Master's Degree Program in
Data Science and Advanced Analytics

Business Process Management

Assignment 2: Saúde+ Health Center Process Analysis

Group E

Elcano Gaspar, 20241021

Rui Ten Jua, 20240854

Rute Teixeira, 20240667

Santiago Ravara, 20230431

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

June, 2025

Table of Contents

1. Introduction 2
2. ETL Process: Data Preparation and Main Insights 2
   2.1. Data Sources Overview 2
   2.2. Data Integration and Model Definition 3
   2.3. Data Transformations 4
3. Dashboarding pages 5
   3.1. Variant Explorer 5
   3.2. Process Explorer 6
   3.3. Case Explorer 7
   3.4. Business Analysis 8
   3.5. Regional Trends 9
   3.6. Patient Segmentation and Satisfaction 11
   3.7. Time Duration Analysis 12
4. Conclusion 13

1. Introduction

This report presents a data-driven analysis of the Patient Appointment Process at Saúde+ Health Center, leveraging process mining techniques. The primary objective is to gain transparency into the "As-Is" process, identify key inefficiencies, quantify negative outcomes such as patient dissatisfaction and extended wait times, and provide actionable insights for improvement.

To tackle this, we applied process mining techniques using *Celonis*, working with real patient-level event logs. This allowed us to:

- Reconstruct the AS-IS process of patient appointments.
- Identify variability and inefficiencies.
- Analyze performance metrics like throughput time, cost and cancellations.
- Explore behavior by patient segments and regions such as insurance status and chronic conditions.

The results provide insights into where time is lost, where patient satisfaction might be impacted, and what parts of the process are most prone to rework or cancellation.

The primary goal of this solution is to support business users in operational management and decision-making based on meaningful analytics that provide clear and actionable insights into Saúde+ Health Center performance.

2. ETL Process: Data Preparation and Main Insights

2.1. Data Sources Overview

The analysis is based on two primary data tables, provided during handout by our client:

- Event Log (Event_Log_Assignment_2.csv): This table contains the sequential records of every event that occurred during a patient appointment. It includes:
    - case id: Unique identifier for each appointment instance.
    - timestamp: The exact date and time an activity was recorded.
    - activity: The name of the specific action or step performed (e.g., "*Check-in at Reception*", "*Orthopedic Consultation*").
- Case Table (Case_table_Assignment_2.csv): This table provides contextual attributes for each unique appointment case. It includes:
    - case id: Matches the case id in the Event Log, allowing for linkage.
    - Patient_age, Patient_gender, Patient_region: Demographic information.
    - Has insurance: Boolean indicating insurance status.
    - Total cost: The total cost associated with the appointment.
    - Chronic condition: Information about any chronic conditions the patient may have.
    - Satisfaction score: Patient's satisfaction rating (0-5).

- Num_previous_appointments: Number of previous appointments in the last 3 months.
- Cancellation_history: Number of cancelled appointments in the last 3 months.

2.2. Data Integration and Model Definition

To enable the analysis of the patient appointment process, we followed a structured procedure in *Celonis* to configure our workspace, integrate the datasets, and prepare them for use in Studio. Below is a step-by-step breakdown of the technical setup:

1. Workspace Configuration and User Access

To begin, user access was configured via:

- Admin Settings > Users > Invite Users, followed by email invitation acceptance. Our professor was also added to ensure appropriate access for evaluation purposes.

2. File Import and Data Integration

We imported our datasets by navigating to:

- Data > Data Integration > Create from Scratch

A name was assigned to the data pool, and files were uploaded via the "Connect to Data Source" option. Once imported, this formed the foundation of our analysis. The Data Pool was named "*Assignment_2*".

3. Data Transformation

In cases where source data required alteration, such as joining tables, cleaning columns, or enriching records, we used the Data Transformation functionality. This step will be further examined in the next section of the report.

4. Data Model Creation

With the data sources ready, we created a data model by:

- Going to Data Pools > Create Data Model, naming it "DM", and selecting all tables.

During configuration, we assigned the primary keys, activity and timestamps, required for process mining. A foreign key was added to the Event Log to address the presence of multiple *CASE_ID*s, while a unique identifier was defined for the Case Table.

