ESADE

CATEGORIZATION AND RANKING ANALYSIS

SUPERGLOVO

Capstone Project
Master in Business Analytics
Academic year 2018/19

Arianna Capella
Adrián Guell
Mengjie Tang

TEAM MEMBERS

Arianna Capella

Business administration studies, now pursuing a Master in Business Analytics.

- Data cleaning and analysis
- Visualization
- Categorization day to day strategy

Adrien Guell

Business administration studies, now pursuing a Master in Business Analytics

- Benchmark in categorization
- Best practices
- ML algorithm

Mengjie Tang

Business administration studies, now pursuing a Master in Business Analytics

- Ranking analysis
- Ranking benchmark
- Best practices
- Algorithm solutions

OUR OBJECTIVES

1

CATEGORIZATION

Analysis of why categorization is important and best practices in the market

2

GLOVO'S SITUATION and VISUALIZATION

Insight from the dataset in order to better understand Glovo's situation, through the use of python's analysis and visualization with Tableau

3

CATEGORIZATION and RANKING ANALYSIS

Analysis of the actual categorization and ranking methods used by Glovo, Identification of main problems and mistakes

4

RECOMMENDATIONS and CONCLUSIONS

Recommendations on how the categorization can be improved, both at an IT and business level

DATASETS

ORDERS' DATASET

Detailed order information:

- Order ids
- Customer ids
- Delivery information
- Products in the basket
- Delivery fee
- Ratings

PRODUCTS' DATASET

Product categorization data:

- Super-Collection
- Collection
- Sections
- Product name
- Prices
- Stores names

Dropping missing values

Eliminated rows with more than 5 missing values – those rows are not interesting for our purpose.

Dropping duplicates

Eliminated double rows in order to reduce the size and eliminate useless information.

Names unification

Lowered all names (collections, products..) and capitalized the first letter, in order to standardize them.

ID cleaning

Removed the characters element from the IDs , keeping only the numbers.

Europe and Latin America table

Filtered the orders' table according to the city. We created two tables, one for Europe (containing Barcelona and Madrid) and one for Latin America (containing Buenos Aires and Lima) in order to simply the analysis.

CATEGORIZATION

PRODUCT CATEGORIZATION

WHY IS IT SO IMPORTANT?

Your app's category tree is one of the primary ways shoppers browse your inventory to find products they're looking for. It allows you to onboard inventory more efficiently by categorizing your products faster and at a higher quality than even before.

WHAT ARE THE BENEFITS?

ENHANCE
Improve the speed at which people find products

CONVERT
Increase conversion rates and boost online sales

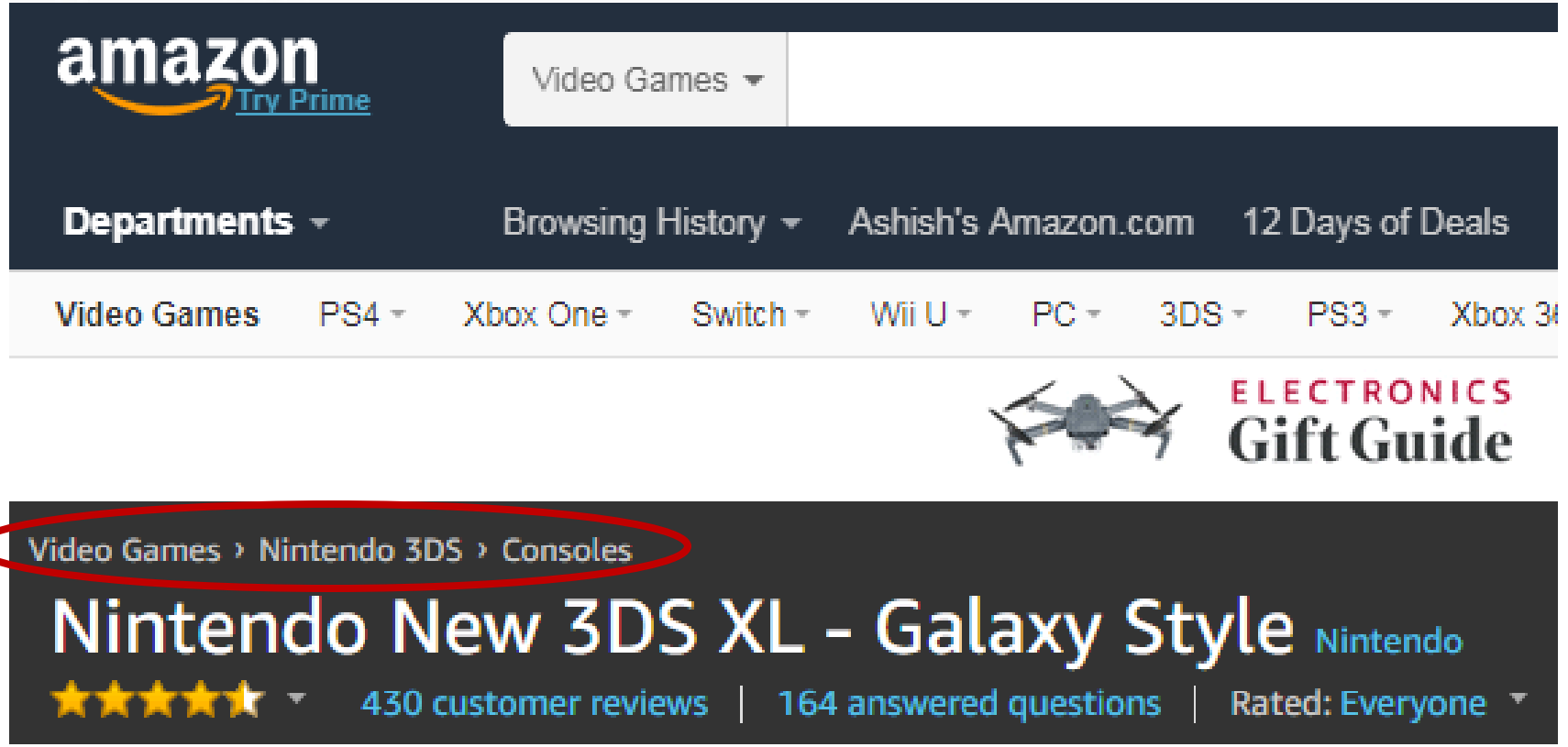
ORGANIZE
Establish an intuitive structure

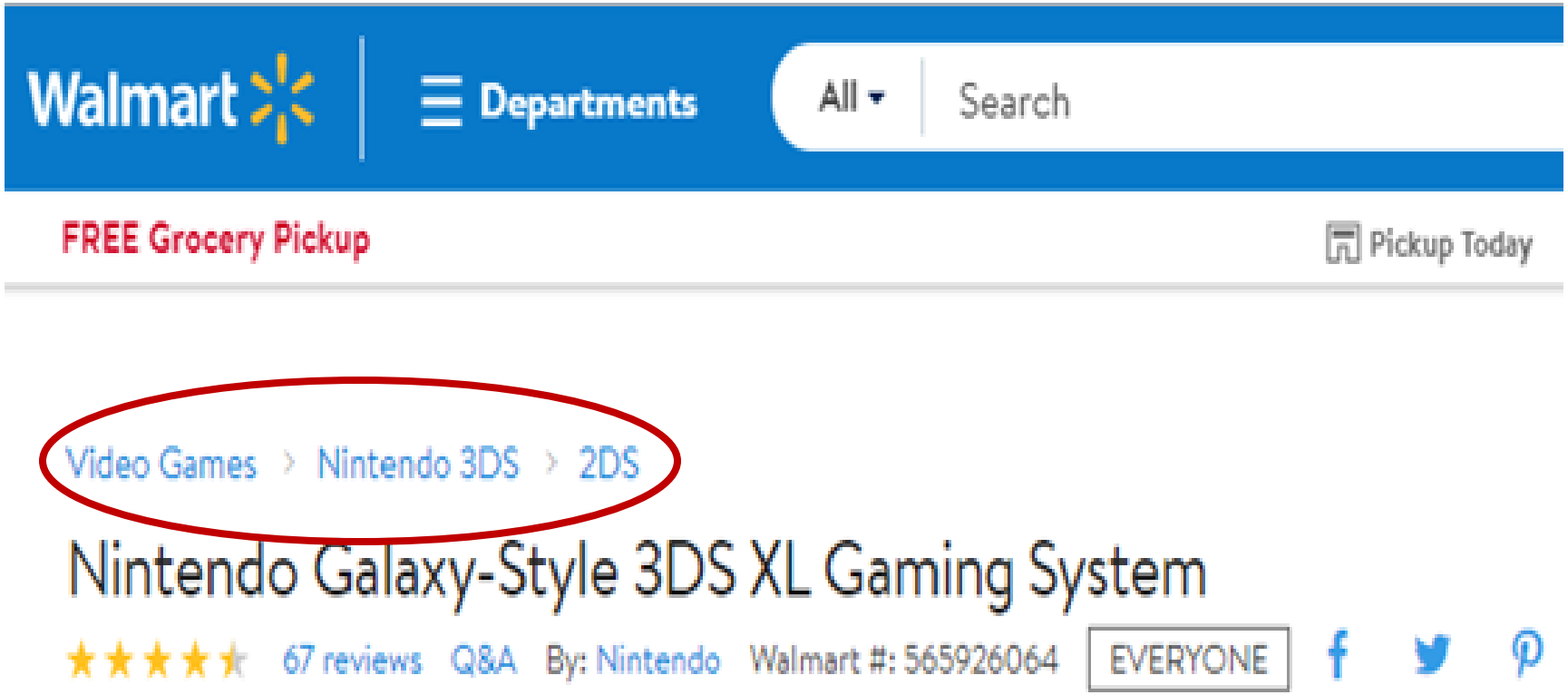
OPTIMIZE
Give search engines a clear hierarchy to find and index your pages

UNDERSTANDING THE PRODUCT CATEGORIZATION PREDICAMENT

- Therefore, product categorization is indispensable for commerce apps/websites. It makes free-text searches faster and provides better user experience by highlighting top categories upfront. Every retailer like Amazon, Walmart etc. tag all their products with a product category.

