d) “IN GATE F 2021-10-30 21:15:35” became “IN” e) … Some events did not require cleaning up: a) “Client Enable NFC” b) “Basket abandoned at Staff” c) …	Not only the event column becomes much cleaner and harmonized, easing event comprehension and perception of the event log, but it was also needed to have the products description to merge the event log with the products data.
5	Using specific product keywords from the log data, the goal was to find the corresponding item – or the most similar one – on the article look-up table, and then input the respective products' information on the main dataset. When no similar or equal product was found, the corresponding event was dropped (19 events).	These products were considered either errors or almost-never-bought products, and therefore represented such unique paths that were of no interest
6	Kept only events directly related to the shoppers' action-generated touchpoints, and, therefore, to their journey: removed “Other”, “Basket States”, “Basket Audit”, “Go 24/7”, and “App Settings/Profile”	Keeping these events would not be relevant for the analysis, and would increase the complexity of the customer journey maps

7	Cleaned wrongful entries and exits	The cleaner the data, the better the shoppers' paths can be conveyed
8	Removed the duplicate rows that resulted from the merging of the event log with the products look-up table by identifying the products that had one description for two different codes, look for the rows corresponding to those products, and remove the duplicate lines	These duplicate rows were wrongly created and therefore needed to be removed
9	Removed outliers and erroneous records on the data	The less noise and erroneous data points in the log, the better one can represent cleaner customers' journeys

IV. Snapshots of Final and Intermediary Tables

Table 5 - Raw Data

Criado por	Cliente	Customer Country Code	Criado em	Tipo de Evento	Evento	Descrição do evento	Session UUID	Basket Id	Basket GUID
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:55:56	CheckOutRequestAllowed	Check out allowed	OUT GATE E 2021-10-30 20:55:56			
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:55:51	ClientReceiptReceived	Client receipt received from POS	Client Receipt Received		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
System	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:50	ClientReceiptReceived	Client receipt received from POS	Client Receipt Received		614776	5e1777b2-9ac4-45da-bc8d-2bccbc872a4e
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:55:46	BasketPaidInApp	Payment successful in App	PAID in App 2021-10-30 20:55:46		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
System	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:45	BasketPaidInApp	Payment successful in App	PAID in App 2021-10-30 20:55:45		614776	5e1777b2-9ac4-45da-bc8d-2bccbc872a4e
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:55:44	BasketCalculateRisk	Basket Calculate Risk	RISK 0		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
User	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:27	CustomerAppFrontEndError	Pingo Doce And Go NOVA Error	An errored in the front end of the PingoDoceAndGoNO'			
System	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:23	BasketValuation	Basket valuation	Basket Valuation		614776	5e1777b2-9ac4-45da-bc8d-2bccbc872a4e
System	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:22	ClientAddProductScan	Client Add Product Scan	ADD GOMAS REGALIZ PICA VIDAL 100 GR SCAN 202		614776	5e1777b2-9ac4-45da-bc8d-2bccbc872a4e
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:55:17	BasketFinalizeShopping	Basket Finalize Shopping	Basket finalize shopping with success.		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
User	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:55:11	CustomerAppFrontEndError	Pingo Doce And Go NOVA Error	An errored in the front end of the PingoDoceAndGoNO'			
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:54:53	BasketValuation	Basket valuation	Basket Valuation		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:54:52	ClientAddProductScan	Client Add Product Scan	ADD GOMAS TALHADAS DE MELANCIA 100GRS SCA		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:54:49	BasketValuation	Basket valuation	Basket Valuation		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:54:48	ClientAddProductScan	Client Add Product Scan	ADD SNACK APER MATUTANO DORITOS TEX MEX 12		614777	8d380747-37ce-400e-900d-c5e7eefab1aa
User	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:54:25	CustomerAppFrontEndError	Pingo Doce And Go NOVA Error	An errored in the front end of the PingoDoceAndGoNO'			
System	9eee61e4-d43d-423d-9bf3-da97a4acc834	'+351	2021-10-30 20:53:42	CheckInRequestAllowed	Check in allowed	IN GATE G 2021-10-30 20:53:42			
System	0550e0b7-00b8-43eb-b138-a921abd67bb9	'+351	2021-10-30 20:53:35	CheckInRequestAllowed	Check in allowed	IN GATE F 2021-10-30 20:53:35			