5. Setting the Event and Case Tables

In the Data Model Canvas:

- We assigned “Event_Log_Assignment_2” as the activity (event log) table.
- We designated “Case_Table_Assignment_2” as the default case table using the *Celonis* built-in functionality, ensuring no autogenerated table was created.

6. Loading the Data Model

To conclude, we executed a Full Load:

- In Data Pool > Assignment_2, we selected Execute Load > Full Load.
- Then, in Data Model > Data Loads, we triggered Load Data Model to finalize data availability for Studio.

7. Studio Preparation

To begin the analysis interface setup, we went to Studio, created a new Space, and selected our “DM” data model. Finally, we created a Package containing our visualizations and metrics, with a name and description provided in View Mode.

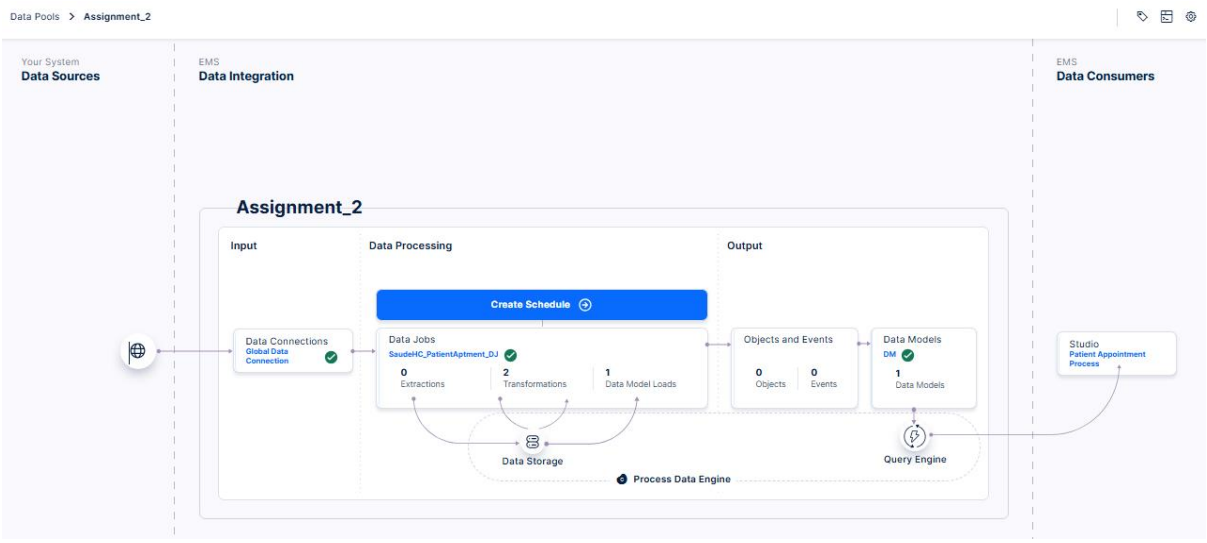


Figure 1: Data Model and Integration Overview

2.3. Data Transformations

To ensure the quality and enrichment of our data, we utilize data transformations whenever necessary. This process involved selecting “Add Data Transformation”, assigning a meaningful name, writing the appropriate transformation logic, and saving the script. One standard transformation applied was the deduplication of the Case Table, performed to eliminate duplicate records and ensure data consistency for accurate analysis.

Although this process is inherently iterative, meant to be revisited as data refinement needs emerge, our ability to fully leverage this functionality was limited by the constraints of the *Celonis* Academic account. As a result, some transformations, such as the one titled “Duration”, were created but remained disabled, restricting our ability to enrich the data as initially intended.

3. Dashboarding pages

Following the data ingestion, setup and loading of our data model, a *Celonis* dashboard was developed in Studio, to deliver a consolidated business analysis for Saúde+ Health Center’s overall process performance and key operational indicators.

The *Celonis* dashboard is organized into several key pages, each designed to provide a distinct view and facilitate comprehensive process analysis. As a foundation, we focused on a global overview that enables exploration by case, process and variant, along with a business overview with key indicators. However, the analysis was expanded to patient segmentation, regional trends and Time Duration analysis, for a more complete study. With scalability in mind, Saúde+ HC managers can expand the analysis with new perspectives as business needs evolve in the coming years.

