# Technical Specification Double-side Page

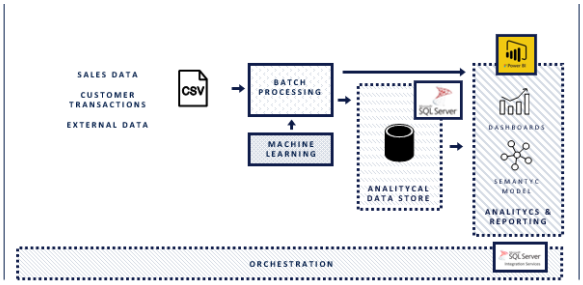
1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

The *Prediction of Consumer Behaviour in Emergency Scenarios (the case of COVID-19)* challenge aimed to help retailers and specifically demand-planners on their daily decision-making, to avoid stockouts, meet the required consumer demand and stay alerted on unusual behaviour and uncertainty situations.

The developed solution provides the final user with a Business Intelligence tool for data analysis. It is composed of several panels, which try to give answer to questions that demand-planners or those in charge of retail stores must tackle on their daily journey to manage inventory correctly and be sufficiently prepared to meet consumer needs.

The dashboards focus on sales volume descriptive analysis, inventory and stockouts analysis, demand planning advises according to an AI based forecast, day-by-day monitoring of current week sales and inventory performance, an anomaly detection and interpretation dedicated report, and basket analysis.

During the EVOLVE stage we review some of the dashboards, specially those focused on anomaly analysis, and finally we split them in two different dashboards: one focused on a global analysis aimed to support with high level anomaly scenarios evolution, and the other one focused on specific product/store analysis to react at an operative level.




2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

For the solution development, we made use of the most widely used resources on Data Science such as: **Python** programming language and its related libraries (numpy, pandas, matplotlib, scikit-learn, etc.), working inside Jupyter ecosystem for data pre-processing, EDA and modelling.

A ML model (based on Gradient Boosting Regressor algorithm) was trained on MIGROS data to forecast weekly demand assuming stable situations. Feature engineering was carried out and the model learned from historical data patterns and seasonality. Not only sales forecasts but also 80% and 90% safety stocks were predicted for each Store-product category aggregation series.

Another ML model was trained on same data, but in this case, seasonality and date features were ignored to better focus on short-time sales lags and predict what now is happening. This model is better understood as *emergency* forecasting.

Finally, an anomaly detection approach was implemented to detect unusual sales volume on every Store-Product Category reference. A time-series daily forecast was built for every series in the dataset using Prophet library. The model was able to well capture seasonality in the data and to set 95% confidence intervals for prediction. It was observed that peak sales normally occurred during weekends. Anomalies were identified based on confidence intervals, if sales volume get considerably out of them under a specific bound constraint.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in?

The solution is built up on different components, all integrated on standard technologies (i.e, Microsoft Power Bi Platform). Different options can be selected under each component to meet specific technical requirements and adapt to future technological needs, for example, well-fitting with pre-existing data storage systems used in this case by retailers.

The tool could be deployed on-premise on in the cloud using iaaS or other specific data services which encapsulate a set of technologies of the same characteristics. Once in the cloud, scalability gets even easier because those services are inherently prepared to scale horizontally to attend more users or deal with more data, i.e. Docker, BI semantic models or ETL processes. Even more, some of the components could be deployed on-premise and work transparently with the rest of the components deployed in a cloud environment. During EXPERIMENT we defined the whole Big data Lambda architecture to deploy the solutions through AZURE services, and during EVOLVE we have continued researching to enhance our flexibility by analyzing equivalent services on AWS (though connecting in the end to PowerBI Service)

Semantic BI technologies and user dashboards are highly customizable with the proposed technologies and the appropriate methodologies for another related sector scalability. The challenge with the provided data is focused on the final level of the supply chain involving the food retail industry, but we aim to make it scalable to the consecutive levels such providers and manufacturing industries. Of course, not on only food-related domain is observed, but also textile, electronics and many other kinds of businesses within the massive commercialization of products or services uniform to large quantities of final customers.

Regarding TRL levels, the solution architecture it's placed on the 9$^{th}$ level as built Power BI dashboards are ready for full commercial deployment. The Artificial Intelligence components developed, the machine learning models trained, remain on the 6$^{th}$ level. Computations have been ran, evaluated and simulations have been carried out on unseen data, but we still didn't have the opportunity to deploy in an operational environment.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR).

The proposed solution does not require any external exposure. Thus, available security infrastructure and policies within the mentioned architecture components will be enough for security assurance on data governance. Policies and rules for user authentication and monitorization can be established through the mentioned analytical tools, i.e. SQL Server security model.

Regarding Legal Compliance and specifically GDPR, private and sensitive data feeding the solution is managed by the retail business itself, and its treatment consent should be handled by them for their legitimate interests. From CYC we'll need to stablish a contract with them which must include the appointment for data processing. On CyC we have already implemented appropriate technical and organizational measures to ensure that our processes meet the GPDR and to guarantee the protection of the right of the individuals. We already work with personal information with other customers, and thus everyone within CyC is committed to confidentiality and our organization has 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 any ethical issue. However, we will be vigilant in that respect and will react with moral responsibility if necessary.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfills the challenge/Theme Challenges and data provider requirements.

CyC has more than 22 years of experience on the design, development, and implementation of BI solutions. We rely on our own procedures and policies to improve and test our solutions and offer high level quality assurance. In addition, CyC is comprised of more than 90 professionals of the SW industry, with 14 directly related to data business lines.

The most challenging issue when developing the solution relied on data quality and data complexity. High volumes of data coming from hundreds of retail references based on different SKU, product categories and store locations made difficult to obtain accurate forecasts and extract patterns from each of them. We had to make balance between getting an accurate solution and entering detail as much as possible. That's why the strategy rely on processing and modelling data through different level of aggregation of time series (aggregating products by category, stores by city, etc.), and finally selecting the best approach which fulfills retailers needs in front of demand planning.

# Annex 1. Means for accessing the MVP

Please, indicate in 1 page indicating the means for accessing the MVP for a potential customer (login information, website address, link to a demo video or whatever means are needed to check that the MVP exists and works).

The MVP demo can be accessed through this link: (439) REACH video - YouTube
Link to the promotional video: (441) REACH CYC Promotional video - YouTube