BU7149: Big Data and AI in Business

Group Members:

Álvaro Tapia

Niccolò Guerzoni Panichi

Ivan Nemtsev

Carlos Ramirez

Jose Ramon Cobos

Date: 5/03/2023

Table of Contents

Introduction

Mercadona is a Spanish supermarket chain well-known in Spain and Portugal, where it started to operate in 2019. By 2023 this company had the plan to expand to Italy. The plan of going abroad came from the fact that the Spanish market became too small for Mercadona. The company even closed some non-profitable stores in Spain. The choice of countries can also be explained easily. Mercadona decided only to use already existing supply chains and that is why expanding further than neighbouring countries would be challenging logistically. Italy is going to be the best option for the next expansion for numerous reasons: it is a neighbouring country, the cultural differences and cuisines are similar to Spain and Portugal being Mediterranean, and finally, even the name of the company arose from there as the transalpine country has been in Juan Roig's (CEO) mind from the very beginning.

Mercadona is also planning to improve its efficiency by increasing sales and customer satisfaction by introducing a delivery service. This is going to be done through gathering and processing various data, testing different options of delivery, and analysing the results.

In order to gain a competitive advantage a competitive analysis will be performed as the second objective of the company.

Since Mercadona is planning to expand to other countries after Italy, a data strategy for analysing which country should be the next is created.

Each of the three objectives is discussed separately and in detail in this document, with details on how and what data will be collected, what will be done with the data, what analysis will be done, and which professions and business activities will be required. The primary objective of the article is to build an analytical architecture and data strategy to support all the objectives. The final data strategy will be presented separately after the objectives, taking into account differences.

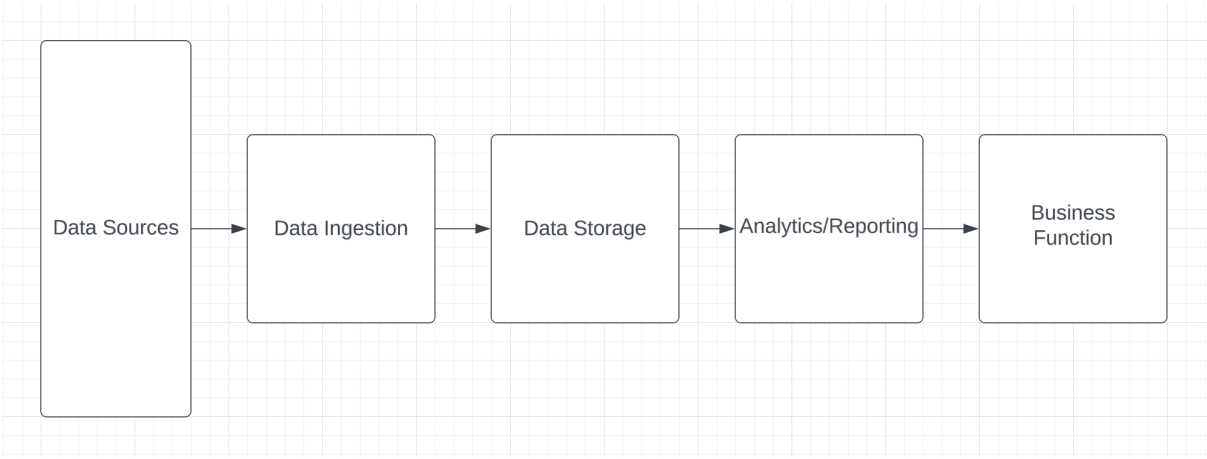

Diagram 1: "Data Strategy"

Objective 1: Increase client satisfaction with a better delivery service.