The resulting analysis is presented in an eight-page interactive report. Each page adheres to a consistent color scheme of purple tones, ensuring a cohesive and visually appealing layout. We will go through each page for a more detailed examination over the following sections of the report.

3.1. Variant Explorer

The Variant Explorer page is designed to provide a transparent view of all the unique sequences of activities (variants) that patients follow during their appointment journey. It visualizes these different paths, allowing us to identify the most common flows, discover deviations, and understand their frequency and performance characteristics.

Main Insights:
The Variant Explorer reveals the diversity of paths patients take within the appointment process, highlighting critical areas for investigation:

- **Variant #1**: which accounts for a substantial 23% of all cases but shows an extremely short average throughput time of just 6 minutes. This path typically involves a "Check-in at Reception" followed immediately by "Return Another Day," suggesting that patients are being redirected without completing a full consultation. This pattern may be the result of overbooking, limited physician availability, triage decisions, or administrative delays. Due to its high frequency and low completion, this variant represents a potential efficiency and satisfaction concern, and further investigation is needed to identify its root causes and mitigate its impact on both patients and staff resources.
- **Variants #3 to #10:** represent longer-duration paths, each contributing to 2% to 5% of the total cases. These variants span between 36 and 58 minutes on average and are likely to correspond to full appointment workflows. However, the variability in duration hints at

underlying inefficiencies such as unnecessary steps, repeated activities, or long wait times between process steps. These can be considered "slow lane" flows that warrant closer examination to streamline patient journeys and eliminate rework.

- **"Others" Category:** although accounting for a smaller share of 3%, shows the highest average throughput time, exceeding 60 minutes. This group likely includes exceptional or highly complex cases, such as those requiring multiple consultations, extensive administrative handling, or unresolved issues that force re-engagement. Despite their low frequency, these cases can consume significant operational effort. As such, drilling into individual cases in this category can help uncover opportunities for process automation, escalation protocols, or dedicated case-handling workflows to prevent these inefficiencies from affecting overall performance.

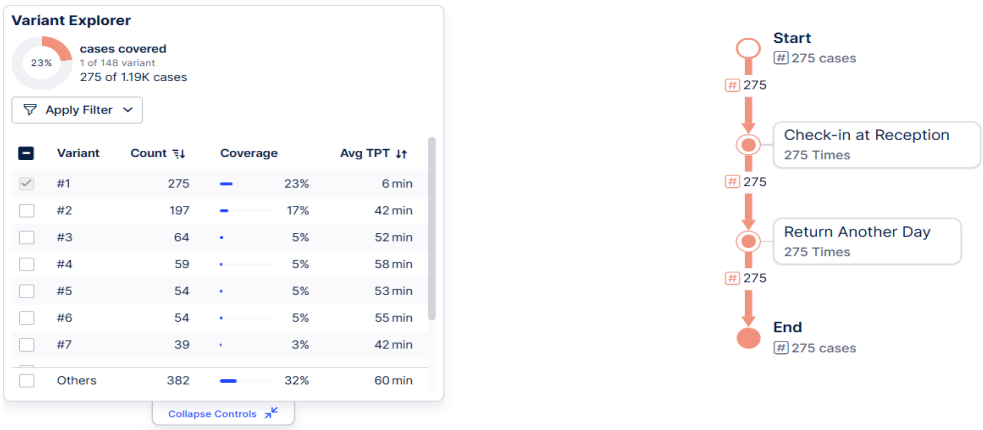
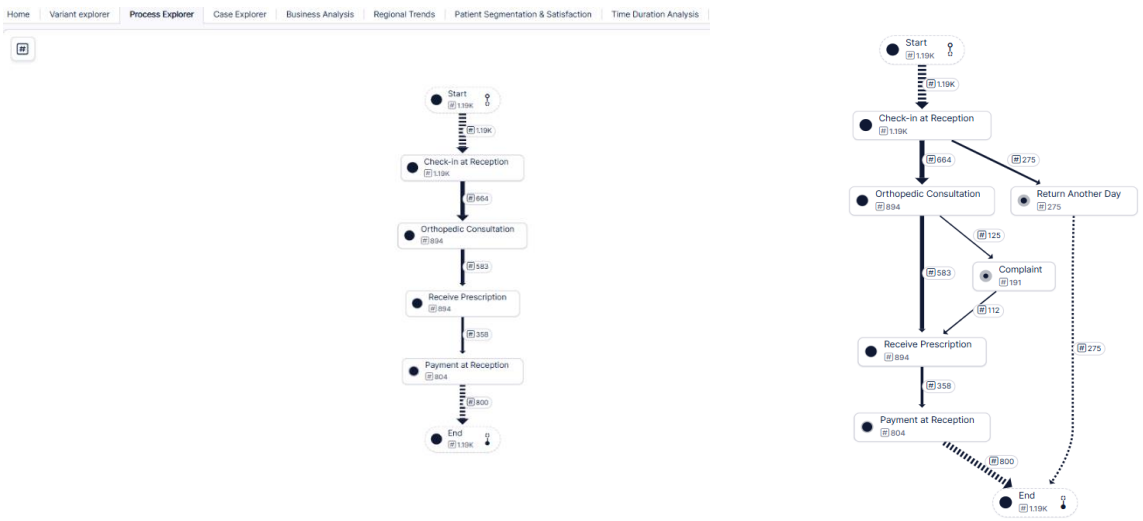