- Let's see what happens if we search, for instance, for a new Nintendo console at different ecommerce websites:


amazon
Try Prime
Video Games
Departments
Browsing History
Ashish's Amazon.com
12 Days of Deals
Video Games
PS4
Xbox One
Switch
Wii U
PC
3DS
PS3
ELECTRONICS
Gift Guide
Video Games › Nintendo 3DS › Consoles
Nintendo New 3DS XL - Galaxy Style
Nintendo
430 customer reviews
164 answered questions
Rated: Everyone


Walmart
Departments
All
Search
FREE Grocery Pickup
Pickup Today
Video Games › Nintendo 3DS › 2DS
Nintendo Galaxy-Style 3DS XL Gaming System
67 reviews
Q&A
By: Nintendo
Walmart #: 565926064
EVERYONE

Glovo

Best Buy › Video Games › Nintendo 3DS › Nintendo 3DS Consoles

Nintendo - New Galaxy Style New Nintendo 3DS XL

Model: REDSUBAA **SKU:** 5588518 **Release Date:** 09/04/2016 4.7 (5,551)

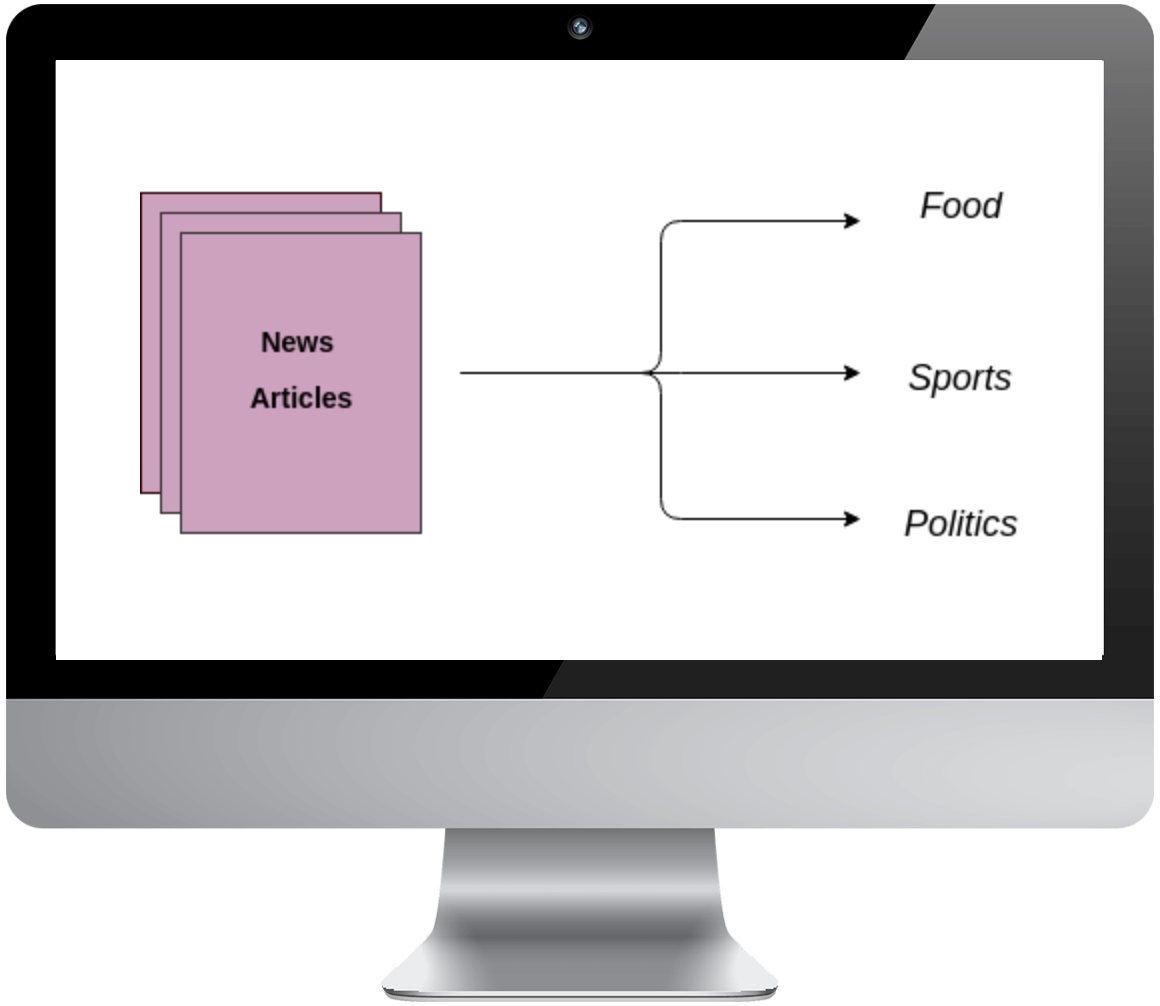
14 Questions, 41 Answers

While this categorization works well on retailer websites (they embrace less amount of product listings), it becomes an issue for ecommerce aggregators (which embrace millions of product listings). As is evident in the screenshots previously shown, the product categories have been defined differently by different ecommerce sites for the same product. This creates problems in assigning the exactly same product from different ecommerce sites to same categories of which the quality of search results and user experience suffer.

What is desired is that a particular product is categorized in exactly the same way across retailers. There are universal taxonomy such as Google taxonomy available which can be used to map the products to universal categories across ecommerce sites. But, mapping those categories manually would take an infeasibly huge amount of time plus is prone to human error.

Then, what is the optimal solution?

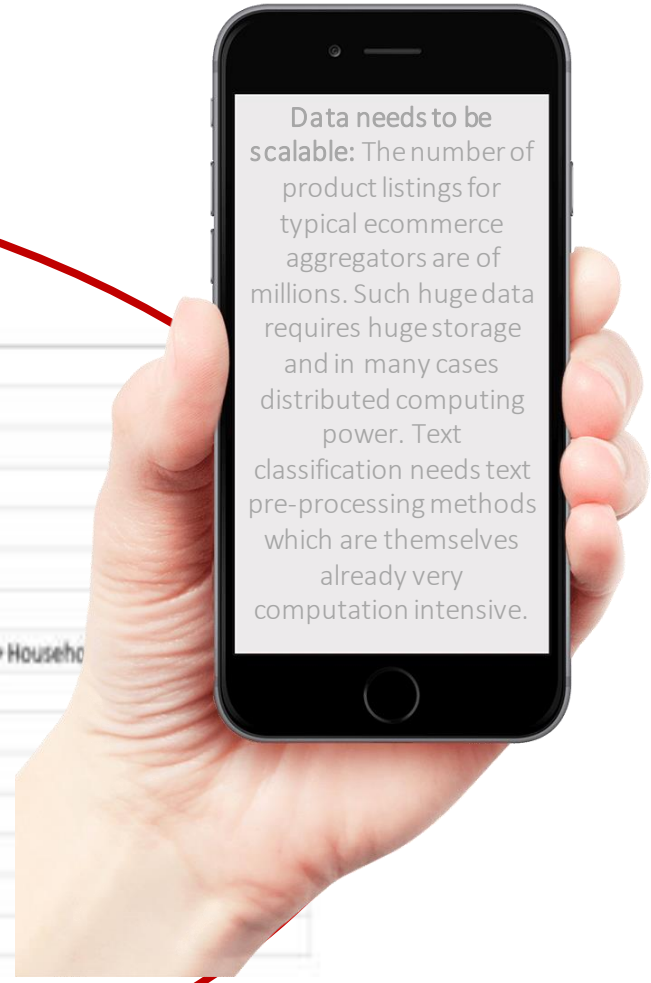
ML APPROACH FOR CATEGORIZATION ANALYSIS



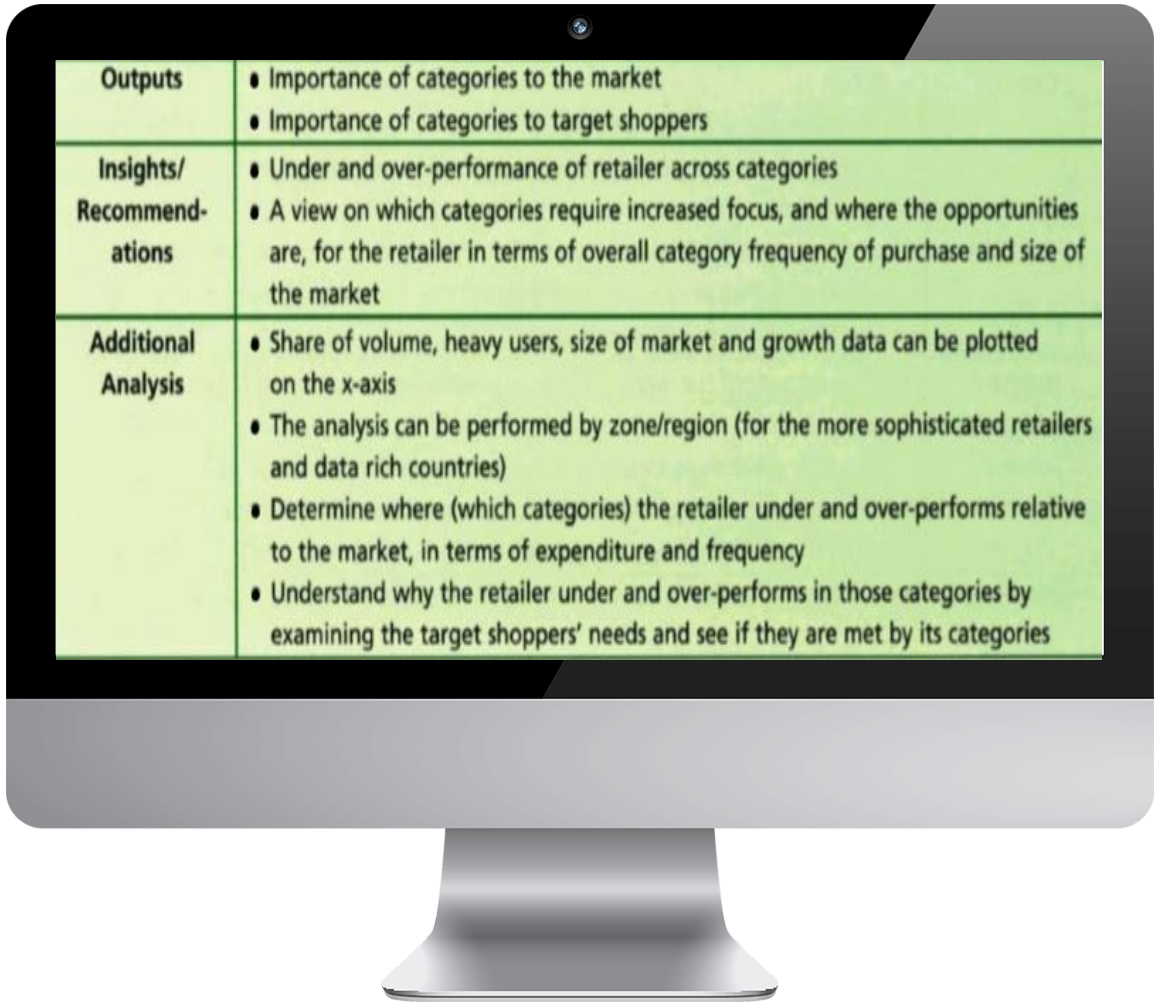
- The Machine Learning (ML) approach uses text classification methods to categorize each product to a category based on product title + description.
- For that purpose availability of training data it's a MUST. ML algorithms need training data to learn the right answers before they start predicting. What is more, that training set must obviously be properly categorized (using for instance the universal taxonomy) so that the algorithm learns from a non-biased dataset and makes category matches well.
- The manual approach is time-consuming and error prone as discussed earlier.
- In the following slide we can see an example of how the training data should look like:

Glovo

Product Title	Product Category
Psycho-Delic Dan by Catherine Holcombe Coaster	Health & Beauty > Personal Care > Hair Care > Combs & Brushes
Tumi Travel Accessory Extra Large Packing Cube	Luggage & Bags > Luggage Accessories
I Already Want To Take A Nap Tomorrow Womens T-Shirt Charcoal Gray Xlarge	Apparel & Accessories > Clothing > Shirts & Tops
Sterling Silver RESPECT THY MOTHER Word Necklace on an 18 inch Box Chain	Health & Beauty > Jewelry Cleaning & Care > Jewelry Boxes
Laguna 5 Drawer Chest	Furniture > Bedroom Furniture > Dressers
Villa Borghese Canvas Art - (36 x 24)	Home & Garden > Decor
Ara Reversible Small Storage Basket Color: Sand/Taupe	Home & Garden > Household Supplies > Storage & Organization > Household
Zodaca Women Sexy Hot Lingerie Lace Piece Suit Underwear Sleepwear with Eye Mask - Black	Health & Beauty > Personal Care > Sleeping Aids > Sleep Masks
Duvet Cover Size: King	Home & Garden > Linens > Bedding > Duvet Covers
Osram Bare Lamp For Panasonic PT-52LCX15B / PT52LCX15B Projection TV Bulb DLP	Outdoor > Lighting > Projections
Cuddle Angora Platinum Handmade Throw Blanket	Home & Garden > Linens > Bedding > Blankets
Timberland Men's Mumford River Camo Chambray Shirt - Large - Gargoyle YD	Apparel & Accessories > Clothing > Shirts & Tops
Long-Sleeve Fold-Over Off-the-Shoulder Sweater	Apparel & Accessories > Clothing Accessories
ugly Christmas Sweaters - Women's Bells Are Ringing Sweater by Tipsy Elves	Home & Garden > Decor > Seasonal & Holiday Decorations



CATEGORY MANAGEMENT



Outputs	• Importance of categories to the market • Importance of categories to target shoppers
Insights/ Recommend-ations	• Under and over-performance of retailer across categories • A view on which categories require increased focus, and where the opportunities are, for the retailer in terms of overall category frequency of purchase and size of the market
Additional Analysis	• Share of volume, heavy users, size of market and growth data can be plotted on the x-axis • The analysis can be performed by zone/region (for the more sophisticated retailers and data rich countries) • Determine where (which categories) the retailer under and over-performs relative to the market, in terms of expenditure and frequency • Understand why the retailer under and over-performs in those categories by examining the target shoppers' needs and see if they are met by its categories

- Once the categorization process is fully automated thanks to ML, we need to think how to successfully capture business benefits out of it (find those categories which are important to the market and to their target shoppers). How?
  - Developing an understanding of which categories are being purchased in the market.
  - Where possible, further refine the analysis to take into account of the retailer's target shoppers, and hence to identify strategic categories for these target shoppers.

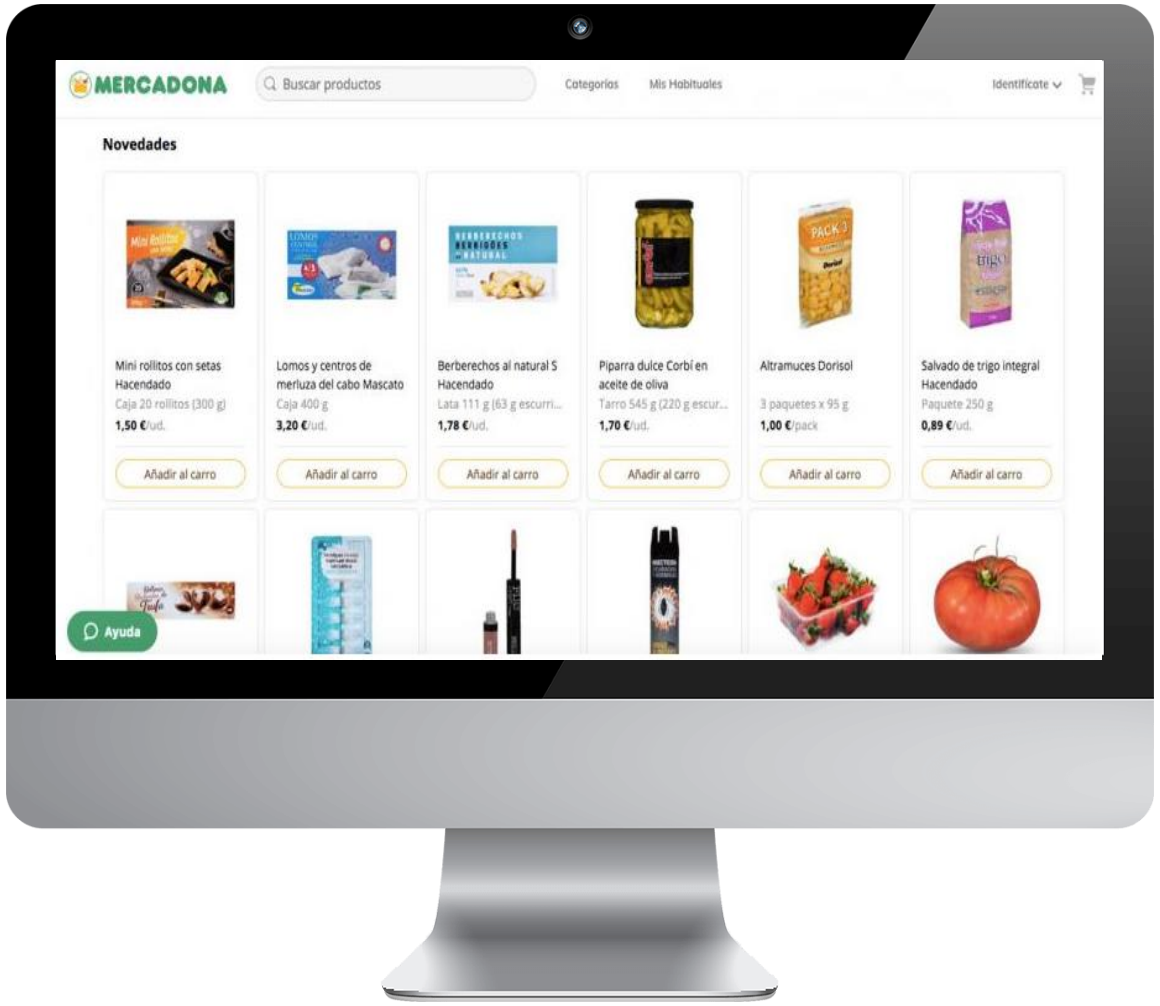
MERCADONA' BENCHMARK ANALYSIS

- Mercadona is one of the biggest companies in Spain (with annual sales of more than 25b €) and the biggest supermarket chain in the country (with an estimated market share of 23%), followed, by far, by Carrefour and DIA.
- The company has recently come up with a new website, which only operates in Valencia so far. Therefore, users not living in Valencia will be directly redirected to the old website.
- This Benchmark analysis is aimed to compare the old vs the new website (specially to analyze the categorization strategy implemented in the new one).
- In the next slides we will go step by step comparing the experience in each website:

MERCADONA' BENCHMARK ANALYSIS

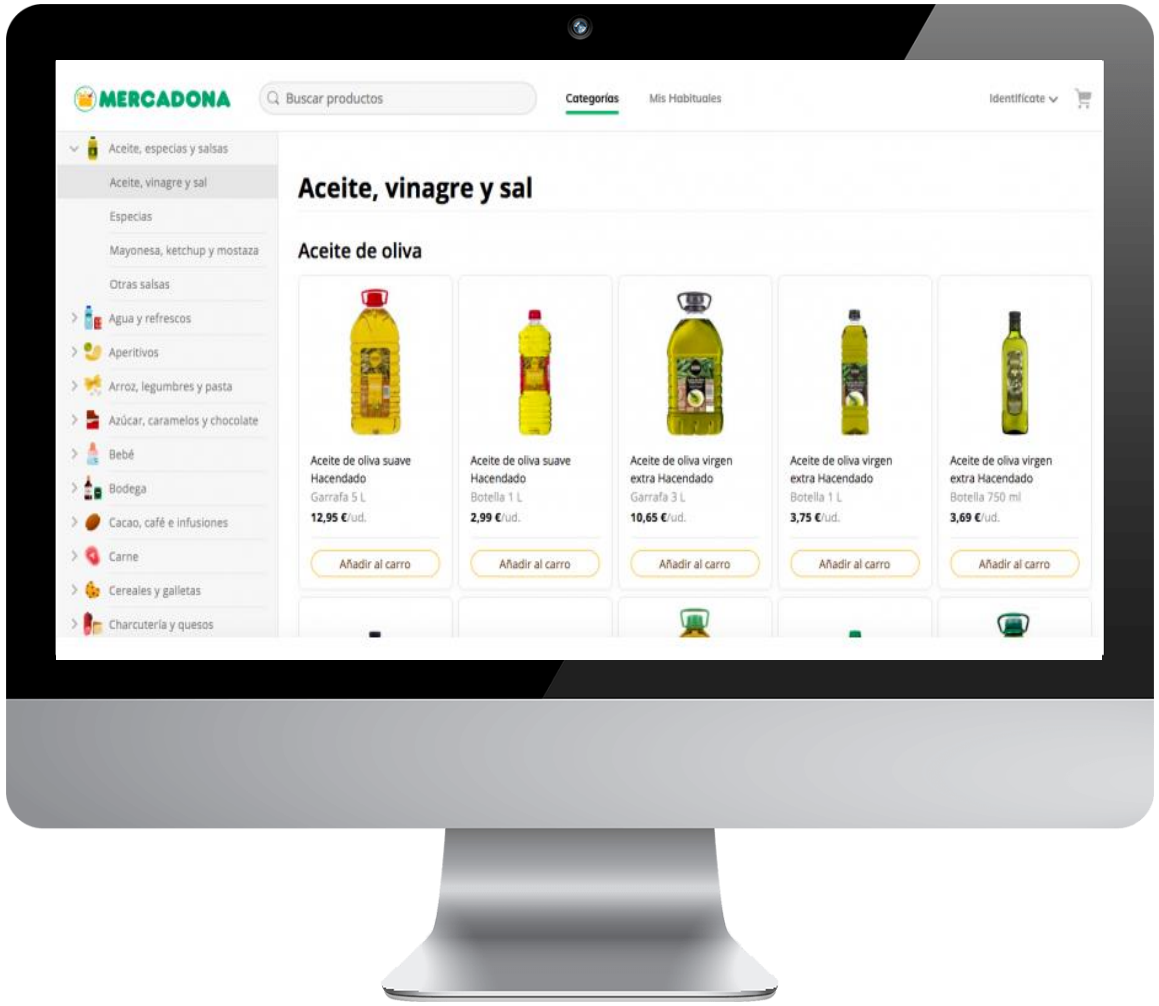
- When you first land on the Mercadona website, you will see a simple, clear and even beautiful homepage. You will then have to enter your zip code: if you are from Valencia you will continue with the new website and, as we have said, if not, you will be redirected to the old one.
- If you stay on the new one, once you enter your zip code, it will ask for your name, email and password, which you will have to confirm later through an email and that's it. From there you will be able to move around the web and start buying.
- You will have to spend more than 50€ to be able to take your basket home (plus the shipping costs)
- Let's see how the new website looks like:

MERCADONA' BENCHMARK ANALYSIS



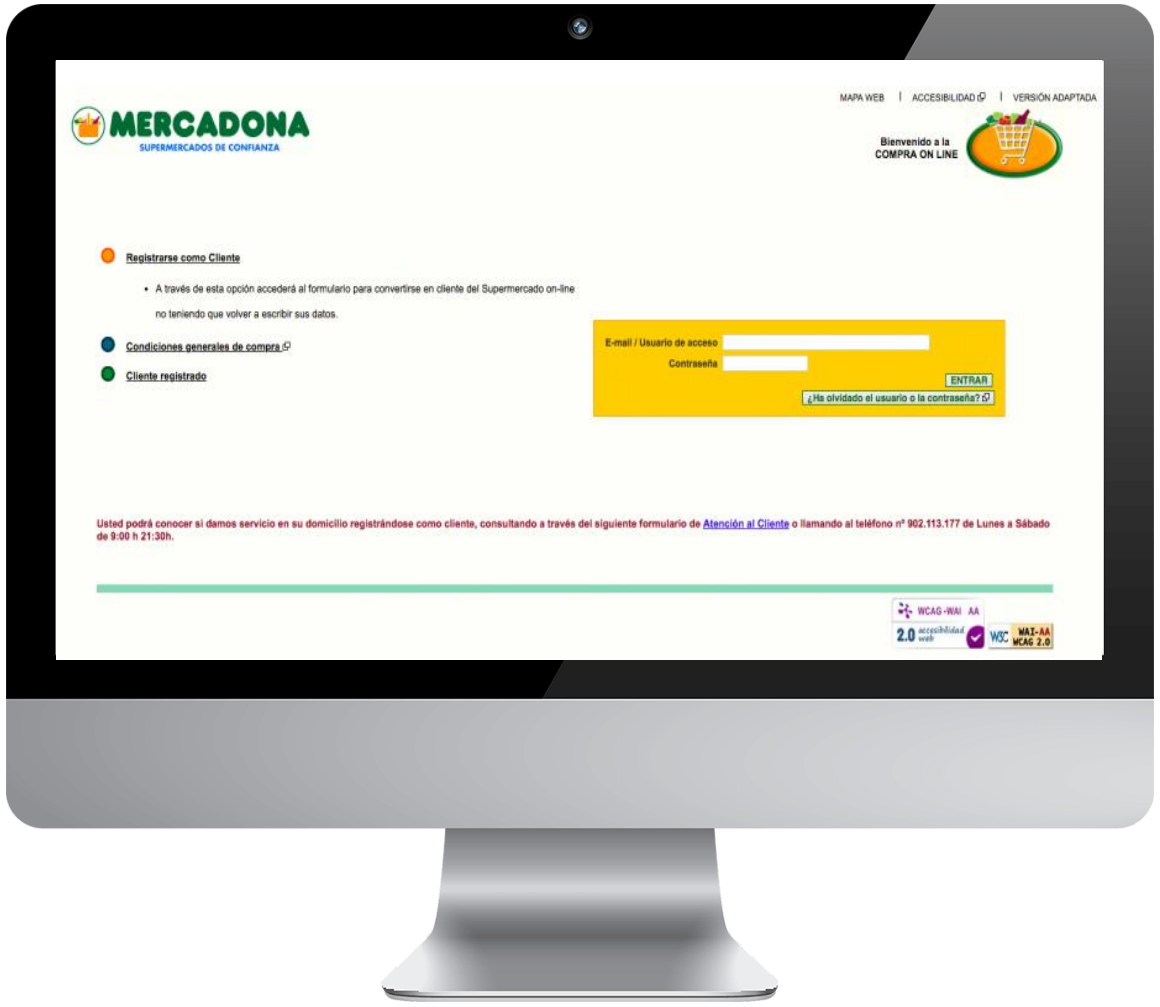
- If you are lucky enough to be able to make your purchase through the new website, you will see how it stands out as soon as you enter it (after entering the postal code). Specially for two sections: "Novedades"(where they tell you about new product offerings) and "Mis habituales" (where you will keep all the things you buy normally through the web.
- You will also have the option to look for exactly what you wish by going to the search button at the header of the website. By selecting a product there, you will not only find that product but also the ones related to it (some recommendations).
- Let's see how to find the different product categories:

MERCADONA' BENCHMARK ANALYSIS



- You will find the different categories in the dropdown of the same name (just at the right of the Search button).
- You will see how this section is perfectly ordered, super easy to find and very intuitive within the different subcategories. Products with high resolution photos, detailed descriptions, option to add it to the basket without having to get into each item …
- You can also see how many things you have in the basket by checking the icon located at the upper right part of the screen, so we don't have to constantly get into it.
- Let's see how the old website looks like instead:

MERCADONA' BENCHMARK ANALYSIS



- Just by landing into the old website (if we are not from Valencia) we can realize of the huge difference… (it looks we went back 20 years from now).
- Also, when accessing the web, instead of only asking for your name, address and password, it asks you for everything (name, two surnames, ID, street, phone … which slows down the buying process.
- Let's see what happens once we fill up all the required info and get to the Home page:

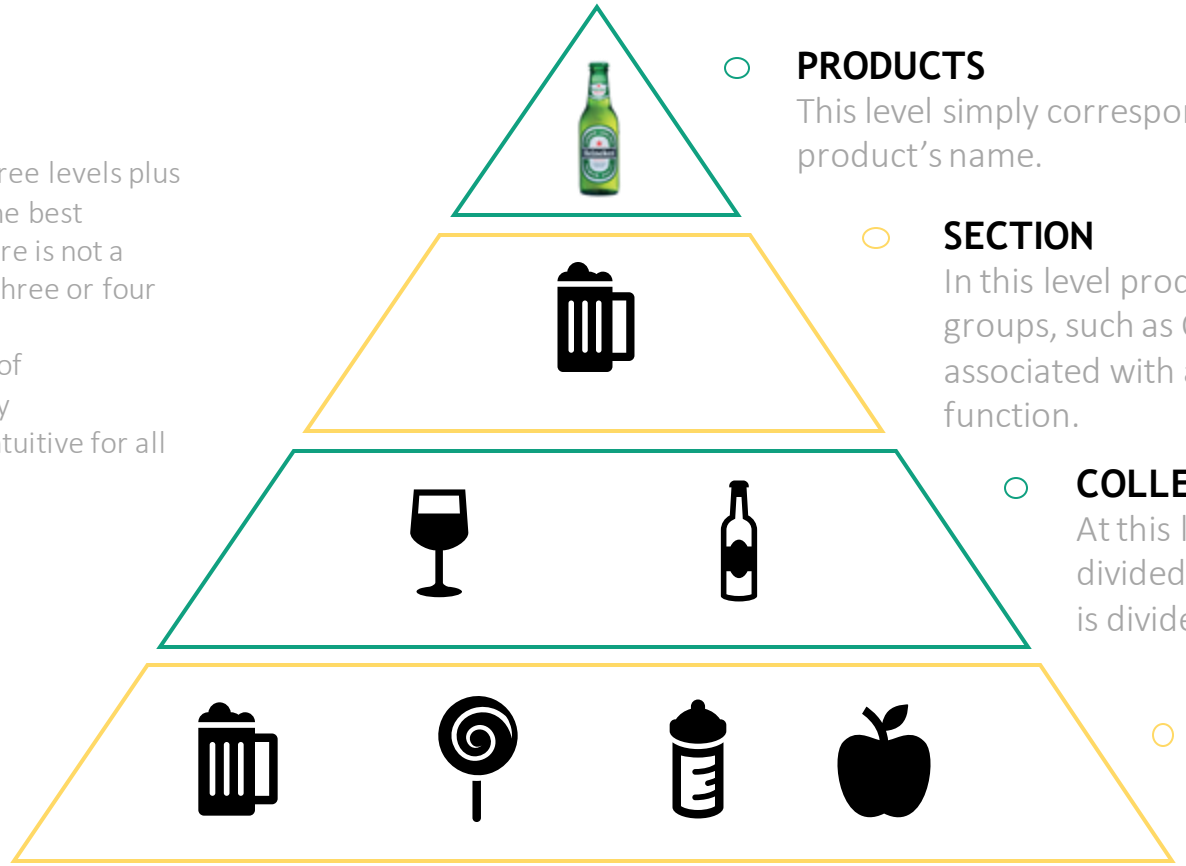
MERCADONA' BENCHMARK ANALYSIS

- First of all, we believe that it is already weird that they make us a shopping guide ….
- Then, we have again the different categories in a dropdown on the left side of the screen. But nothing to do with the previous categorization. The categorization as a whole is worse, not intuitive at all, a thousand sub-sections, no photos that tell you what you are buying beyond a short description and not detailed description.
- We can also see "an ugly" ticket at the right side to show us what we are buying.
- So, to conclude, it seems pretty obvious the benefits of the new vs the old Mercadona's website (by just looking at the screenshots you can already see the huge difference between them).