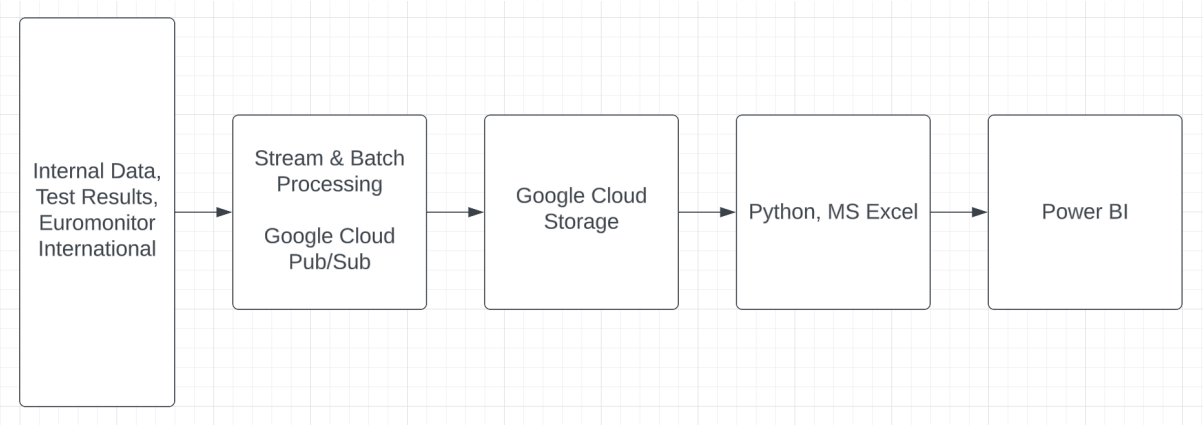

Diagram 2: “1 Objective Data Strategy”

1.1. Data Source

A company like Mercadona, which wants to continue growing, must be able to attract new customers and retain the ones it already has. That is why improving its delivery service is one of its main objectives, as more and more people prefer to shop online. The first thing to study in order to know how to improve is who our customer is.

For this purpose, the data that need to be analysed are:

- Customer data: to determine where their customers are located, what they typically order, and when they typically place their orders. This will help the company optimize its delivery routes and schedules and ensure that they have the right types of products available when its customers want them.

- Demographic data: to obtain more detailed information about the characteristics and behaviours of the people who live in the areas where they provide delivery services. This could help the company tailor their marketing and product offerings to better meet the needs and preferences of its customers.

Once Mercadona has an idea of who wants to target and what are the preferences of targeted groups, the delivery service can start. In order to optimize its operations and provide the best possible service to its customers the company needs to analyse a different set of data that will be gathered during the operations.

These data are:

- Inventory data: to determine which products are in stock and which ones are running low. This will help the grocery store ensure that they have enough products on hand to fulfil delivery orders and avoid stockouts.

- Sales data: to determine which products are selling the most and which ones are not selling well. This will help the grocery store adjust its inventory and product offerings to better meet its customers' needs.

- Delivery data: to determine how long it takes to deliver orders to different areas, which routes are the most efficient, and which drivers are the most reliable. This will help the grocery store optimize its delivery processes and ensure that it can deliver orders as quickly and efficiently as possible.

- Traffic data: The grocery store should analyse traffic data to determine which routes are the most congested and which times of day are the busiest. This will help the grocery store optimize its delivery routes and schedules to avoid traffic congestion and ensure that they can deliver orders on time.

The orders for the delivery can be made through the app or website. In this way, Mercadona will gather more data that needs to be analysed:

Website or app data: Mercadona should analyse these types of data in order to improve the customer experience, improve the customer journey, and be able to sponsor specific products to specific people.

Reviews: Analyzing reviews would allow Mercadona to have better reputation management, improve quality controls, and increase overall customer satisfaction.

1.2. Unstructured Data

Most of the data that Mercadona needs to analyse are structured data: demographic data, consumer data, sales data, inventory data, and delivery data are typically structured data as they are usually organized and have a clear data model. The data gathered from reviews, instead, will result in unstructured data.

By using sentiment analysis, unstructured data of text can be transformed into structured data that can be used to gain insights and make informed decisions. Once the sentiment analysis has been performed, the data can be structured and categorized based on the sentiment scores, making it easier to analyse and visualize the data. This structured data can then be used for further analysis, such as identifying trends or patterns in customer sentiment, or for creating predictive models based on historical data.

1.3. Obtaining the Data

Before starting the delivery service in Italy, Mercadona can consider buying data from other companies to gain insights into the Italian market and customer behaviour. By buying data from other companies, Mercadona can supplement its own market research and gain a deeper understanding of the Italian market and customer behaviour. This will help Mercadona to make informed decisions about its delivery service and increase its chances of success.

The data that Mercadona needs to buy from other companies are consumer, demographic, and competitor data. All these types of data can be purchased from market research companies that operate in Europe, like Euromonitor International, Nielsen, Gfk, Kantar, and Ipsos.