Figure 2: Variant Explorer - Most Frequent Process Flow

3.2. Process Explorer

The Process Explorer represents a standard patient flow in the health center clinic, beginning with check-in at reception, followed by an orthopedic consultation, receiving a prescription, making a payment at reception, and ending the visit. Most patients follow this path from Start to End.



U

Figure 3:Process Explorer - End-to-End Patient Journey

bout

Figure 4: Process Explorer - Detailed Process Flow with Deviations

the process flows with deviations that highlight operational inefficiencies and patient behavior trends.

Several significant deviations from this main path are worth noting. 23% of patients leave after checking in by opting to "Return Another Day". This early drop-off may point to issues such as long wait times, lack of appointment availability. Addressing this issue could significantly improve the customer experience.

Another deviation occurs during or after orthopedic consultation. A total of 191 patients filed complaints. They may indicate systemic issues within the consultation experience, such as perceived quality of care and communication problems. These concerns need further due diligence to maintain patient trust and satisfaction.

Other bottlenecks are notable, only 664 cases move directly from check-in to consultation, and just 583 continue from consultation to prescription.

The last activity pair of activities that are payment and exit are highly efficient. With 804 cases passing smoothly through payment to end and only a small number of exceptions, this phase of the process certainly is the most reliable. This reflects positively on the center's billing procedure.

These figures suggest potential process delays or patient-specific complexities that could be addressed by better pre-visit preparation.

3.3. Case Explorer

The Case Explorer is designed to analyze each individual case in detail, particularly to identify which activities the patient completed during a visit.

Through the Case Explorer, it is possible, for example, to observe, by sorting cases by consultation time, that appointment durations range from a minimum of 3 minutes to a maximum of 2 hours. While the minimum duration may suggest a quick interaction, it is often an indicator of dissatisfaction, as it typically corresponds to outcomes such as "Return Another Day" or "Missed Turn".