GLOVO'S CATEGORIZATION

CATEGORIZATION

Glovo categorization is based on three levels plus the product names. According to the best practices we have investigated, there is not a perfect number of categories, but three or four are considered the best options.
The most important characteristic of categorization is that must be easily understandable by everyone and intuitive for all different kind of clients.



PRODUCTS
This level simply corresponds to the product's name.

SECTION
In this level products are grouped in smaller groups, such as Colas inside sodas. Could be associated with a brand as well or a specific function.

COLLECTION
At this level thee super-collections are divided into smaller groups, such as drinks is divided in beer, sodas, water..

SUPER-COLLECTION
This is the biggest category, it divides products into groups such as drinks, meat, hygiene products…

CATEGORY STRATEGY

RETAILER STRATEGY

Consist in defining the retailer strategy of Glovo. Identify the drivers of the strategy and perform analysis.
To define the category is important to understand which are the retailer's target consumers in order to create categories which target the right profiles.

DEVELOP CATEGORY PLAN

It's an 6 steps process in which, according to the insight gained from the analysis of the previous step, it is possible to identify the best category for the retailer purpose.
In this case we performed an analysis of the existing one, proposing some changes and improvements.

IMPLEMENT CATEGORY PLAN

This is the phase of the implementation. Consist in applying the new categorization to the existing products. It is often difficult for companies such as Glovo because categorization is linked to the supplier's one.

REVIEW CATEGORY PERFORMANCE

This task should be performed continuously in order to evaluate whether the new categorization is efficient and to control that everything is working properly. In addition is important to control that new products are correctly classified since the beginning.

RETAIL STRATEGY

	Insight
Who are the retailer's target shopper and is the retailer reaching them?	• Young tech-savvy • Young workers without a lot of time • Young/new families
Who is the retailer's competition for target shoppers?	• Ulabox • Deliberry
What opportunities exist to improve against key competitors ?	Glovo is already better than competitors under many aspects. Glovo offers cheaper and faster delivery, it does not require a minimum basket amount and offers products from multiple supermarkets all in the same app.
What are the strategic categories in the market?	• Promo → for people economical concerned, would be good to always have a promotion section • Local products → young professionals and in general new generations are attached to the environment, both for health and ecological reasons, so they prefer local products • Fresh → health conscious people like daily fresh products

CATEGORY PLAN

CATEGORY DEFINITION

Definition of the categories level. Glovo has already 4 levels categorization, with Super-collections, collections, sections and products.

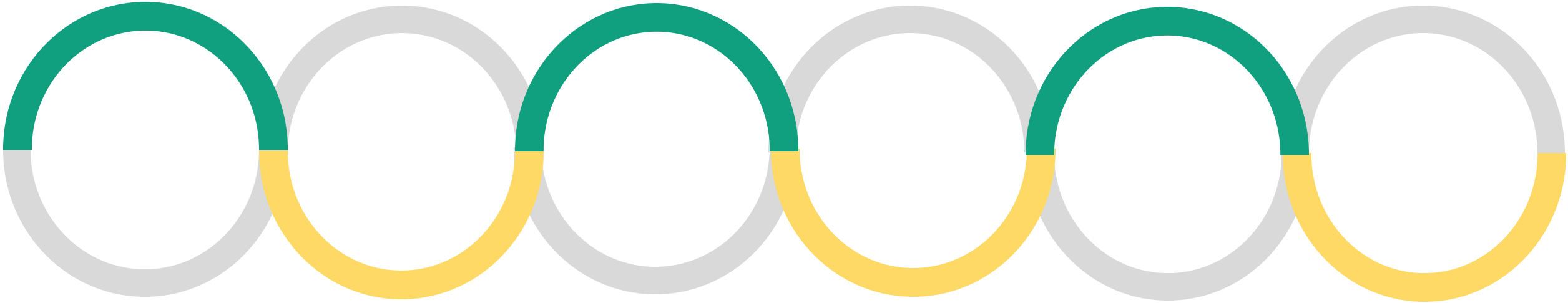
CATEGORY ASSESSMENT

Assess whether the categories are matching the target shoppers. This could be checked in different ways:

- Questionnaires to shoppers
- Comparing the performance of the old categories with the one of the new ones
- Studying competitors categories

CATEGORY STRATEGIES

The strategy of the categories should be in line with the previous analyzed retailer strategy.



CATEGORY ROLE

The role of Glovo's categories is DESTINATION (It means categories are created on the base of which products is contained inside). As we will suggest later on, the creation of CONVENIENCE and ROUTINE categories could be useful.

PERFORMANCE MEASURES

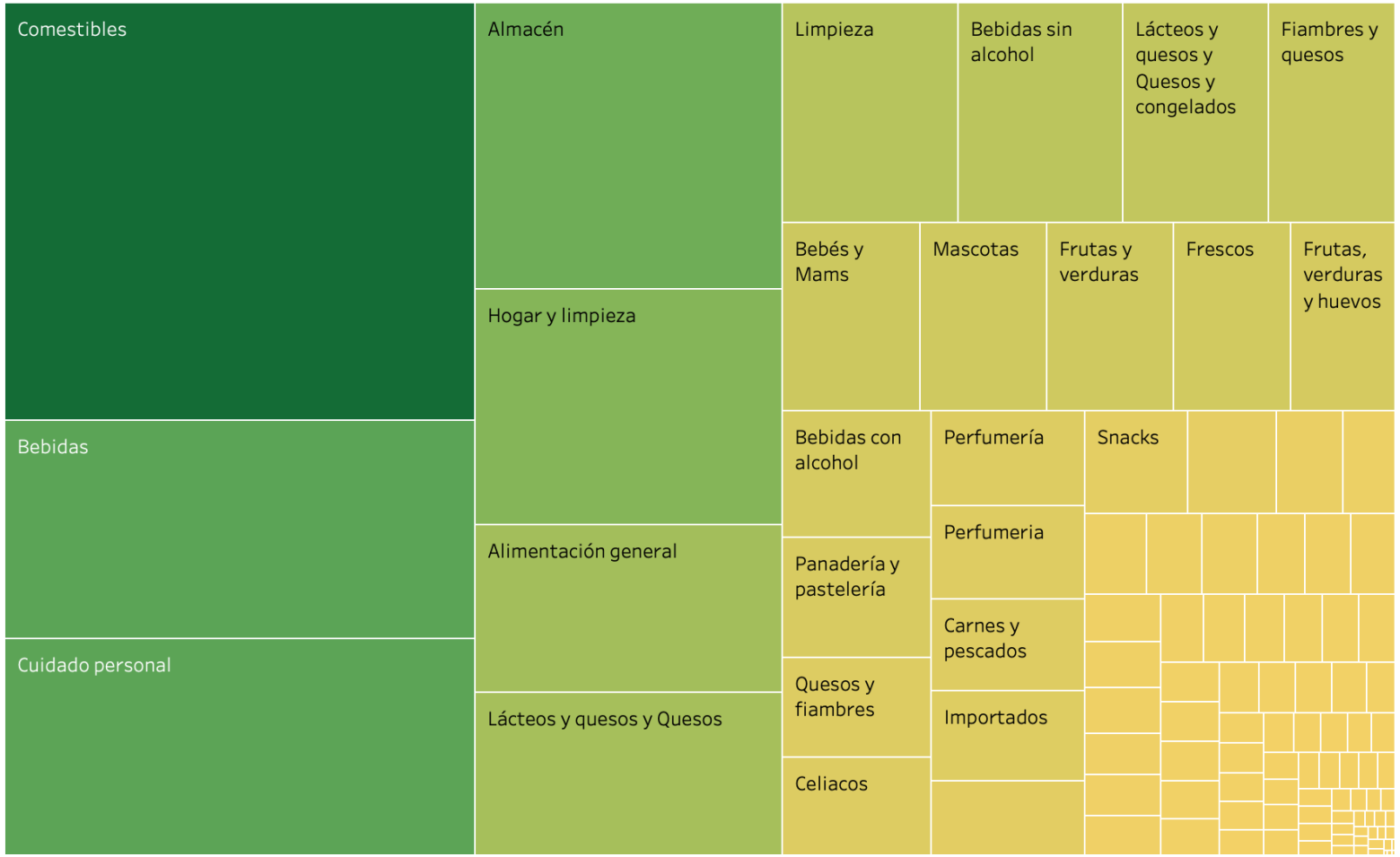
This step has to be performed continuously in order to evaluate how the categories are performing. This can be done thanks to metrics that can be found in the conclusions.

CATEGORY TACTICS

This step consists in the tactical mix to satisfied the target shopper. Include the study of pricing, promotions and merchandising.

DATASET VISUALIZATION

SUPER-COLLECTION DISTRIBUTION – EUROPE AND LATIN AMERICA TOGETHER



Collection Group. Color shows count of Collection Group. Size shows count of Collection Group. The marks are labeled by Collection Group.

Count of Collection Gro..
1 11.159

MANY SUPER COLLECTIONS

The graph represents all the super-collections in Europe and Latin America

DISORGANIZED

Confusing and disorganized, same super-collection with different names

BIGGEST CATEGORIES

Size is shown according to the count of collections inside each super-collection→ logically the biggest categories are the most vast (Comestibles and Bebidas)

EUROPE

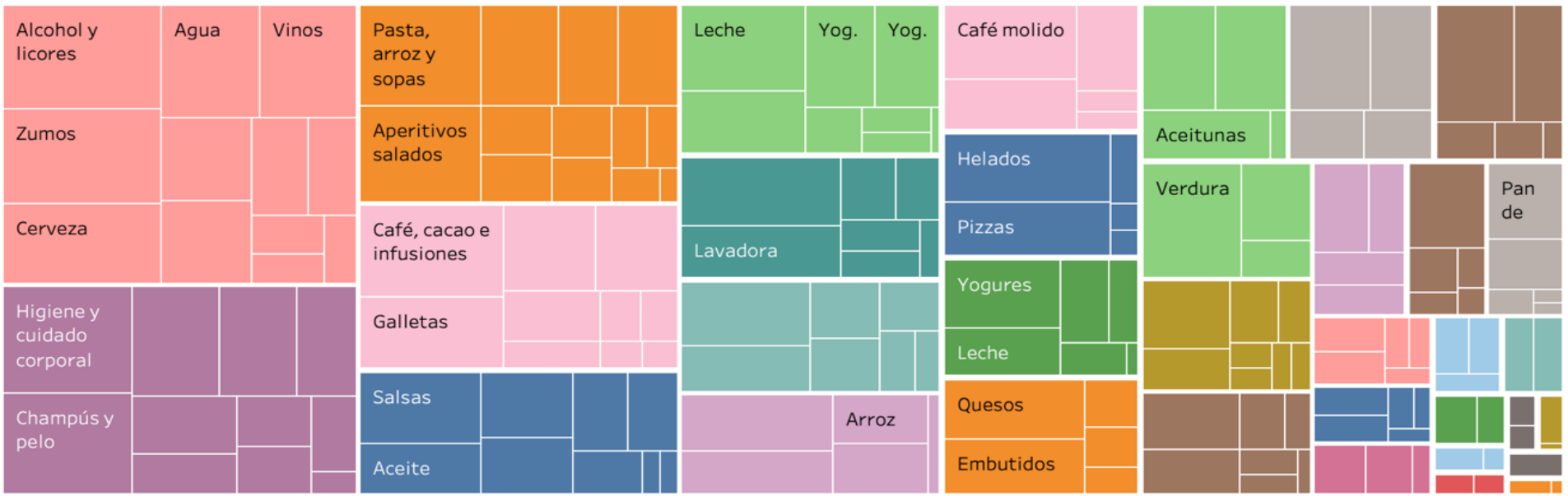
BARCELONA & MADRID

Revenue by super-collection (EU)