Table 6 - Final Data

VISIT_ID	CLIENT_NATIONALITY	EVENT_DATETIME	EVENT_WEEKDAY	TIME_OF_DAY	EVENT_DESC	EVENT_CATEGORY	AREA	DIVISION	FAMILY	CATEGORY_AGG	SUBCATEGORY_AGG
1	Portuguese	2021-10-07 15:19:56	Thursday	Afternoon	Check in allowed	Check In/Out	Entrace	Entrace	Entrace	Check in allowed	Check in allowed
1	Portuguese	2021-10-07 15:23:41	Thursday	Afternoon	Check out allowed	Check In/Out	Exit	Exit	Exit	Check out allowed	Check out allowed
2	Portuguese	2021-10-11 15:14:38	Monday	Afternoon	Check in allowed	Check In/Out	Entrace	Entrace	Entrace	Check in allowed	Check in allowed
2	Portuguese	2021-10-11 15:18:58	Monday	Afternoon	Add Coffee Machine Products	Add/Remove Products	PERECIVEIS ESPECIALIZADOS	PADARIA/PASTELARIA	CAFE E BOLOS	Add BEBIDAS	Add CAFE
2	Portuguese	2021-10-11 15:20:47	Monday	Afternoon	Check out denied	Check In/Out	Exit	Exit	Exit	Check out denied	Check out denied
2	Portuguese	2021-10-11 15:20:58	Monday	Afternoon	Check out denied	Check In/Out	Exit	Exit	Exit	Check out denied	Check out denied
2	Portuguese	2021-10-11 15:23:11	Monday	Afternoon	Check out allowed	Check In/Out	Exit	Exit	Exit	Check out allowed	Check out allowed
3	Portuguese	2021-10-11 17:54:34	Monday	Afternoon	Check in allowed	Check In/Out	Entrace	Entrace	Entrace	Check in allowed	Check in allowed
3	Portuguese	2021-10-11 18:00:59	Monday	Afternoon	Add Coffee Machine Products	Add/Remove Products	PERECIVEIS ESPECIALIZADOS	PADARIA/PASTELARIA	CAFE E BOLOS	Add BEBIDAS	Add CAFE
3	Portuguese	2021-10-11 18:03:17	Monday	Afternoon	Add Coffee Machine Products	Add/Remove Products	PERECIVEIS ESPECIALIZADOS	PADARIA/PASTELARIA	CAFE E BOLOS	Add BEBIDAS	Add CAFE
3	Portuguese	2021-10-11 18:05:01	Monday	Afternoon	Check out allowed	Check In/Out	Exit	Exit	Exit	Check out allowed	Check out allowed
4	Portuguese	2021-10-12 08:52:45	Tuesday	Morning	Check in allowed	Check In/Out	Entrace	Entrace	Entrace	Check in allowed	Check in allowed
4	Portuguese	2021-10-12 08:55:33	Tuesday	Morning	Add Coffee Machine Products	Add/Remove Products	PERECIVEIS ESPECIALIZADOS	PADARIA/PASTELARIA	CAFE E BOLOS	Add BEBIDAS	Add CAFE
4	Portuguese	2021-10-12 08:57:16	Tuesday	Morning	Check out allowed	Check In/Out	Exit	Exit	Exit	Check out allowed	Check out allowed

V. Visualizations for Support of the Exploratory Data Analysis

Period of the Day	Time Horizon	Class Breaks (Approx. Peaks)
Morning	Entries until 12 a.m.	9.15 and 10.45 a.m.
Lunchtime	Entries from 12 p.m. to 3 p.m.	12.15 and 2.15 p.m.
Afternoon	Entries from 3 p.m. to 7.30 p.m.	3.45, 5.15 and 6.45 p.m.
Dinnertime	Entries from 7.30 p.m.	8.15 p.m.