The best choice for Mercadona would be Euromonitor International for several reasons. Euromonitor International offers comprehensive coverage of the food and beverage industry, consumer goods, and retail, which are key industries for Mercadona. Euromonitor International is known for its high-quality research and analysis, and for offering customizable solutions that can be tailored to meet the specific needs of its clients.

1.4. Data Ingestion

For what concerns Data Ingestion, Mercadona will need both stream processing (since you need real-time, for data like inventory, delivery, and traffic) and batch processing. Stream processing involves processing data continuously as it arrives, without waiting for all data to be collected before processing begins. The stream processing will be used for those data that are needed in real-time, like inventory data, delivery data, and traffic data. Batch processing, on the other hand, is the processing of data in a batch, where data is ingested and processed in discrete chunks, usually collected over a period of time. In batch processing, data is collected, stored, and processed in large volumes. Mercadona will use stream processing for demographic, consumer, and sales data.

Google Cloud Pub/Sub is a good choice for data ingestion for Mercadona because of:

- Real-time data processing: Google Cloud Pub/Sub is designed for real-time data processing, which means it can handle high-speed data ingestion and processing. This is essential for Mercadona as it needs to process data in real-time to make quick decisions and respond to market changes.

- Reliable messaging: Google Cloud Pub/Sub provides reliable messaging, ensuring that data is delivered to the intended destination even if there are issues with the underlying infrastructure or connectivity. This means that Mercadona can trust that its data is delivered and processed correctly, minimizing the risk of data loss or errors.

Google Cloud Pub/Sub can also be used for batch processing, but it may not be the most efficient or cost-effective way to handle batch processing jobs. Batch processing typically involves processing large amounts of data that have been accumulated over time, which is not the primary use case for Google Cloud Pub/Sub.

Instead, Google Cloud offers other services like Cloud Dataflow, Cloud Dataproc, and Cloud Composer that are specifically designed for batch processing of data. These services can be used in conjunction with Google Cloud Pub/Sub for end-to-end data processing pipelines that include both real-time and batch processing.

1.5. Data Storage

For the storage, Mercadona should use another Google service since it would ensure compatibility and it would make it easier to integrate and manage the data processing pipelines, and it would reduce the latency. Some cloud service providers offer reduced costs for data transfer and storage when using their own services. By using the same provider for data ingestion and storage, companies can potentially reduce their overall cloud infrastructure costs.

Google Cloud Storage is a good choice for Mercadona because of the following reasons:

- Scalability: Google Cloud Storage is highly scalable, which means it can store a large amount of data and handle a high volume of requests. This is important for Mercadona as it will likely have a large amount of data to store and manage as it expands its business and launches a delivery service in Italy.

- Durability: Google Cloud Storage provides high durability for data, ensuring that data is protected and available even in the event of hardware failures or other issues. This means that Mercadona can trust that its data is safe and available when it needs it.

- Security: Google Cloud Storage provides strong security features, including encryption at rest and in transit, access controls, and audit logging. This means that Mercadona can store its data securely and meet regulatory requirements for data protection and privacy

1.6. Analytics

For the Analytics a combination of Python and Excel will be used. This is due to two reasons: Python has a larger set of libraries for handling Big Data and skilled professionals on Python are easier to find.

Big Data will be used in the analysis in particular for what concerns demographic and customer information.

Artificial Intelligence will play a big role in this strategy. Some techniques used will be:

- Route optimization: AI algorithms can be used to optimize delivery routes, taking into account factors such as traffic, weather, and road conditions, to reduce delivery times and fuel costs.

- Predictive analytics: Predictive analytics can be used to forecast demand for delivery services, enabling companies to plan their resources and schedules more effectively.

- Natural Language Processing (NLP): NLP can be used to automate customer service and support, by analysing and responding to customer inquiries and complaints through chatbots or virtual assistants.

- Demand forecasting: AI can be used to predict demand for grocery delivery services, based on factors such as time of day, day of the week, and seasonal trends. This can help stores plan their resources and schedules more effectively and reduce the likelihood of stockouts or overstocking.

- Customer recommendations: AI-powered recommendation engines can be used to suggest products and promotions to customers based on their purchase history, preferences, and behaviour. This can help stores increase sales and improve customer loyalty.

- Inventory management: AI algorithms can be used to optimize inventory levels, by analysing sales data and predicting which items are likely to be in high demand. This can help stores avoid stockouts, reduce waste, and improve profitability.

