

# CAPSTONE

## Client Prioritization for route optimization in a CPG Company

### Introduction

Distribution in the Consumer Package Goods Industry is a process that requires high optimization. The cost of goods and margins are low, while the number of deliveries might be high, making every cent count in the balance.

In this project you will design a model to optimize the number of contacts and deliveries, focusing on profitable customers and increasing the margin of the less profitable.

For this project you will use a real dataset from Pascual, a leading CPG company in Spain, with data from 2024.

### Context

Pascual is a well-known Spanish company, primarily recognized for its production of dairy products, beverages, and other food-related items. Pascual has grown into one of Spain's leading food and beverage companies. It started with dairy products in 1969 but has grown into a lot of other categories such as plant-based beverages, water, coffees, juices, and it also distributes an important number of other brands in the CPG industry.

### Pascual Logistics

Pascual has around 120.000 clients in Spain, and they are divided by delegations (cities) and channels:

- HR -> Restaurants and bars (e.g. Casa Dani)
- HH -> Restaurant Chains (e.g. San Gines, Makkila)
- AR -> Small Shops (e.g. Typical convenience Shops open 7 days-)
- DM -> Retail supermarkets (e.g. Mercadona)

For the matter of this challenge, we will only focus on HR and AH Channels (which are the most complex to manage due to large numbers)

***In these Channels each client has a pre-established number of contacts per week.*** These contacts are the number of times a client is contacted or visited by a promotor (agents that take the client's orders) within a month. Once the order is taken, it is routed in within the accorded date of delivery.

Comentado [MP1]: Dias y veces al mes

There are some concepts you should familiarize yourself with to understand the Dataset and challenge:

- **Median Ticket:** is the median of the income generated by the orders of a certain client within the selected period. *If the period only includes 2 values, it will take the average.*
- **Number of Orders** represents the total number of orders of a client within a certain period.
- **Frequency:** Is the median number of orders a specific client has within a period. *This attribute will not be given, it must be calculated.*
- **Number of Contacts:** number of times a promotor contacts the client within the established period. You will see 2:
  - o **Contacts prom month:** these contacts are made physically
  - o **Contacts tel month:** these contacts are made by phone
- **Net Income:** total amount of money that comes from the specific client (comes from the sum of the total amount of each order within the established period)
- **Efficiency:** number of orders / contacts. *This attribute will not be given, it must be calculated.*
- **Volume:** general weight of all the products within the order, *it should not be considered as quantity.*
- **Promoter ID:** Unique ID of the promoter (visitor)

Comentado [MP2]: Por pedido (no poner mensauk)

***In theory, the number of contacts and orders should be aligned,*** meaning each time a client is contacted by a promotor, an order should be taken. But it is not always like this. There are some situations that are not ideal:

- The number of orders is lower than the number of contacts meaning team inefficiency.

- The medium ticket for the client's orders is equal or lower than 80€ (a ticket below 80 can compromise order rentability)
- Both of the above (worst scenario)

## Objective

The main objective is to make the orders system more efficient by reducing logistical and labor costs. ***To do so, you should establish a new frequency and/or number of contacts for the clients based on historical data and estimate the savings.*** To do so, the following facts should be taken into consideration during the prioritization procedure:

- The logistics cost per order is 10€
- The cost of a promotor visit is 15€
- The efficiency should be focused on the promotor visits mainly, we do not aim to reduce the number of calls.
- A client that has a certain number of orders and is being contacted/visited more times is not ideal.
- A client with high frequency and low median ticket is not ideal.
- Each order has a transportation cost, and even though we charge for delivery we want to avoid over shipping, as it leads to inefficiencies such as a truck having to make more stops or arriving late to a destination.
- For Pascual it is important to deliver a quality service, so the changes you suggest should not generate a noticeable impact on the client. We want to maintain order volumes while optimizing the logistics.
- The storage capacity of the clients is not unlimited, so the changes you suggest should not be extreme
- The behavior across channels and geographies varies a lot.

## Explainability

After the creation of the new frequencies, Pascual team would like to understand the changes made to the clients and the efficiency obtained within those changes.

In order to do so, you should design an LLM tool (using RAG or similar) that when asked about a specific client, explains using a chain of thoughts and other indicators, the change made in the frequency/Number of contacts.

Take this as an example:

- The client 653025 has a median ticket of 80€
- It had a frequency of 3 times per month and 4 contacts pre-established.
- The model offers 1 visit every 2 weeks as an optimization measure.
- These should guarantee an alignment between frequencies and contacts and hopefully an increase in the median ticket.
- With these changes, XXX€ will be saved in X period of time

This tool, using an LLM model will justify the suggested frequency based on the chosen model.

## Procedure

You are free to use the model you consider better to solve this problem while you guarantee a good solution to it.

For the model you have a white canvas, you can do a mix of models/optimization tools to deliver a solution that makes sense for the business. *If you consider important to do a segmentation to understand the clients you can do so.*

## Dataset

*Orders\_Master\_Data.csv*

The client's information provided in the data set is encrypted, nevertheless the data is real, so you must guarantee the correct use of it.

The granularity is in days, meaning each row represents a client on a specific date. The results must be aggregated, a row per client.

## Results

It is important that you execute an exploratory data analysis to make sure you understand the dataset and make some findings to conduct the prioritization process.

You should deliver:

*Capstone - Route Optimization PASCUAL*

- Table with each client and the new suggested monthly frequency and number of contacts in csv format
  - You may include other relevant information or suggestions
- Notebook of the model in *.ipynb* format
- Slides with EDA, findings, procedure, and results
- LLM tool that supports solution.