Table 7 - Time Horizons for each period of the day and included Visits Peaks

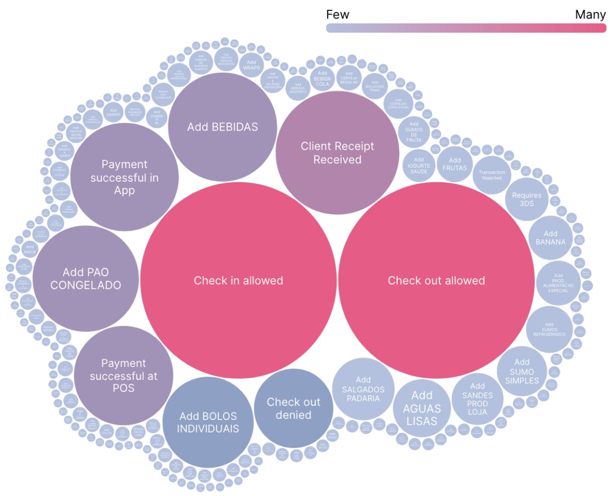

Figure 14 - Frequency of Events in the Morning, on Class Days

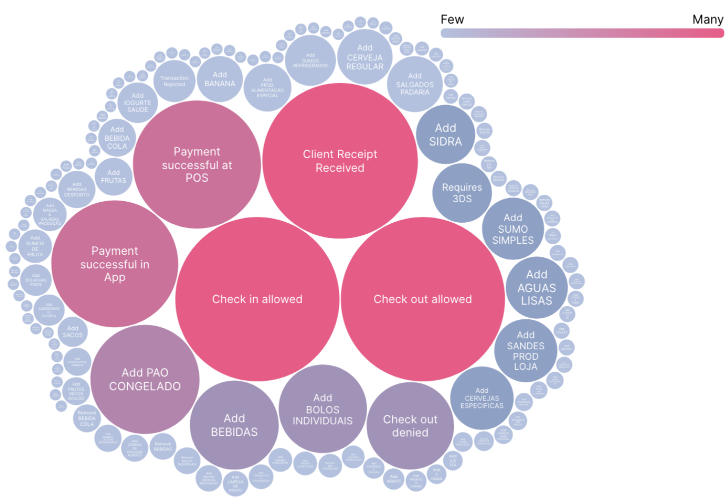

Figure 15 - Frequency of Events in the Morning, on Saturdays

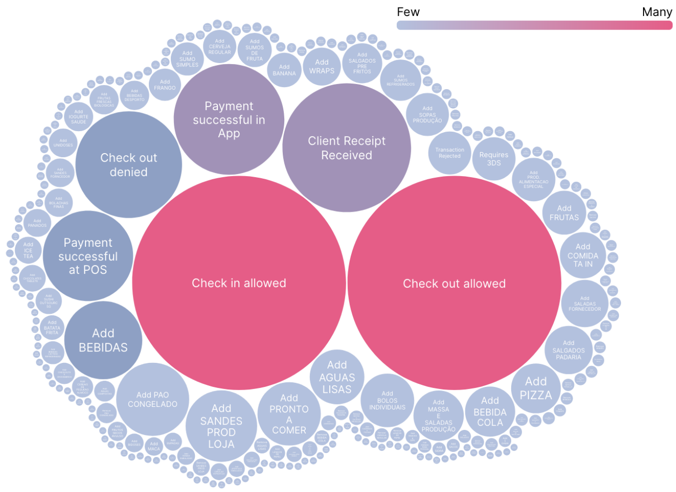

Figure 16 - Frequency of Events at Lunchtime, on Class Days

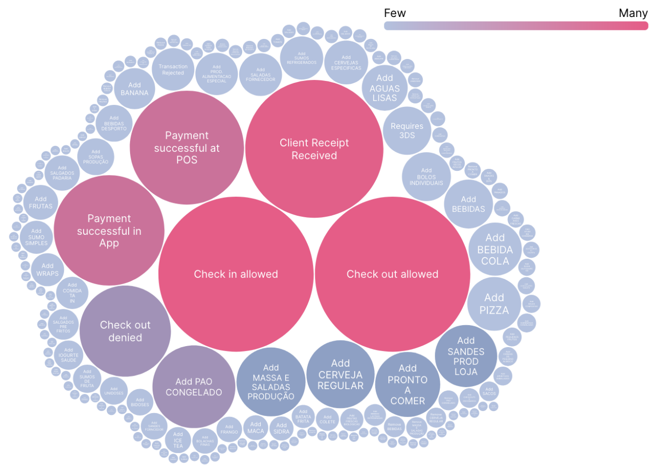

Figure 17 - Frequency of Events at Lunchtime, on Saturdays

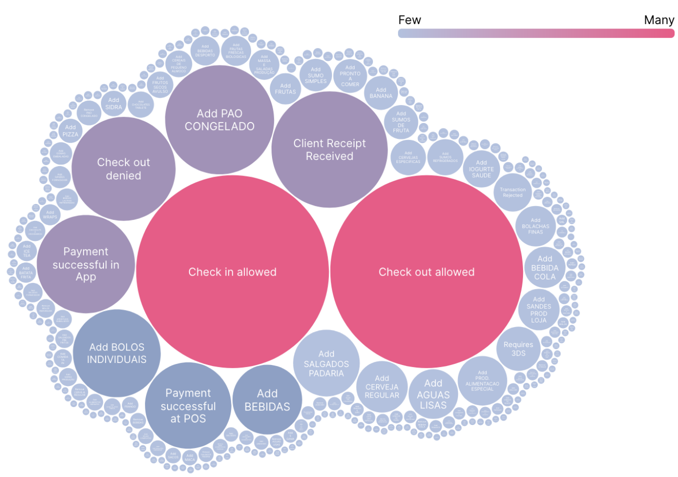