Categories hierarcy

Collection Group	Collection	
Aceites y condimentos	Aceite	Abc
	Especias	Abc
	Harinas y levaduras	Abc
	Ketchup, mostaza y mayo..	Abc
	Masas y hojaldres	Abc
	Sal y azúcar	Abc
	Salsas	Abc
	Tomate frito	Abc
	Vinagre	Abc
Agua	Con gas	Abc
	Sin gas	Abc
Alimentación general	Aceite y vinagre	Abc
	Aperitivos salados	Abc
	Conservas	Abc
	Dietéticos	Abc
	Especias	Abc

Collection by super-collection in EUROPE



INSIGHT EUROPE

34 SUPER-COLLECTIONS

In Europe we can find 34 super-collections, while in the app in Barcelona you can only find 17.

BEST PERFORMANCE

Best selling Super-collection by revenue is Bebidas, followed by Bebes, Mamas y Higiene.

BIGGEST SUPER-COLLECTION

The best performing super-collection by revenue corresponds also to the biggest super-collection by size (Number of collections contained).

REAPETED PURCHASES

Most of the customers buy more than once. The average amount of purchases per consumer is almost 12, with a maximum of 1721 orders and a minimum of 1.

BASKET SIZE

The average price of the basket is 12,69 €. The biggest order account for over 500 €, while the small one, few cents.

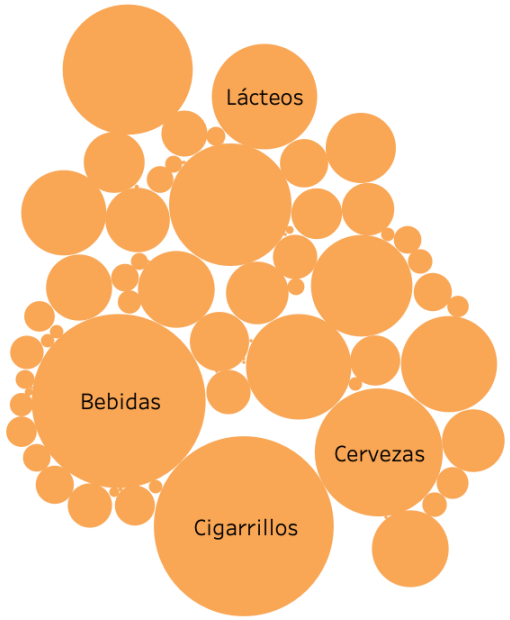
DELIVERY TIME

The average delivery time for Europe is 40 minutes (as we can also notice most of the ETA is around 30 min). Higher delivery time is 5 hours.

LATIN AMERICA

LIMA & BUENOS AIRES

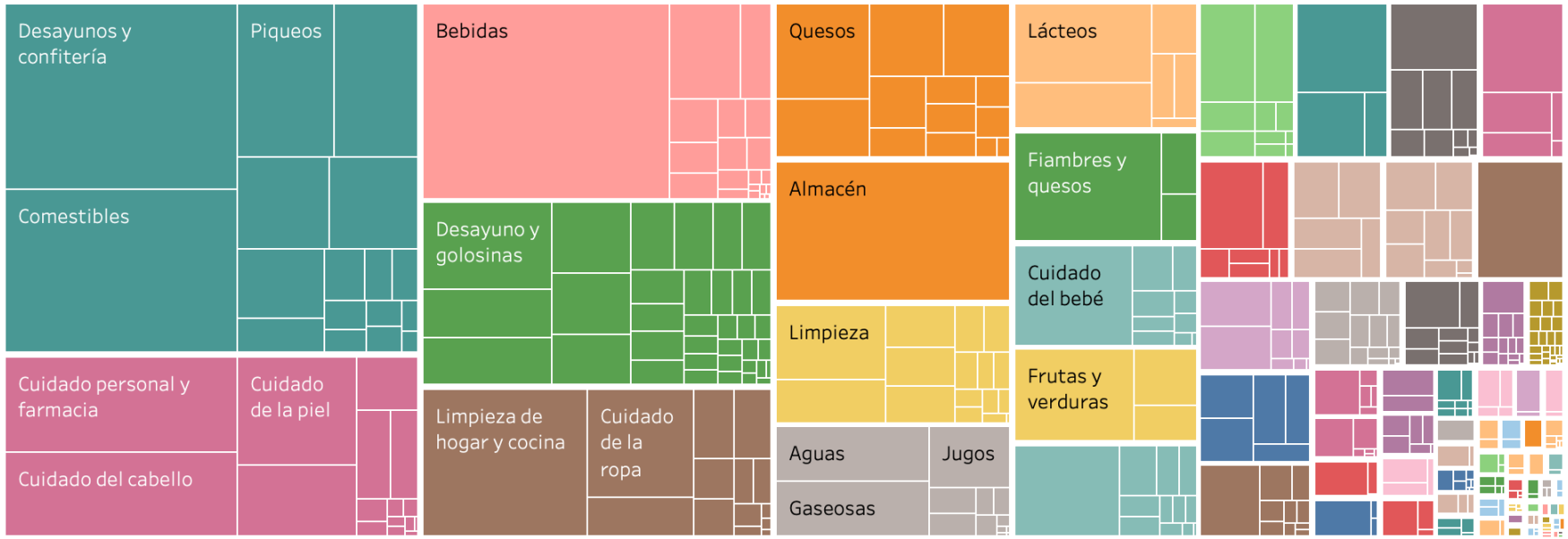
Revenue by category (LATA)



Categories hierarcy

Collection Group	Collection	
Alimentación general	Almacén	Abc
Alimento húmedo	Alimento húmedo (gatos)	Abc
	Alimento húmedo (perros)	Abc
Almacén	Aceites y aderezos	Abc
	Aceites y vinagres	Abc
	Aceitunas y encurtidos	Abc
	Aderezos	Abc
	Aderezos, salsas y untabl..	Abc
	Arroz y legumbres	Abc
	Conservas	Abc
	Conservas y encurtidos	Abc
	Desayuno y golosinas	Abc
	Desayuno y merienda	Abc
	Elaborados	Abc
	Endulzantes	Abc
	Fideos, arroz y legumbres	Abc

Collection by super-collection in LATA



INSIGHT LATIN AMERICA

82 SUPER-COLLECTIONS

In Latin America we can find 82 super-collections, while in the app in Lima you can only find 12.

BEST PERFORMANCE

Best selling Super-collection by revenue is Cigarillos, followed by Bebidas.

BIGGEST SUPER-COLLECTION

The best performing super-collection by revenue corresponds also to the biggest super-collection by size (Number of collections contained).

REAPETED PURCHASES

Most of the clients use the app multiple times to order products. The average number of orders is 6, while the maximum is 876 and minimum 1.

BASKET SIZE

The average price of the basket is 116,69 pesos. The biggest order account for over 7444, while the small one, few cents.

DELIVERY TIME

Average delivery time in Latin America is 37 minutes. Average ETA shown in this region is 30-35 minutes. Although there are some exceptions, indeed maximum wait time is 10 hours.

CONSUMER BEHAVIOUR

SMALL ORDERS

Most of the orders, especially in Latin America, are small. Average order size in Latin America is slightly higher than 2, while in Europe is 4.

NO FAMILY GROCERIES

Looking at the most sold products (or categories) we can assess that the typical SuperGlovo clients are not families, but mostly young people.

LOYAL CLIENTS

As seen analyzing the average number of purchases per consumers, we can notice that most of the clients are loyal, and purchase repeatedly on SuperGlovo.

NOT PRIME CLIENTS

Most of the consumers, both in Europe and Latin America are not prime. This can be explained by the low delivery fees that are applied also for non prime consumers.

ESTIMATED ARRIVAL TIME

Most of the products ordered have an estimated arrival time of 25/35 minutes or 30/35 minutes.

CATEGORIZATION PROBLEMS

- **No homogeneous categorization** between the different continents(Latin America and Europe) and even among different countries inside the same continent (Madrid has different categories compared to Barcelona).
- **Many similar super-collections** (such as Bebes and Bebes y Mamas or Bebidas, Bebidas sin alcohol and Bebidas con alcohol..) which contains the same collections and even the same products. → Confusing for the consumers
- **Huge differences among the size of super-collections and collections** (sizes vary from 2 to 40) and huge differences in products inside the same super-collections (vary from 1150 products inside comestibles to 1 inside Vodka..)
- **Sometimes inconsistent content** → products names are not clean, we can find same product with different names and different prices, even in the same city.
- **Unscalable content** → since the products are not well organized, impossible to automatically categorized new products when inserted.
- **Missing information** or descriptions on products --> impossible to apply a Machine Learning algorithm (either random forest or clustering)
- **Presence of confusing categories** such as Carnes y Embutidos, Carnes y pescados, Carnicería y pescadería, in which consumers don't know where they can exactly find the products.
- **Many sections with only one product** inside, which is confusing for both suppliers and for consumers. Sections shouldn't be too specific.

EXAMPLES OF TRANSFORMATIONS PROPOSED

BEBIDAS

Bebibas

Bebidas con Alcohol

Bebidas sin Alcohol

Bebidas con Alcohol

Bebidas sin Alcohol

CARNES, PESCADOS…

Carnes y Embutidos

Carnes y pescados

Carnicería y pescadería

Carnes

Pescados

Embutidos

CIGARILLOS, CIGAROS..