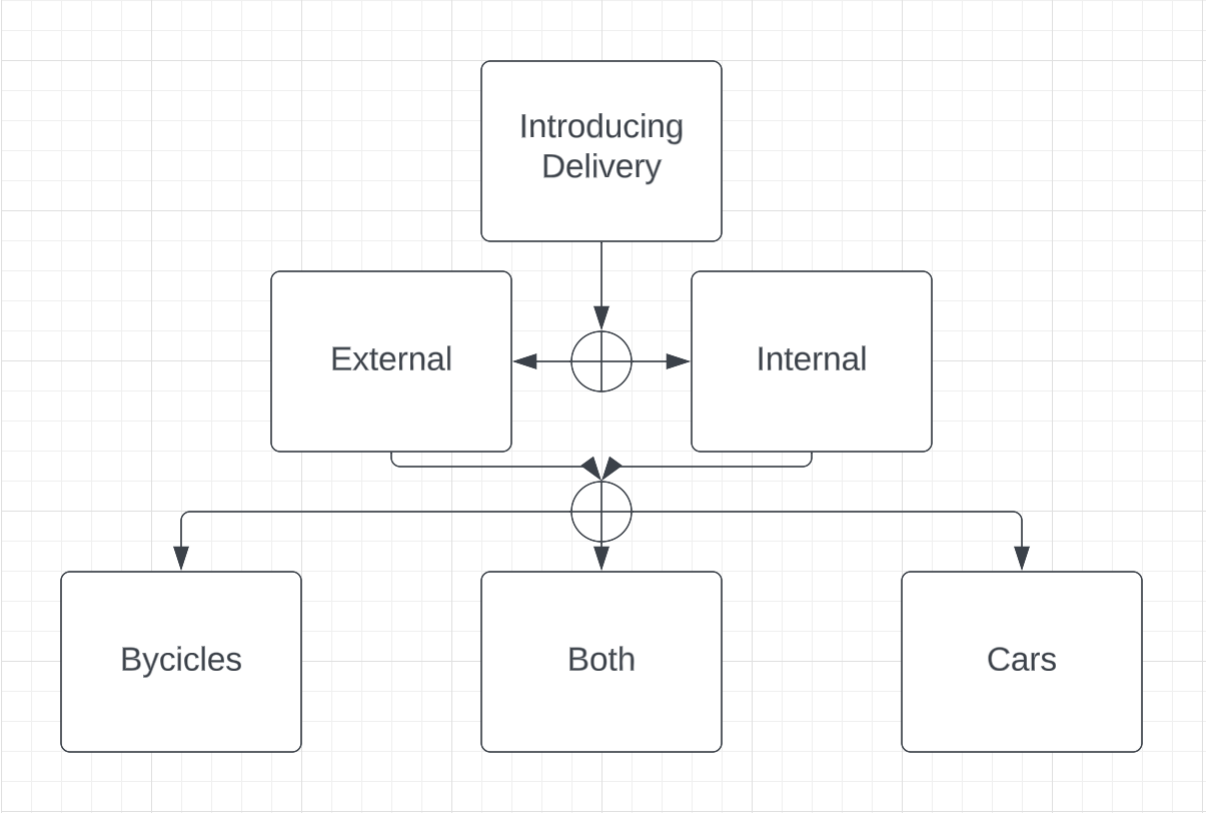

Diagram 3: “Delivery Service Options”

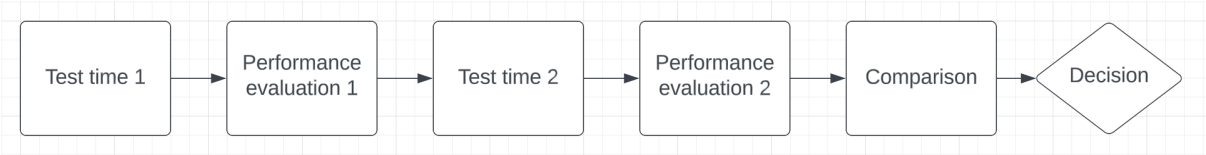

Diagram 4: “Decision-making Process for Delivery”

1.7. Business functions

For this particular objective, the best choice for data visualization will be Power BI. This is due to the fact that Power BI has built-in real-time analytics capabilities, allowing businesses to monitor and analyse data in real-time. This can be especially useful for the grocery store's delivery service, as it can help track orders and inventory in real-time.