Figure 18 - Frequency of Events in the Afternoon, on Class Days

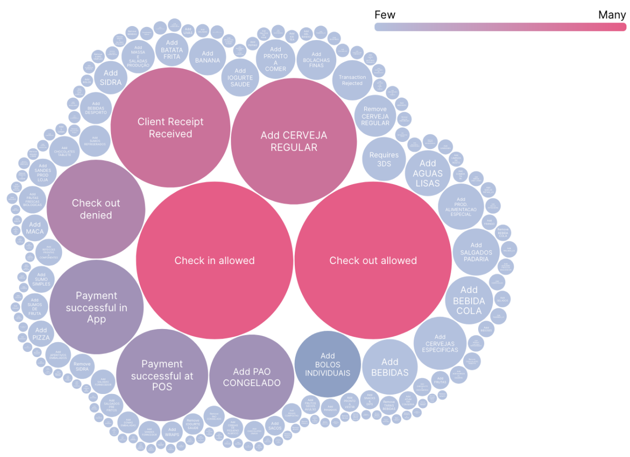

Figure 19 - Frequency of Events in the Afternoon, on Saturdays

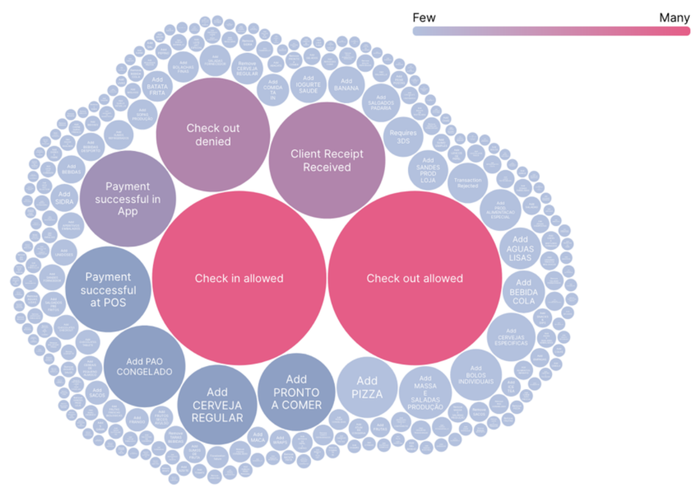

Figure 20 - Frequency of Events at Dinnertime, on Class Days

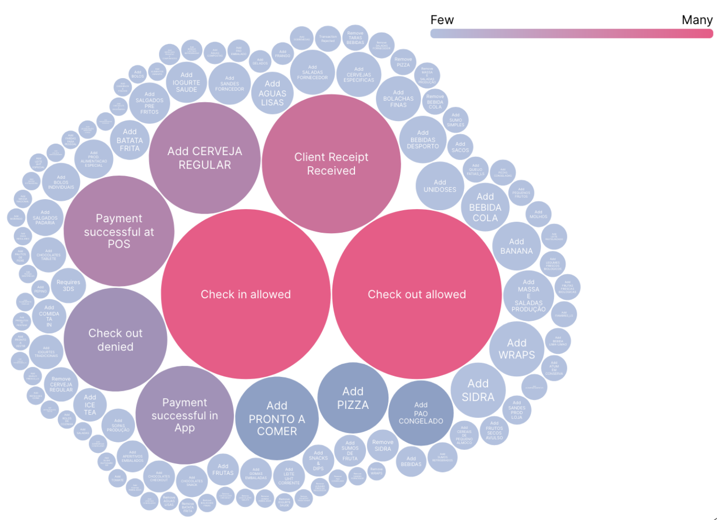

Figure 21 - Frequency of Events at Dinnertime, on Saturdays

VI. Complementary Event-log Visualizations

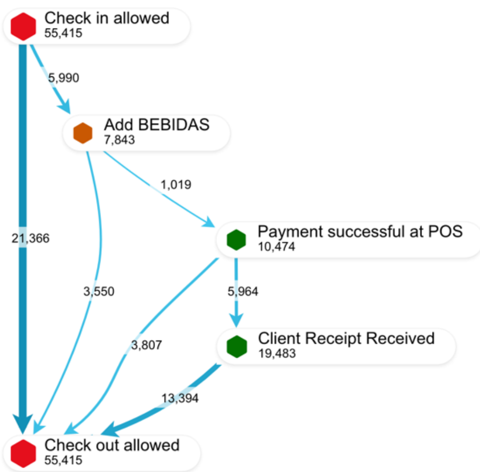

Figure 23 - Example Visualization of the Event log (Visit Frequency View)