Cigarrillos

Cigarrillos y Preservativos

Cigarros

Cigarros y Complementos

Cigarros y Preservativos

Cigarillos

Cigarros

Preservativos

WHY REORGANIZATION?

As alternative of adopting a Machine Learning algorithm as suggested in the conclusions, would be possible to complete a manual reorganization of products in order to simplify the purchasing process. These are few examples we decided to examine in order to clearly explain what we mean by transformations.
The three categories for Bebidas, could be reduced to two, avoiding confusions.
Same goes for the categories relative to meat, fish and salami and the ones related to cigars, cigarillos.. many mixed categories create confusion on where the desired product is actually available; by creating 3 separate super-collections the problem would be solved.

PRODUCTS' NAMES

1

ONLY ONE NAME PER PRODUCT

Instead of having Coca Cola 33cl, Coca Cola 1l, Coca Cola 2l.. Just have one COCA-COLA product in order to avoid confusions and to classify the product in an easier way

2

INSERT POSSIBILITY TO CHOSE THE SIZE AFTERWARDS

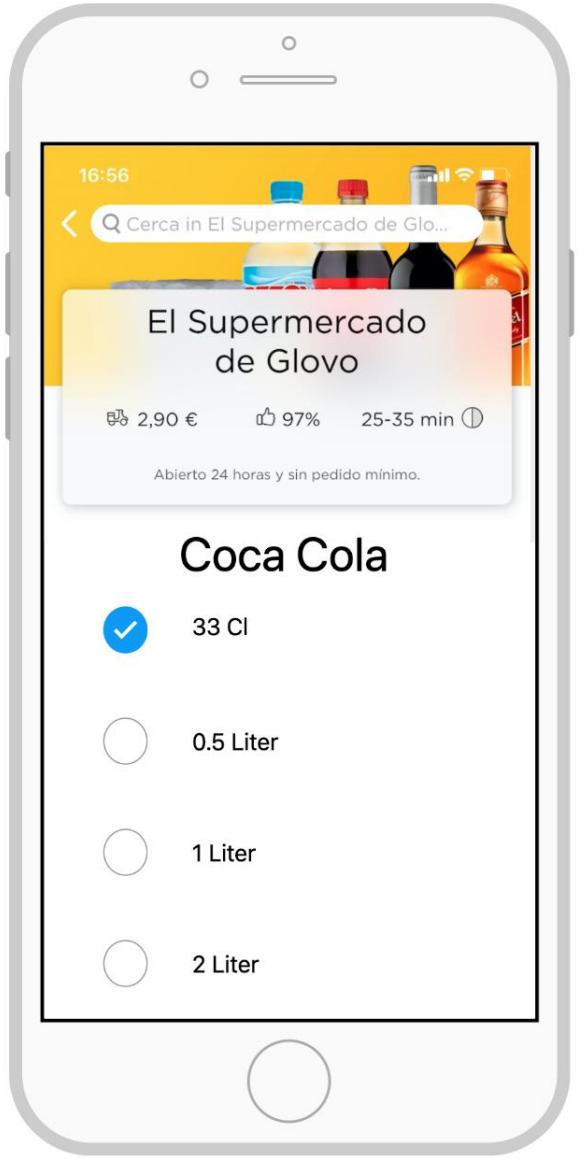
After selecting the product the user will be able to select the product size/quantity he desires.

EASIER PRODUCT CLASSIFICATION

This transformation will result in decreasing the number of products name and in avoiding having the same product in different size divided in different category

BETTER CONSUMER EXPERIENCE

The consumer will have a better consumer experience, it will be easier to find the products and easier to understand the size, since in many cases now the name of the product is not entirely shown because of space missing therefore the size is not easily readable



RANKING

WHY RANKING MATTERS?

The order in which the products are shown is always important in the customer journey. Most of the time, especially in the case of buying from the smartphone, consumers tend to focus on the first products shown, to be faster and reduce the time spend on the app. For these reasons, raking in of vital importance for Glovo. The main goal is always to **maximize the profits per consumers,** indeed by making easier for consumers to find products and by showing before determinate products, the consumer will spend more, being more satisfied, thus **increasing the visibility of the app**, the relevance in the market and, finally **increasing the conversion rate.**

Categorization

Rating

Sponsored Placement

Promotions

Membership

Best Seller

amazon

All drink

24/7 Customer service. Learn more

Departments Browsing History 's Amazon.com Today's Deals Gift Cards Registry Sell Help

EN Hello Account & Lists Orders 0 Cart

,000 results for "drink"

Sort by: Featured

Books
- Gastronomy History
- Alcoholism Recovery
- Sociological Study of Medicine
- Self-Help
- Women's Studies
- See more

Grocery & Gourmet Food
- Bottled Iced Tea
- Fruit Juice Beverages
- Juice Smoothie Drinks
- Almond Milks
- Flavored Drinking Water
- See more

Kindle Store
- Alcoholism
- Health, Fitness & Dieting
- Women's Studies
- Teen & Young Adult Biography eBooks
- Memoirs
- See more

See All 28 Departments

Customer Review
- & Up
- & Up
- & Up
- & Up

Brand
- Suja
- Capri Sun
- GoodBelly
- Califia Farms
- Drinking
- Drinks
- Good Karma
- International Delight
- Izze
- Kool-Aid
- Lipton
- Sparkling ICE
- bai
- LifeWay
- Gatorade

Book Language

SPONSORED BY SANPELLEGRINO FRUIT BEVERAGES

Shop the delightful fruit flavors of Sanpellegrino

Shop now›

Sanpellegrino Pomegranate 2,067

Sanpellegrino Prickly Pear and 2,067

Sanpellegrino Orange 2,067

Ad feedback

Drink: The Intimate Relationship Between Women and Alcohol
by Dowsett Johnston, Ann
122
Paperback
$13.38 $16.99
Get it as soon as Fri, May 24
More Buying Choices
$2.35 (103 used & new offers)
Other formats: Kindle , Audible Audiobook , Hardcover

Best Seller

24 Ounce Cocktail Shaker Bar Set with Accessories - Martini Kit with Measuring Jigger and Mixing
1,903
$16.95
Eligible for Shipping to Spain
More Buying Choices
$14.95 (7 used & new offers)

Drinks
Jul 6, 2016
91
App
FREE
Available instantly on compatible devices.

Drink In My Hand
by Eric Church
49
MP3 Music
Listen with Music Unlimited Or $1.29 to buy MP3
Or $1.29 to buy MP3

I'm Just Here for the Drinks: A Guide to Spirits, Drinking and More Than 100 Extraordinary
by Sother Teague and Robert Simonson
222
Hardcover
$14.68 $24.99
Get it as soon as Fri, May 24
More Buying Choices
$9.05 (59 used & new offers)

Drink your food, chew your water
Closed Caption
1
Prime Video
from $0.99 to rent
Or $0.00 with a Prime membership

Best Seller

REBUILD RANKING CRITERIA

QUERY

Popularity
Conversion Rate
Number of searching outcomes

USER PREFERENCE

Price
Brands
Location
Historical order

QUERY ⟷ DEAL/ PROVIDERS

Contest similarity,
Intention similarity,
Product similarity…

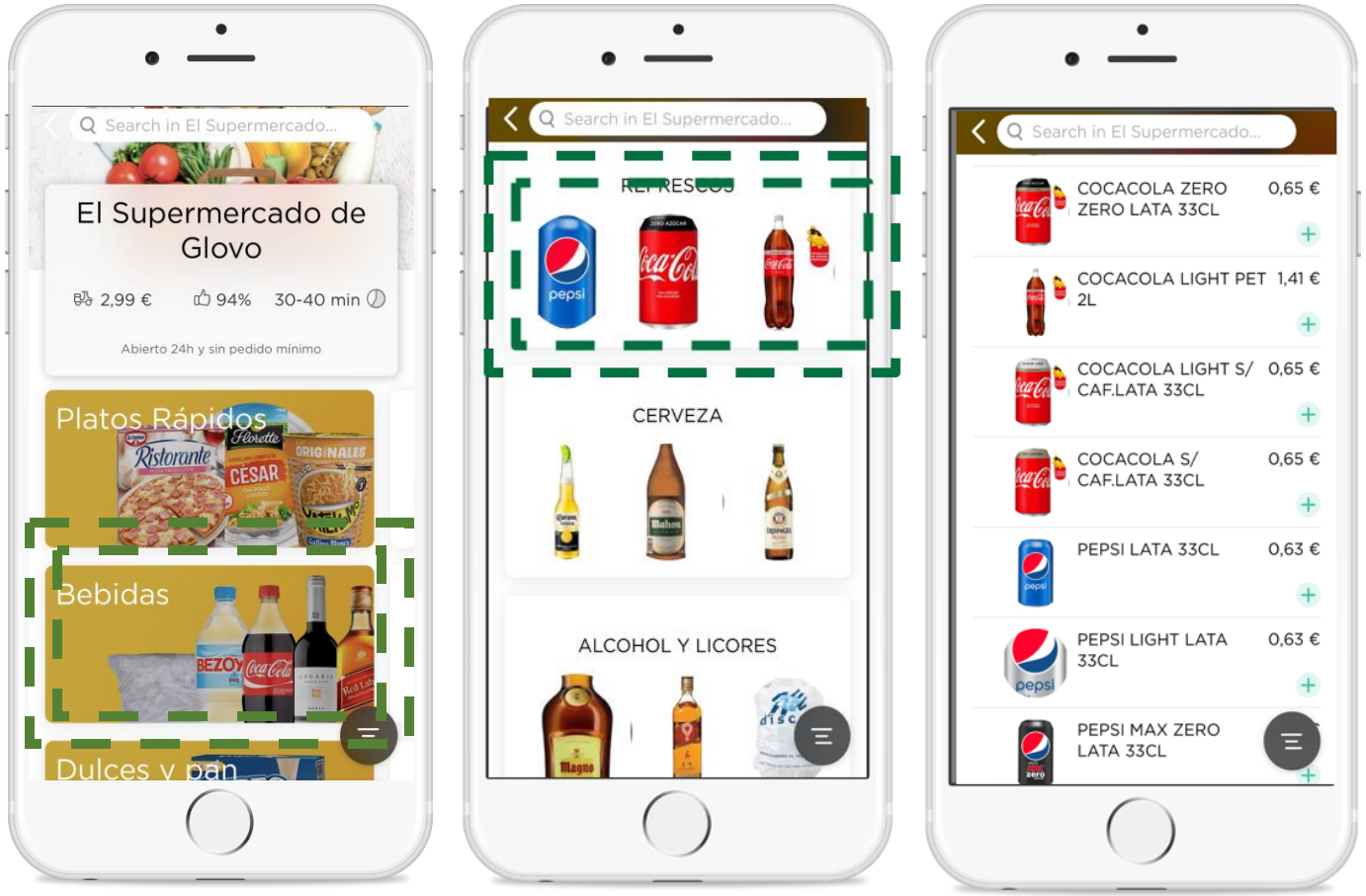
DEAL/ PROVIDERS

Total sales
Ratings
Coupon

USER ⟷ DEAL/ PROVIDERS

Delivery distance,
Browsing rate vs Order rate,
Order preference (brands, amount),
Ratings…