Objective 2: Increase foreign market share by competitor analysis.

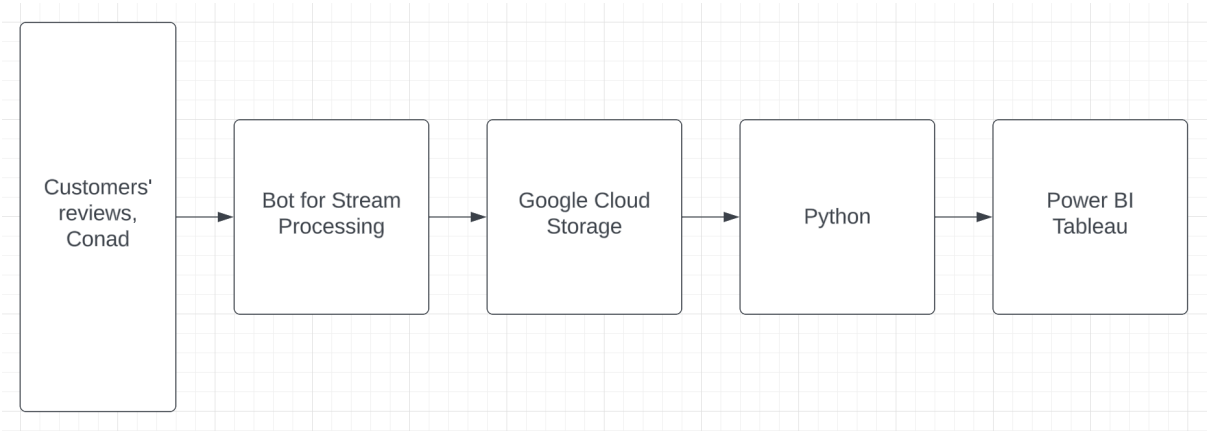

Diagram 5: “2 Objective Data Strategy”

The desire has driven Mercadona's expansion of its operations abroad to increase market share through competitor analysis. The company's bold strategy to open new stores in Italy comes with a robust analytics infrastructure. Mercadona has decided to implement technological processes to establish its business in Italy and has done so by retrieving and analysing available information from a relatively similar-sized Italian company.

Analysing Mercadona's future competitor involves following a series of data-driven procedures and maintaining a solid framework for the foreseeable future. Before stating the purpose of the data, it is essential to report on the Italian competitor to be analysed. Conad is an Italian retailer with a wide range of supermarkets throughout Italy. According to Ozbun (2022), Conad leads the supermarket chain sector in Italy with a 23.4% market share. This company not only outperforms the competition in the market due to a high sales volume but also because of the technological development reflected in its online business.

Conad's website has nothing to envy from a global giant such as Amazon or Walmart. The platform's interface is user-friendly and more importantly, equipped with a data-rich feedback section on each product sold. "Mercadona is the leading supermarket chain in Spain with 27.4% market share" (Eloise Trenda, 2023). It is obvious to choose Conad as a model to imitate and potentially steal market share from, as both companies are leaders in their respective countries and invest heavily in the digital race.

2.1. Data source

The objective is to get a reliable and abundant source of information to collect data from the customer reviews of our main competitor in Italy, Conad to carry out the

analysis; this source is the reviews section of Conad's own website. This information is unstructured, so a pre-processing of the data, data parsing, will be carried out to decompose the data into smaller components that can be quickly processed and understood by the program.

The type of reviews Mercadona could rely on to develop the analysis are:

- Product reviews: Data related to customers' opinions about competitors' products.
- Customer service reviews: Data related to the sentiment of our competitor's customers regarding customer service.
- Facility status reviews: Data related to the customers' opinion of the condition of the facilities, such as cleanliness, organization, and accessibility.

This data is necessary to evaluate in general terms our competitor's services in Italy and thus offer its customers better services in these aspects.

2.2. Data ingestion

Mercadona would most likely need to use stream data ingestion, which would require processing data in real-time or near-real-time. Customer reviews are typically generated continuously and in high volume and processing them in batches would result in significant delays and lower the accuracy of the analysis.

Stream data ingestion would allow Mercadona to collect and process customer reviews in real-time or near-real-time, allowing for more timely insights and decision-making. Using stream data ingestion, Mercadona could also potentially detect emerging trends or issues in customer sentiment as they arise and respond quickly to address them.