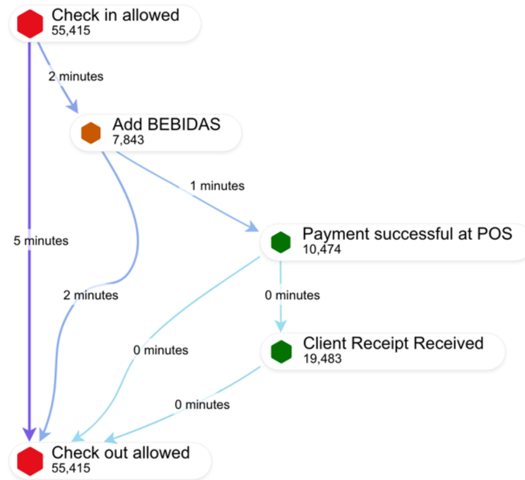

Figure 22 - Example Visualization of the Event log (Throughout Time Average View)

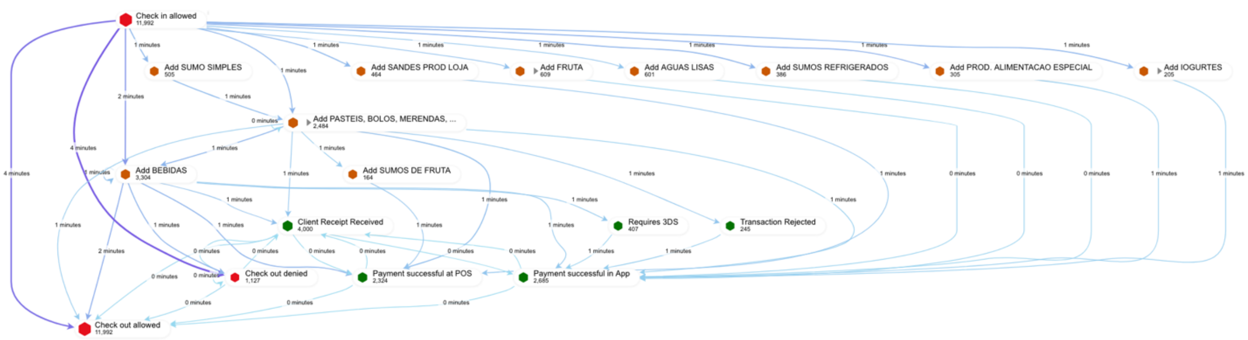

Figure 24 - Customer Journeys in the Morning (Throughout Time Average View)

[93.4% Activities; 73.7% Connections]

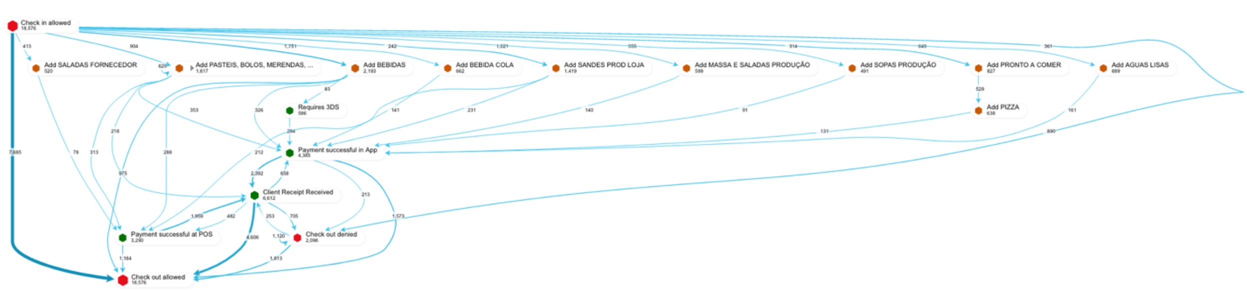

Figure 25 - Customer Journeys at Lunchtime, on Class Days (Visit Frequency View)

[88.2% Activities; 65.5% Connections]

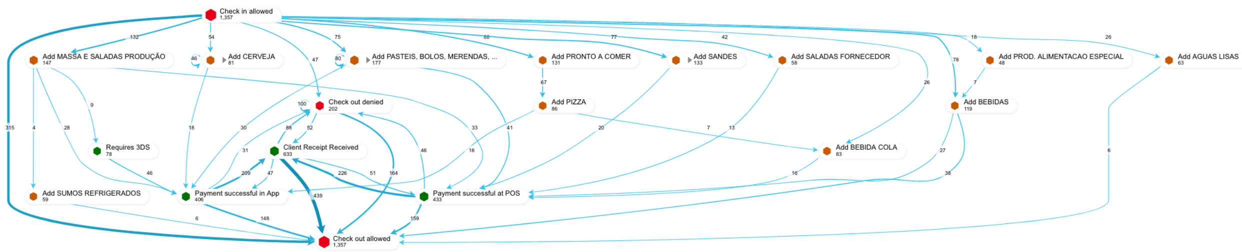

Figure 26 - Customer Journeys at Lunchtime, on Saturdays (Visit Frequency View)