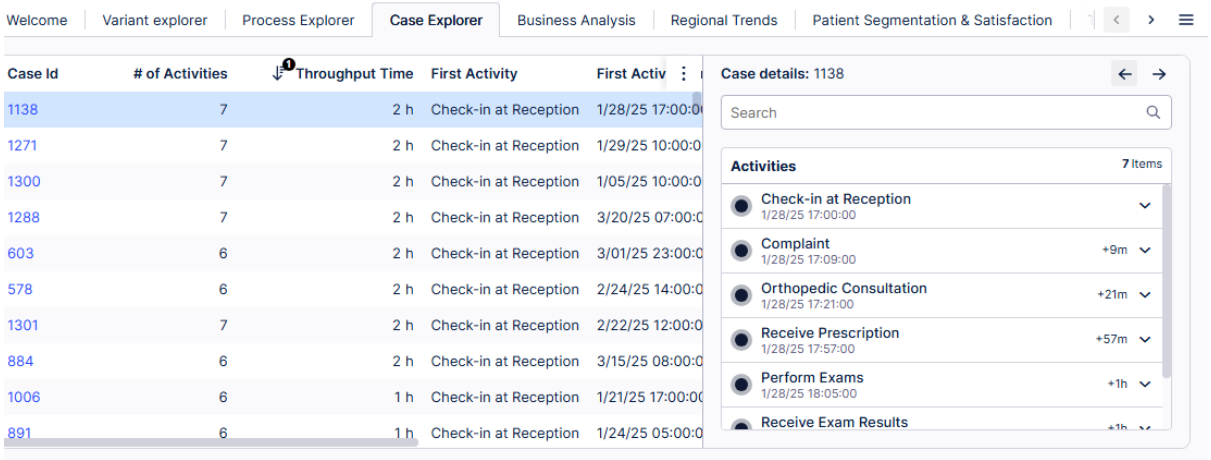


Figure 5: Case Explorer - Individual Case Details and Activity Log

3.4. Business Analysis

The fourth page of our report provides a summarized business overview, highlighting key performance indicators (KPIs) that either hinder or enhance the efficiency of the appointment process and overall service quality.

The layout of this page can be described by the display of 2 KPI lists on top. The first list aggregates values by average:

- average wait time from check-in to consultation (9 min)
- average cost of an appointment (61,35€)
- average satisfaction score (3.59).

These figures reveal that the in-person waiting time is notably low, especially when compared to the 27 min average reported in UK studies for the same specialty. This KPI appears in a green tone, based on the color mapping applied using that benchmark as a threshold. Conversely, the satisfaction score indicates room for improvement, falling short of the target value of 4 or higher, and is therefore displayed in orange tones. It is concerning that even with a low in-person waiting time, the perceived service quality is still below excellent scores. This disconnect suggests that delays alone aren't the issue- instead, inefficiencies may stem from the quality of consultation itself, staff-patient interactions or even how appointments are managed.

Moving to the second KPI list, numbers are presented in totals:

- total number of appointments (1190) - the basis of our analysis
- total cancellation history (611)
- total insured appointments by percentage (43%)
- total complaints during the process (16%).

The complaint rate is considerably high, especially when compared to the national benchmark of 6-8% official complaints for orthopedic services in Portugal, as reported by *Portal da Queixa.* As a result, this KPI is displayed in an orange tone, following the color mapping configured using this benchmark as the threshold, helping to visually highlight performance areas in need of attention.

Immediately below, a combined column and line chart visualizes the distribution of cancellations and appointment history by patient age. The data shows that patients aged 38-52 and 64-77 have the most consistent attendance, with no cancellations and frequent prior visits. Similarly, patients under 24 and over 79 display reliable patterns, positioning these age groups as potentially more loyal and satisfied customers.

In the bottom-left corner, a horizontal bar chart was used to display the full set of possible activities in the patient journey - all processes begin with Check-in at the Reception. We can also observe that all consultations lead to a prescription and payment. Notably, system errors are rare, occurring in only 0,04% of the cases. In 6% of the cases, the patient is waiting for the doctor, and in 4% of the cases the consultation is cancelled. While these rates are currently low, they should be closely monitored to ensure they remain minimal.

In the bottom right corner, a pie chart reveals that most payments are made at the reception, with only 10% occurring via App. While this may not directly impact perceived service quality, it suggests the mobile payment option is underused. Considering that 28% of patients are under age 35 - data in *Patient Segmentation and Satisfaction page*, a demographic more receptive to digital solutions - better advertising should be considered to increase adoption.