Batched types of data ingestion can be very costly and require a high level of professional expertise to maintain the servers. To perform this task, it would be necessary to develop a bot that would automatically collect in real time all the reviews from the Conad website. Mercadona will need two data scientists. These professionals are assigned to collect the data influx and ingest the dataset alongside a streaming service provider.

2.3. Data storage

Mercadona could consider using a cloud-based storage solution like Google Cloud Storage.

A cloud-based storage solution is better for the company in terms of scalability, durability, and security. All these reasons are explained in detail in the first objective.

2.4. Analytics / Reporting

The data analysis and reporting phase will be staffed by another dual team of professionals, combining the necessary technical and business expertise. A computer scientist and a business analyst will work on analysing, retrieving information, interpreting, visualizing, and presenting the data. The computer scientist will import the structured data into the Python software, and this is because:

- Python has a vast ecosystem of libraries and tools designed specifically for NLP and sentiment analysis.
- Python has a large community of developers who contribute to libraries, tools, and forums, which makes it easy to find solutions to problems.
- Python integrates well with other technologies commonly used in data analysis.

Sentiment analysis will be performed through text analysis, supported by an opinion lexicon, to extract valuable information and general knowledge about our competitors' services, products, and establishments. The following quote explains this technology in more detail. "Sentiment analysis, also known as opinion or sentiment mining, captures the polarity of text, often classified as positive, negative, or neutral. Moreover, the association of sentiment and emotion with text spans different levels, such as sentences, paragraphs, and documents" (Vijay Kanade, 2022).

2.5. Business functions

The Business Analyst will be responsible for reporting the sentiment results and data from the Python executions. Tableau will be the software needed for data visualization. Tools like Power BI are very popular and probably less expensive compared to others, so they tend to be a favourable alternative for many organizations. However, Tableau may be better suited for visualizing complex concepts such as sentiment analysis and data-driven results. Another important use is that this visualization software has many graphic displays, which is convenient because when presenting to different audiences, it requires different graphical forms to present the data. Tableau has long-term membership packages that make it cost-effective in the long run, especially since this analytics project will be an ongoing effort.

Objective 3: Expand to new countries.

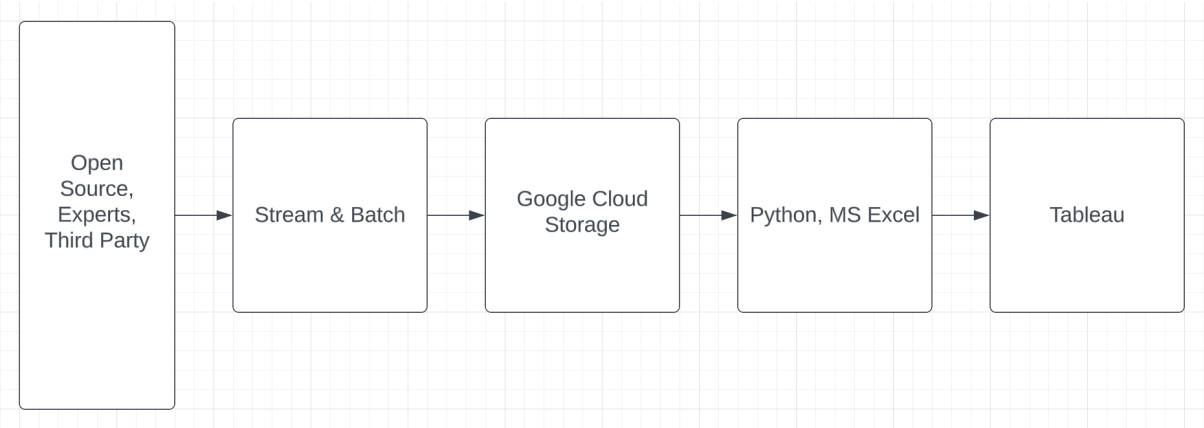

Diagram 6: “3 Objective Data Strategy”

3.1. Data Source

Mercadona has managed to expand throughout the Iberian Peninsula and now wants to expand to Italy, another of its objectives is to expand into as many countries in the European Union as possible. To this end, it has been concluded that it requires the following data:

- Legislation and regulations: Obtain from the web pages of the different countries where the company wants to expand the business. The data is unstructured and will need a law expert to summarize the main laws and regulations that can affect us.