[86.4% Activities; 59.6% Connections]

Figure 27 - Customer Journeys in the Afternoon, on Class Days (Visit Frequency View)

[90.6% Activities; 69.6% Connections]

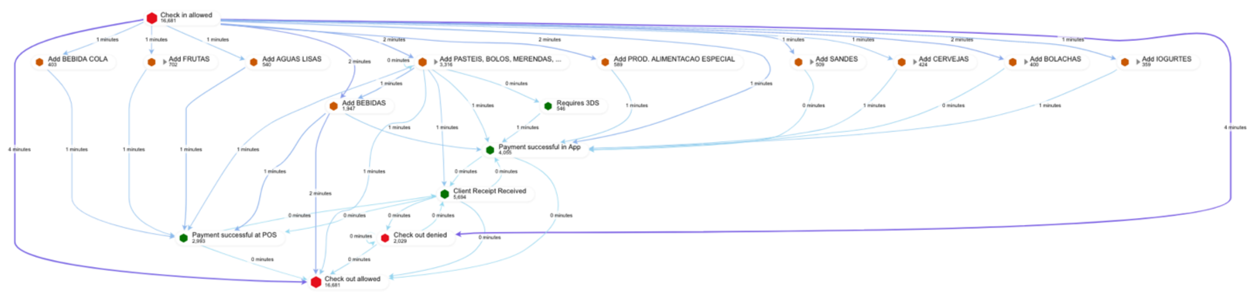

Figure 28 - Customer Journeys in the Afternoon, on Class Days (Throughout Time Average View)

[90.6% Activities; 69.6% Connections]

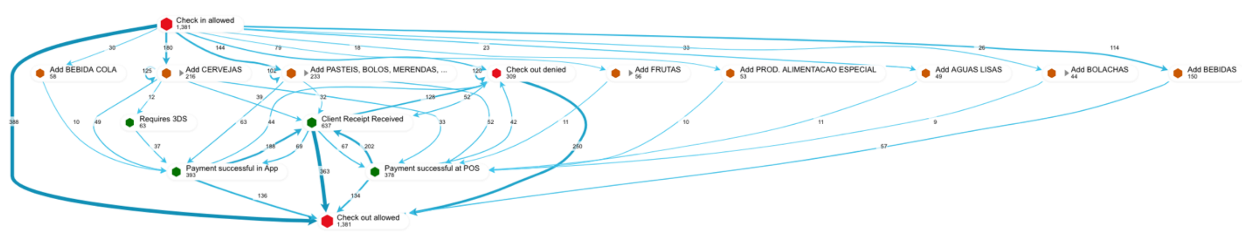

Figure 29 - Customer Journeys in the Afternoon, on Saturdays (Visit Frequency View)

[85.7% Activities; 64.3% Connections]

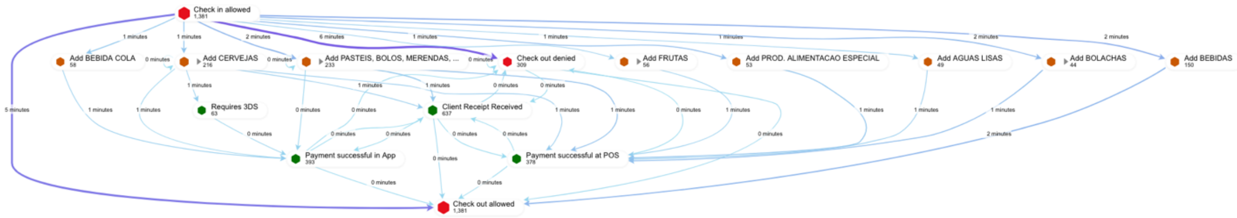

Figure 30 - Customer Journeys in the Afternoon, on Saturdays (Throughout Time Average View)

[85.7% Activities; 64.3% Connections]

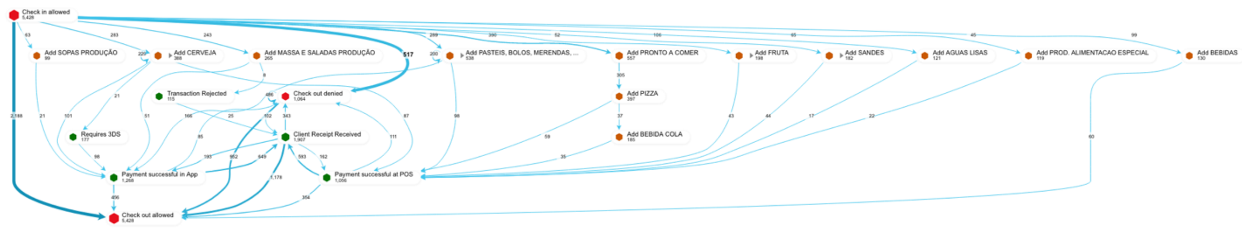

Figure 31 - Customer Journeys at Dinnertime (Visit Frequency View)

