Technical Specification Double-side Page

1. TECHNICAL SCOPE:

'IMPROVING STORES' EFFICIENCY USING CLIENTS SHOPPING TIMES' challenge aims to identify improvement opportunities both at the commercial and operational levels. We have been solving these problems within the industry by computing the 'time in store' per customer according to their shopping basket.

(1)During the EXPERIMENT phase we have created a series of Analytic Dashboards that allows the Retailer to better understand their customers , by different DATA perspectives oriented to Stores, Customers and Products. Our solution uses a combination of Clustering algorithms, Product Recomendation algorithms and Basket Analysis to provide the best insights possible

First, our proposed solution provides the Retailer with a methodology + Dashboards which give different perspectives to give a full understanding of retail stores and different buyers. Using this methodology, the first step would be to identify MIRROR STORES with a clustering analysis made among their different Retail branches (Hiper, Super, Proximity).

(2)During the EXPERIMENT phase we have evolved our solution in two different areas: Analytics + Data Pipelines. With the added Dashboards, the Retailer can identify stores and users with similar patterns, and create promotions for their users using our promotional Reward-Engine (Roiward.io). The Analytic Dashboards provide insights about how each of those promotions affect the customer base, allowing the Retailer to understand not only their users behaviour, but also know how different promotions affect their customer base.

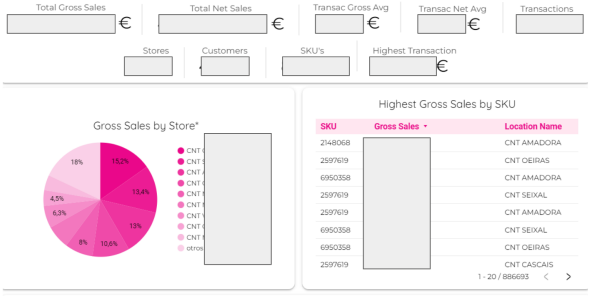
This will help to identify the stores with similar patterns (customer behaviour and operations of the store), to help us plan and execute different experiments. This way, the impact could be measured after applying the analytics + actions. Therefore the impact of the solutions provided can be measured with A/B testing with mirror vs target stores in an automated data flow. With this layer of BI Dashboard together with the data experiments + strategy, we will suggest different experiments that we have applied successfully in the past with other clients based on our Auto-ML algorithms. We will try to actuate in two different levels: Operational improvement (reduce time in-store), Sales improvement (increase sales)

As a value-added feature to the proposed solution, our promotional Reward-Engine (Roiward.io) can be linked to their SIGA platform to send different rewards to customers to improve shopping performance (sales and time). As well we can import the feedback of the reward engine campaign data to provide original customer data, giving the ability to the end-user to analyse the effect of different campaigns of the retail consumers

.

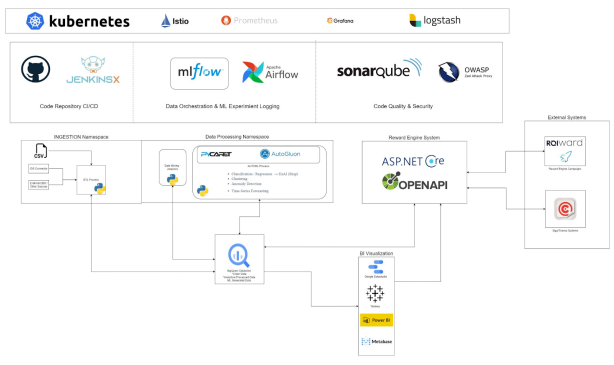
Segment	Number of locations	Number of transactions	Number of customers
1	8	643.706	43.650
3	7	124.054	12.506
2	27	25.384	4.297
5	11	37.888	5.651
4	28	5.149	1.359
Total	81	836.181	67.463

1 - 5 / 5



SKU recommendations by Store

SKU	Category	Subcategory	Location Name	Predicted Rating
4950225	0304 - águas	030401 - águas lisas	CNT MAIA	393434.30592721346
5197332	1601 - pão tradicional	160105 - pão trigo	CNT SEIXAL	166602.06605246724
4318832	1601 - pão tradicional	160105 - pão trigo	CNT MAIA	153138.8671047444



2. ALGORITHMS, TOOLS AND CONCLUSIONS:

During the EXPERIMENT phase, we have implemented a series of analytical and ML algorithms to improve the Sales of our customers:

1) Analytic Data pipeline that processes the data given by the provider, generating a series of different data tables that feed the Dashboards. At this point, we are using Google Bigquery as a DataStore, applying a set of SQL queries to process the data