- Demographics, economic and educational situation: We need to know the possible customers of the business, so we need to study the:

    - Population density: It needs mapping software but perhaps is very expensive for the company to have if we are going to use it only at the start so it’s a better idea to buy it from an outside company specializing in this kind of research.
    - Economical and educational situation: Buy from a third-party data of possible customers.

It's going to need a data analyst that joins both types of data to find the best places to open the new shops.

- Logistical information: To know where to build a distribution centre which is capable of efficiently supplying all the shops to be opened in the area. The data to be entered should be geographic, real estate, road, and connection data. Transform the semi-data-like maps and roads into structured ones.

- Competitive environment: Data about the competitors that the company will face if they finally open the shops. Stock market data and also the public opinion of those companies in social media to discover the weakness that it can try to use to its advantage and what we should avoid doing. Data from social media is semi, we need to convert it.
- Cultural data: Finally, to understand the culture and traditions of the country to see if our business model can fit in the country and what things the company should change to be able to adapt better to the country. In this case, it can use again the users' data and use an expert that can create a list of topics to be aware of.

3.2. Data Ingestion

It can introduce the data from third-party companies by streaming into the company but will need the help of a data analyst to batch the rest of the data. It's going to be a big task at the start of the project because it introduces all the data on legislation, population, logistics, and cultural data. Perhaps the company needs to create a team for this and later keep some of the staff for the competitors' data. It will introduce data about the rest of the companies but it's data that can change very fast, so we need to keep updating it.

3.3. Data Storage

Because the data is not something that the company needs to use every day it can be placed in the cloud. It is going to be cheaper and easier to manage than having a server in the company. It can save money because there is no need for a room for them and the extra security that they need.

For this reason, it´s going to use Google Cloud which offers the opportunity to store all of its data online and offers good scalability, durability, and security. It also contains Dataproc, a managed Hadoop environment. Hadoop is the most commonly used program for data storage because of the tools it offers and because it allows converting semi-structured data into structured data, so an analyst that knows how to use it can also work with Google Cloud.

3.4. Analytics / Reporting

For the analysis of the data that has been collected, it will use a mix of Excel and Python. Excel is a very commonly used software that can display unstructured data, even if it is not the best software that can be used to start working with the data, making it more accessible for the rest of the departments of the company that perhaps need to use some of this information.

But it needs to use a language like Python to be able to obtain a meaningful output of all the data. Python is one of the best options for data analysis, with an ever-increasing number of users as well as a huge number of libraries that can be used to

speed the process of analysis. This is useful because it is easier to find data analysts with Python backgrounds and being such a popular language, it is easier to find solutions to the possible difficulties they may find.

The objective of the analysis is to find if entrance to a country is possible, which are the areas with the highest population with the same economic power as the average customer in Spain who are not satisfied with their current situation. In this way, it will be easier to settle in those places and to get to know each other among the citizens so that when shops open in places with more competition it will not be so difficult to start.

3.5. Business Functions

Finally, they are going to use Tableau to show the findings of the project to the main departments that can be involved in the expansion of the company, logistics, and marketing. Tableau is harder to learn in comparison to other software like PowerBI or Excel, but it has no data limits and an experience data analyst will be able to represent the data in a more meaningful way. For this, it is going to need a data analyst with experience in Tableau but nowadays a lot of them have both Python and Tableau experience.

3.6. Data Strategy

After the different teams have done research on the data and the different programs, they need to be able to process it, they should be put together to find out what common elements can be shared in order to reduce the costs and the work required to process so much data.

The data needed for each objective is different but can use some common data. For example, demographic data that requires mapping software or competitor data obtained from different social media. It has been decided to select Euromonitor International as our primary provider of consumer, demographic, and competitor information.

For the data ingestion, we have decided to do most of it by stream. The companies contracted to provide data will do it on a constant basis and we will create several bots to extract data from the social networks automatically. But we will need a couple of data analysts to enter initial data for expansion to another country.

It has been decided that the most efficient and cheapest way to store data will be to contract Google Cloud services.

For the analyses, Excel and Python will be used for the reasons previously stated and a team of professionals will process all the data.

Finally, it has been decided that they will use Tableau to present the findings to the different departments. Although it may be a bit more complex due to the large amounts of data we may need to display, we prefer this option.