[88.1% Activities, 64.5% Connections]

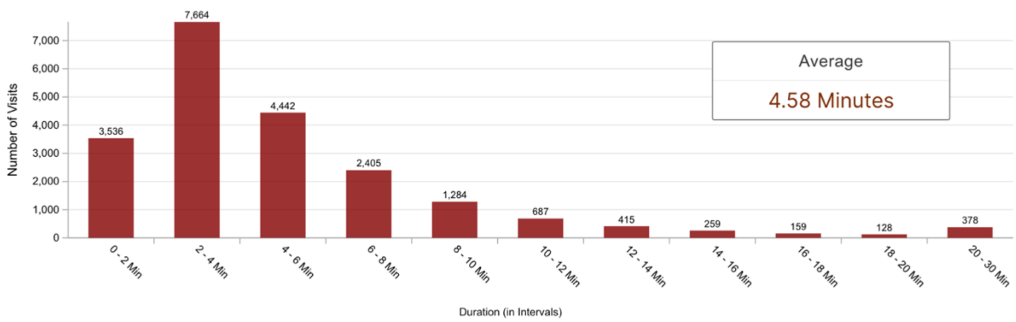

Figure 32 - Distribution of Visits' Duration, considering Visits with no Middle Events

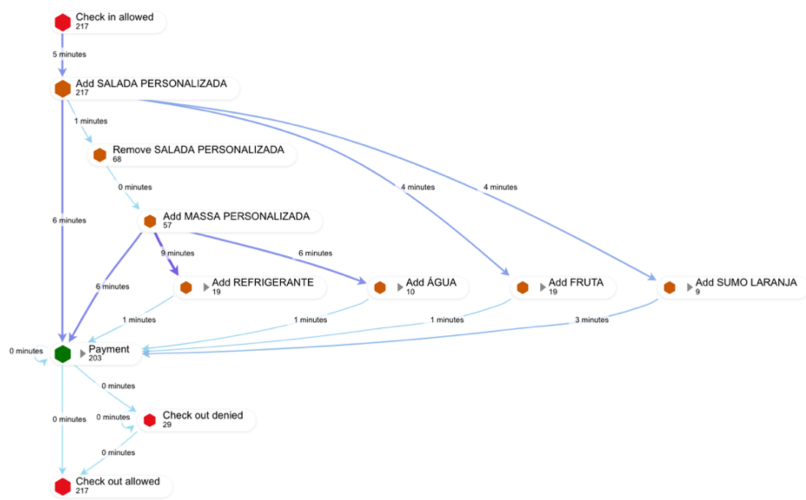

Figure 33 - Customer Journeys including Personalized Salads (Throughout Time Average View)

[88.9% Activities; 63.9% Connections]

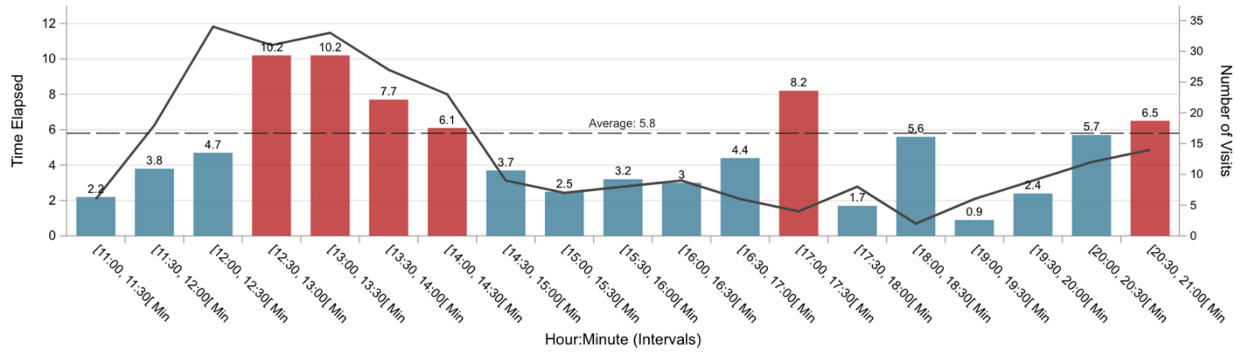

Figure 34 - Distribution of Ordering Times for Personalized Salads, by Hour:Minute Interval (Class Days)