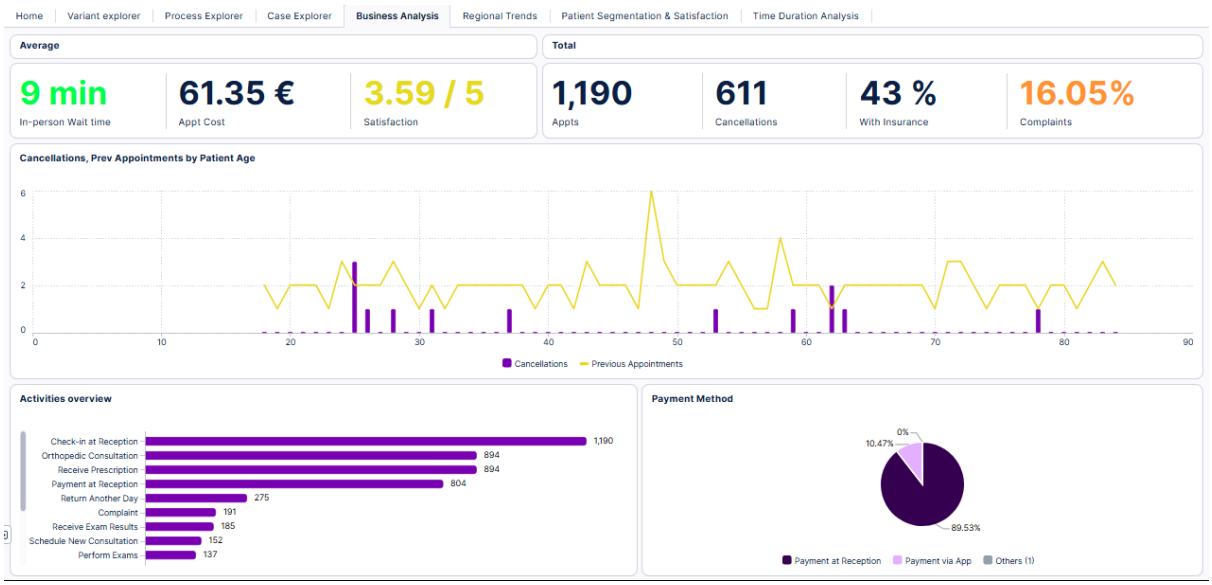


Figure 6: Business Analysis Dashboard - Overall Performance Metrics

3.5. Regional Trends

This page helps Saúde+ Health Center have an overview over the real performance of the patient appointment process, by breaking down appointment behavior and key KPIs by region, the dashboard allows us to interpret the AS IS situation, highlight geographic discrepancies, and uncover where negative outcomes like cancellations or dissatisfaction are concentrated.

At the top left, the line chart "Monthly Appointment Trends by Region" shows how appointment volume evolves over time for each region. The Center region consistently receives the highest number

of patients, followed by South, which also peaks in February. These trends indicate that both regions concentrate most clinical activity and are likely where the operational pressure and patient experience risk are greatest.

The "Total Appointments per Region" bar chart (top right) confirms the volume hierarchy: Center leads with 491 appointments, while South, Islands, and North follow closely around 230.

More critically, the "Cancellations by Region" chart (bottom left) highlights that Center also has the highest number of canceled appointments (228). This should sound our alarms, because while Center is the busiest, it's also losing the most patient interactions. The high cancellation rate directly ties into the lack of transparency cited by management and suggests that the process isn't just overloaded but misaligned with patient expectations or capacity constraints.

Next to it, in the "Satisfaction per Region" chart (bottom center), we observe that no region reaches an average satisfaction score of 4.0 (the quality benchmark). Center again leads (3.72), but the gap to the lowest-performing region (South, at 3.39) is narrow. The generally modest satisfaction across all areas reinforces the company's concern that it is not isolated but generalized.

Finally, the "Insurance per Region" donut chart (bottom right) adds another layer of insight. It reveals that the Center region has the highest proportion of insured patients (31.87%). Conversely, South and North show lower insurance coverage, which may affect both service access and patient follow-through.

Although average cost per appointment was not displayed here, prior analysis confirms the difference is marginal (less than €3 between regions), meaning cost does not explain satisfaction or outcome variations.

Overall, this page materializes what was previously only a suspicion that the appointment process suffers from underwhelming patient experience. These findings provide a clear direction for targeted interventions in Center and South to optimize their operational planning.