To be able to keep the company's information up to date with all the data it collects, a team of at least 10 data analysts is required. But at the beginning, it is

recommended to create a larger team that is able to put everything in operation because at the beginning it will be when more data will be introduced and processed.

Conclusion

Mercadona, the Spanish giant supermarket chain, wants to focus its efforts and analytical techniques to achieve several corporate goals. The strategy accounts for three data-driven objectives followed by a series of phases. First, Mercadona wants to increase client satisfaction with a better delivery service. An objective committed to tracking customer/demographic data, using batch and stream processing techniques, and using many AI algorithms to analyse. Second, Mercadona wants to increase foreign market share through competitor analysis. This objective focuses on sentiment analysis in Python from the online reviews of the Italian supermarket chain Conrad. Third, Mercadona wants to expand into new countries. This objective will obtain the data from many sources and legislation, ingest it in a batched type and analyse it in Excel and Python. Finally, based on the objectives' similarities and efforts needed to keep them running in the foreseeable future, the company has decided to create a new analytics department. The new department will be equipped with 10 data analysts.

Reference List

1) Project Pro, 02/02/2023 *How Food Delivery Apps are leveraging Big Data Analytics?* https://www.projectpro.io/article/how-food-delivery-apps-are-leveraging-big-data-analytics/197

2) *Starting a Delivery Service - Doing Your Competitive Analysis*. Roundtrip Blog https://blog.roundtrip.ai/how-to-start-a-delivery-service/

3) Henny Jones, 29/07/2021 *The Role of Data Science and Data Analytics in Food Delivery Startups*. CodeX. Medium https://medium.com/codex/the-role-of-data-science-and-data-analytics-in-food-delivery-startups-2f8359f5d151

4) Rahul Saini, 17/09/2021 *How Data Analytics Is Transforming Delivery Management Business* https://networkon.io/resources/blog/data-analytics-is-transforming-delivery-management-business/

5) Google Cloud. *What is Pub/Sub?* | Cloud Pub/Sub Documentation | Google Cloud https://cloud.google.com/pubsub/docs/overview

6) Ozbun, T. (2022). *Leading supermarket chains in Italy in 2021, by market share.* Statista. https://www.statista.com/statistics/1285229/leading-supermarkets-italy/#:~:text=The%20top%20five%20supermarket%20chains,by%20Selex%20with%2022.22%20percent.

7) Trenda, E. (2023). *Market share of the biggest supermarket chains in Spain in 2022*. Statista. https://www.statista.com/statistics/761781/sales-share-of-the-biggest-supermarkets-in-spain/.

8) Kanade, V. *What is Sentiment Analysis? Definition, Tools, and Applications.* Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-sentiment-analysis/

9) Tondak, A. (2022b, December 6). *Structured, Semi Structured and Unstructured Data*. https://k21academy.com/microsoft-azure/dp-900/structured-data-vs-unstructured-data-vs-semi-structured-data/

10) *How to start a distribution business in 9 steps* | indeed.com. (n.d.). Retrieved March 5, 2023 https://www.indeed.com/career-advice/finding-a-job/how-to-start-a-distribution-business

11) Team, C. B. I. (2023, February 15). T*ableau vs Power BI: Which is better?* BI Connector Blog. https://www.biconnector.com/blog/tableau-vs-power-bi/

12) Hillier, W. (2023, February 24). T*he 9 Best Data Analytics Tools for Data Analysts in 2023*. CareerFoundry. https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/

13) Haan, K. (2023, January 10). *Best Data Analytics Tools & Software* (2023). Forbes Advisor. https://www.forbes.com/advisor/business/software/best-data-analytics-tools/

14) *On-Premises vs. Cloud: Pros and Cons of Each*. (2021, December 8). https://www.teradata.com/Trends/Cloud/On-Premises-vs-Cloud

15) Hassan, F. U. (2023, January 18). *Data Warehouse Solutions—On-Premise vs. Cloud*. Royal Cyber. https://www.royalcyber.com/blogs/servicenow/on-premise-vs-cloud-data-warehouse/

16) Smallcombe, M. (2020, December 2). *Hadoop vs. SQL* — Which is Better for Data Management? Integrate.io. https://www.integrate.io/blog/hadoop-vs-sql/

17) *How to start a distribution business in 9 steps* | *indeed.com*. (n.d.). Retrieved March 5, 2023 https://www.indeed.com/career-advice/finding-a-job/how-to-start-a-distribution-business