2) Unsupervised ML for Clustering of Stores and Customers, mainly using the K-MEANS model. We have implemented 2 different processing pipelines to work with virtually any data size, even with thousands of millions of rows. The first one is our own in-house processing, that can apply not just K-MEANS model for clustering, but also AP / MEANSHIFT / SC / HCLUST / DBSCAN / OPTICS / BIRCH / KMODES, thanks to our auto-ml engine. The second approach uses Google KMEANS ML modelling to process dataset so big that can't YET fit in our standard pipeline

3) APRIORI/FPGROWTH algorithms to provide basket recommendation data, based on historical purchases made by customers

4) Overall product recommendations targeting customers that have never bought an specific product SKU, based on Matrix Factorization algorithms

4) [STILL IN PROGRESS] We have also implemented Time-Series forecasting algorithms, using NAIVE / SNAIVE / ARIMA / AUTO-ARIMA / PROPHET forecasting models. Our solution automatically selects the best performing algorithm for the data provided, and generates predictions for the amount of product needs for the Retailer stores

As discussed with our data provider, our solution has two different areas of focus: Operational efficiency and Sales improvement. We plan to achieve these goals by facing the challenge of covering these different areas of growth with a proposal that lies first on layer Data and experimentation Strategy & Methodology. Here we define a field of data processing flows that set the base of the experiments and metrics in order to set up different growth activities. Once we have defined this flow we will plan and execute the following

3. SCALABILITY AND FLEXIBILITY OF THE SOLUTION:

During the EXPERIMENT phase, we have observed that certain ML processes, mainly the ones that need to factorise huge amounts of data need an amount of resources that need to be managed correctly not to overspend in the use of the cloud infrastructure. We are already started implementing a custom VM manager for our cloud provider (Gcloud), that will allow us to spin up really big computational resources just the time that the process needs it, as other current solutions take too much time to stop the machines, with all the cost that this implies.

4. DATA GOVERNANCE AND LEGAL COMPLIANCE:

The proposed solution does not require any external exposures. Thus, available security infrastructure and policies within the mentioned architecture components will be enough for security assurance on data governance. However, within our ISO27001 certification process, we established internal processes, policies, and rules for user authentication along with monitoring layers that ensure the data governance of the solution complies with legal requirements that could appear during the project (these layers will be explained further in detail). Although we will not have to end-user personal data, regarding Legal Compliance and specifically GDPR, private and sensitive data feeding the solution is managed by SONAE.MC itself, and its treatment consent should be handled by them barring their legitimate interests.

From Dative.io we'll need to establish a contract with them which must include the appointment for data processing. We already work with personal information with other customers, and thus everyone within Dative.io is committed to confidentiality. Our organisation is ISO standard certifies and all our employees have authorization and control measures in place. In addition, it is important to notice that within the provided dataset there is no demographic information or any kind of customer details which could raise ethical issues. However, we will be vigilant in that respect and will react with moral responsibility if necessary, and despite all measures, we do cover ourselves with an insurance policy for up to 1.5M€ in damages due to data breach.

The data governance and legal compliance tools and systems we have in place are listed below:
- ISO 27001 Certification
- Istio.io - Simplify observability, traffic management, security, and policy with the leading service mesh.
- Monitoring: Prometheus & Grafana
- Authentication: Google IAM System
- BigQuery - Accesses controlled by roles, rows and tables

5. QUALITY ASSURANCE AND RISK MANAGEMENT:

Dative.io is composed of more than 15 international professionals with more than 5 nationalities on the team. As well we covered certifications like Google Cloud Architect and we have our own Cybersecurity team so we can take care of the projects starting with the best design principles and tools for monitoring them in real-time.

To minimise the risk and ensure that quality assurance and performance are upheld. We manage our solutions with different levels of software and systems that help us deliver safe and correct services:

- Security & access: Google IAM Auth system with 2FA and server log monitoring
- MLFlow in order to store, control, manage, update and measure the quality of our ML processes
- OWASP ZAP: Vulnerability Assessment, Penetration Testing, Runtime Testing, Code Review
- SonarQube is a Code Quality Assurance tool that collects and analyzes source code, and provides reports for the code quality of your project. It combines static and dynamic analysis tools and enables quality to be measured continually over time.
- Static code analysis within Github code and Docker images

On the other hand, as a referral and reference, part of this technology has been developed by an R&D project financed by EU FEDER Funds under the CDTI Cervera program in collaboration with Tecnalia Research and Innovation.