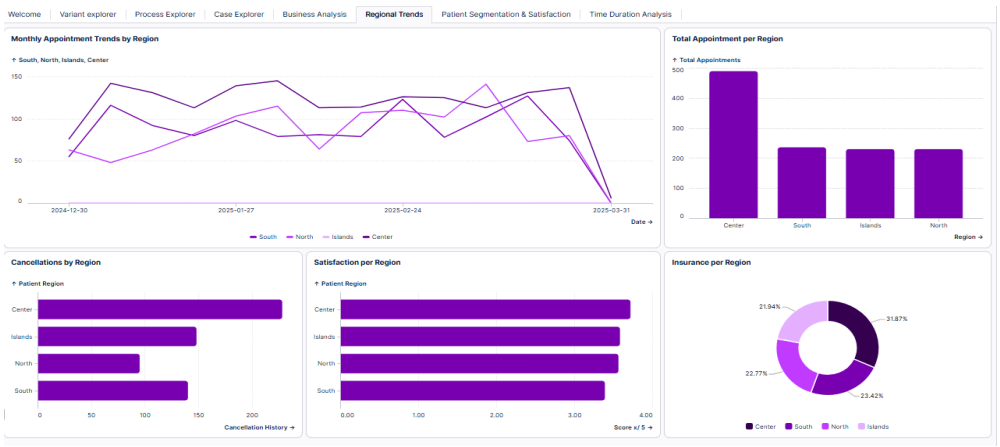


Figure 7: Regional Trends Dashboard - Geographic Performance Comparison

3.6. Patient Segmentation and Satisfaction

This page presents a dashboard dedicated to customer segmentation and their satisfaction with the service provided by the clinic. At the top, a card displays the average overall satisfaction score, which is 3.59 out of 5, indicating a generally positive experience.

A pie chart shows appointment cancellations segmented by age group, where we can conclude that the least consistent group is patients aged 36-55, while the group with the lowest cancellation rate is patients over 76. Next to it, a scatter plot illustrates the relationship between Total Cost and Satisfaction Score. We observe that there is no strong correlation between cost and satisfaction, higher prices do not lead to lower satisfaction, suggesting that the pricing is aligned with patient expectations.

In the lower left section, a bar chart compares satisfaction between patients with and without health insurance. The data suggests that insured patients tend to be more demanding, with slightly lower satisfaction scores. This might indicate that the price paid is actually perceived as a value-adding factor by uninsured patients.

At the bottom center, there is a bar chart that relates whether patients have insurance to their cancellation behavior. In this case, patients with insurance tend to cancel more appointments, which makes sense, as the cost for them is presumably lower.

Lastly, another pie chart shows the distribution of appointments by age group, which appears to be well balanced. However, it also highlights that the most critical segment is patients aged 36-55, who have an average satisfaction score of around 2.87, a value that can be considered only moderately satisfactory.

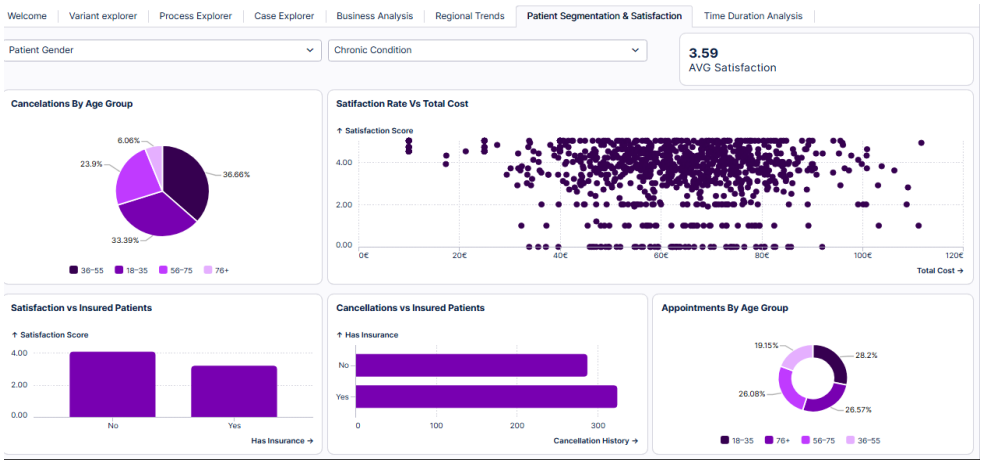


Figure 8: Patient Segmentation & Satisfaction Dashboard

3.7. Time Duration Analysis

This dashboard gives us an analysis of patient journey duration and process efficiency in the healthcare center. Key insights include the visualization on the left that shows that the average process time from check-in to Exams, which stands at 41 minutes, with a modal time of 47 minutes, a minimum of 3 minutes, and a maximum of 80 minutes, indicates significant variability in patient wait times. The top-right visualization reveals that patients with asthma experience the longest average process time (44 minutes), followed by those with hypertension and heart disease (42 minutes), no chronic conditions (41 minutes) and diabetes (40 minutes). The "Check-in to Consultation time" distribution shows that most patients (100 instances) complete this step within a space of 10 minutes, but the frequency drops for longer durations, noticing that the frequency for more than 16 minutes is 137 instances highlighted in grey.