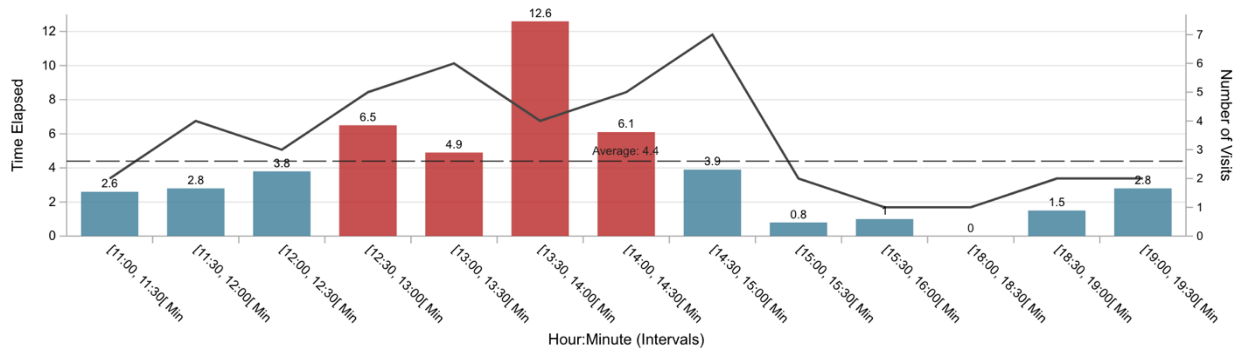

Figure 35 - Distribution of Ordering Times for Personalized Salads, by Hour:Minute Interval (Saturdays)

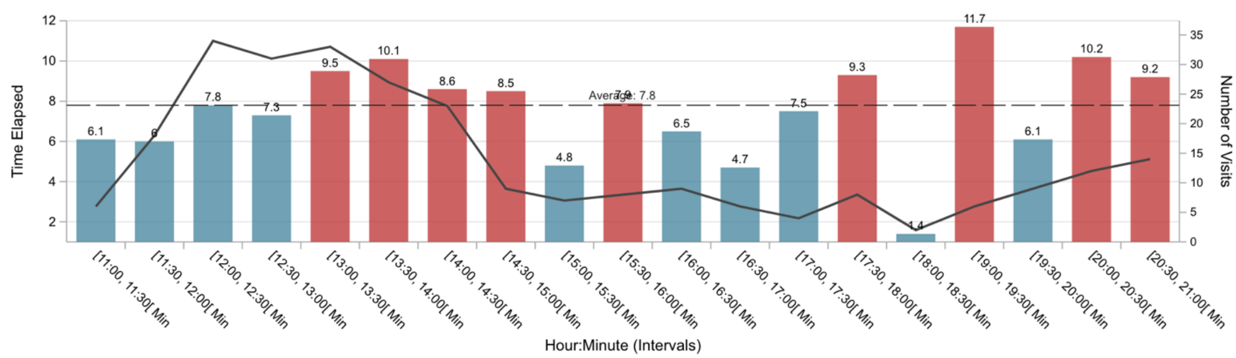

Figure 36 - Distribution of Waiting Times for Personalized Salads, by Hour:Minute (Class Days)

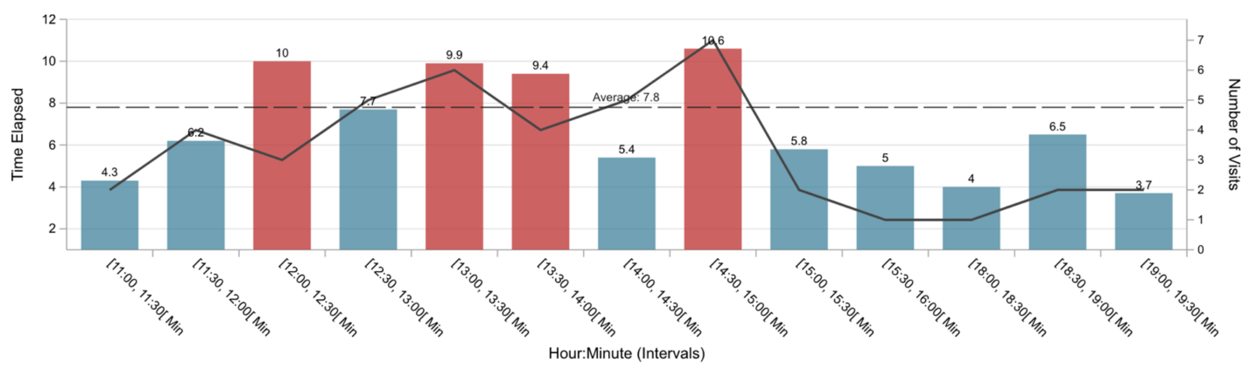

Figure 37 - Distribution of Waiting Times for Personalized Salads, by Hour:Minute (Weekends)