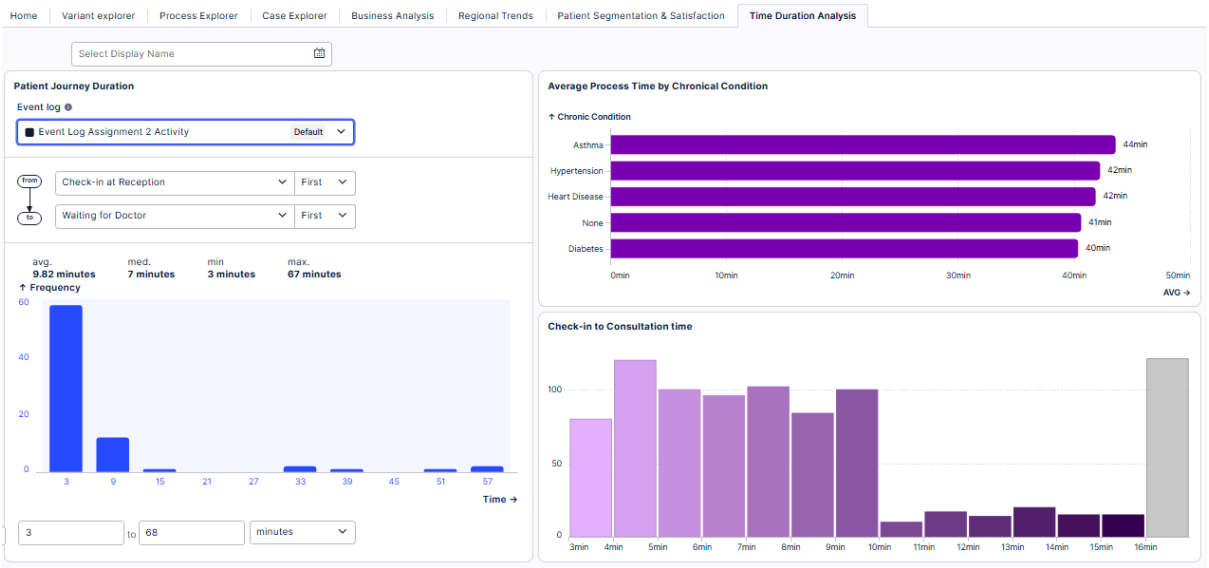


Figure 9: Time Duration Analysis Dashboard - Process Bottlenecks

4. Conclusion

This project successfully applied process mining to the event log of Saúde+ Health Center's patient appointment process, enabling a transparent, data-driven view of how the service is operating. Using evidence-based insights derived from real execution data, we identified critical metrics such as **average throughput time (9 minutes), satisfaction scores (3.59), and complaint rates (16%)** that reflect overall efficiency and patient experience.

The analysis confirmed strong process compliance (check-in → consultation → prescription → payment) and robust data logging, enabling a clear view of the patient journey. However, there are key friction points, such as the **23% drop-off rate at check-in**, **complaints after consultation**, indicating early-stage inefficiencies. Additionally, a grey area in our analysis is understanding how such a low in-person waiting time can still lead to so many complaints and suboptimal satisfaction. To diagnose further, complaints should be analyzed with greater granularity, ideally categorized by justification or activity step.

Segment-specific findings (e.g. lower satisfaction among insured patients and regional disparities) highlight the need for **personalized service adjustments**, while the high cancellation rate among middle-aged patients (36-55) suggests systemic scheduling inefficiencies. To drive measurable improvements, the Health Center should:

1. **Optimize the check-in process** by reducing wait times and implementing real-time patient notifications.
2. **Enhance consultation quality** through staff training and structured feedback mechanisms.
3. **Boost digital adoption** with tailored incentives and usability improvements.
4. **Segment and categorize complaints** to inform root cause for further process analysis.

By acting on these evidence-based insights, Saúde+ Health Center can reduce execution variance, improve patient satisfaction, and strengthen its decision-making with reliable operational data.