CURRENT SUPERGLOVO RANKING



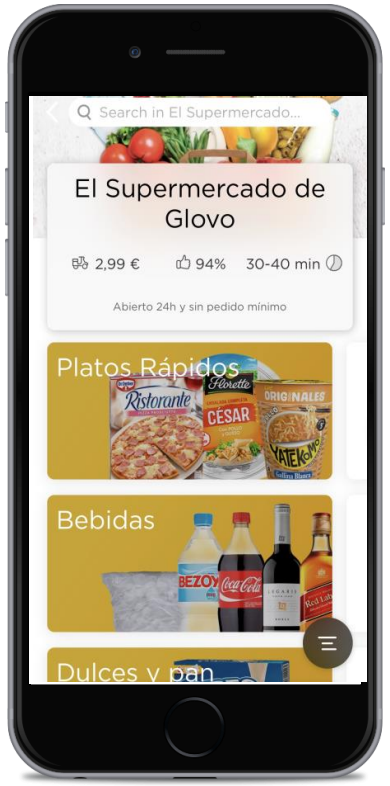
DIFFERS AMONG REGIONS

Lack of universal ranking criteria, which reduces management efficiency and makes impossible to reach algorithm scalability

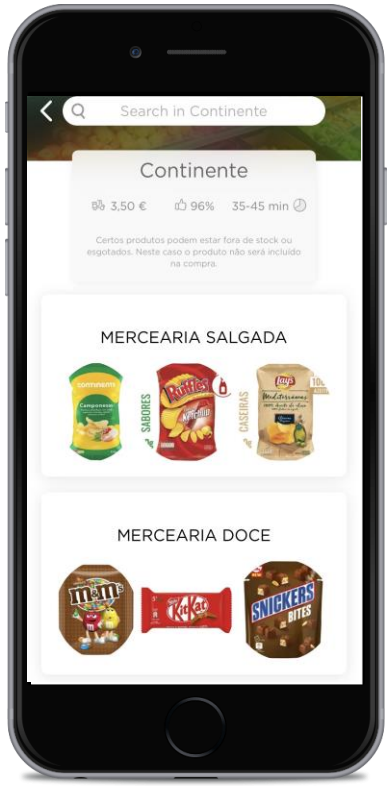
LACK OF CLEAR STRATEGY

There is no clear strategy even among the same region. It creates chaos, reduces the user experience and consequently the conversion rate.

DIFFERENT COUNTRIES, DIFFERENT RANKING



Barcelona (Spain)



Lisbon (Portugal)

Lima (Perù)

- Completely different ranking order between the different countries
- Is the first category shown the most popular? Or in alphabetical order?
- Is it helping the consumers to find easily products or only confusing them?

CONCLUSIONS

&

RECOMMENDATIONS

CATEGORIZATION – IT APPROACH – ML ALGORITHM

Glovo

1 SAMPLING

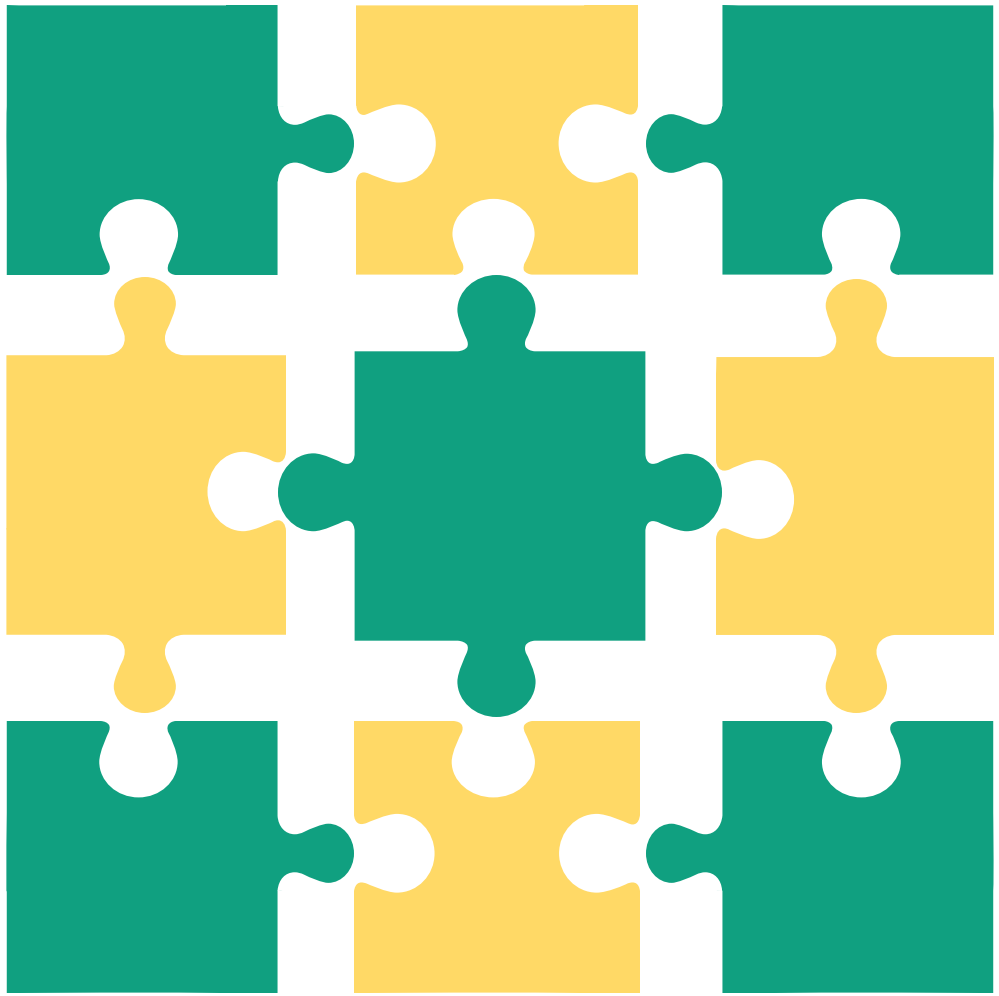
The sampling should be done for two main reasons; firstly to avoid biased and secondly to have a representation from each category.
→ Stratified Random Sampling

2 PRE-PROCESSING

In this step the text is transformed to numeric representation in order to understand the importance of each word.
→ TF-IDF, CountVectorized, stop-words, stemming, lemmetisations

3 TRAINING

Simply training the model, so fit the model to the training data.
→ Naïve Bayes, Support Vector Machines



4 PARAMETER OPTIMIZATION

Find the best parameter to the model. In this step it is important to select the best parameters in order to obtain a functioning model.
→ Grid Search, Cross Validation

5 MODEL NESTING

Train separate models for different levels of hierarchy. In this case it is necessary to create different model for every category level.
→ Groupby, sub-setting, For-loops

6 PRODUCTIONISING

Setting the model to be used in real life. Try before with the existing products and then with new added products that must be categorized for the first time.
→ Pickling, Django, Javascript

CATEGORIZATION – IT APPROACH – ML ALGORITHM

SCALABLE
Once the algorithm is developed, it is going to scale up with the amount of products.

EASY FOR NEW PRODUCTS
It is going to be easy to categorize new products, by just inserting them in the sample.

AUTOMATIC
It will require only someone able to review the performance, categorization will be done in a completely automatic way.

NEED OF ADDITIONAL INFORMATION
As now, with the available dataset would be impossible to create the ML algorithm. New data on products are needed, such as product description or key words

NEED OF CLEANING DATA
Creating a ML algorithm always requires to have cleaning data. This means that all the product data should be cleaned periodically.

POSSIBLE MISTAKES
ML learning algorithms learn with time, thus, especially at the beginning errors are possible and supervision is required.

CATEGORIZATION – BUSINESS APPROACH

DIFFERENTIATION

Given the different consumer behaviour among the selected locations, we would advice two separate categorization for Europe and Latin America.

SIMPLICITY

Bigger and less super-collections in order to identify better the differences between the super collections and the collections. Moreover it would be better to avoid sections containing only one product

TEMPORARY/SPECIAL CATEGORIES

As seen from the retail strategy, temporary categories are valuable. Creating more temporary and special categories can simplify the purchasing process. For example, could be useful to create a category for «Pre-party» or «Calcotada». This kind of special categories can be created with the use of associations rules.

CLUSTERING

As improvement plan, we also would like to recommend the use of cluster for the categories creation. Similar products, with similar kew words would go in the same category.

RANKING – BUSINESS APPROACH

MOST RELEVANT KEYWORDS	• Rank according to the most popular products/category • In Europe would be Bebidas while in Latin America would be Cigarillos • Helps consumers to find immediately what they are looking for (most of the time)
SALES	• Rank categories by the biggest by sales • This category should show the trends of the market • People tend to trust what others consumers buy
SPONSORED PRODUCTS	• Rank first the products or category containing the products the company is sponsoring • Could create extra revenue, by asking for a fee to be listed first
MOST RATED PRODUCT	• Rank product according to their ranking given by the users • Reviews are getting more important with the spread of internet, people trust them more day by day
OTHER FACTORS	• It is possible to use other factors such as prices, consumer satisfaction, margins. Parent-child, associations rules

RANKING - ALGORITHM APPROACH

ITEM-BASED

- Compute similarity among items
- Recommend similar items to those already liked/bought by the user

$$S(u,i) = \overline{r_i} + \frac{\sum_{j \in N} (r_{uj} - \overline{r_j}) \cdot w_{ij}}{\sum_{j \in N} w_{ij}}$$

USER-BASED

- Look for similar users
- Rank items according to what has been liked /bought by similar users

$$S(u,i) = \overline{r_u} + \frac{\sum_{v \in U} (r_{vi} - \overline{r_v}) \cdot w_{uv}}{\sum_{v \in U} w_{uv}}$$

ASSOCIATIONS' RULES

- Create associations rules based on previous orders
- Rank items that are usually bought together

OTHER FACTORS

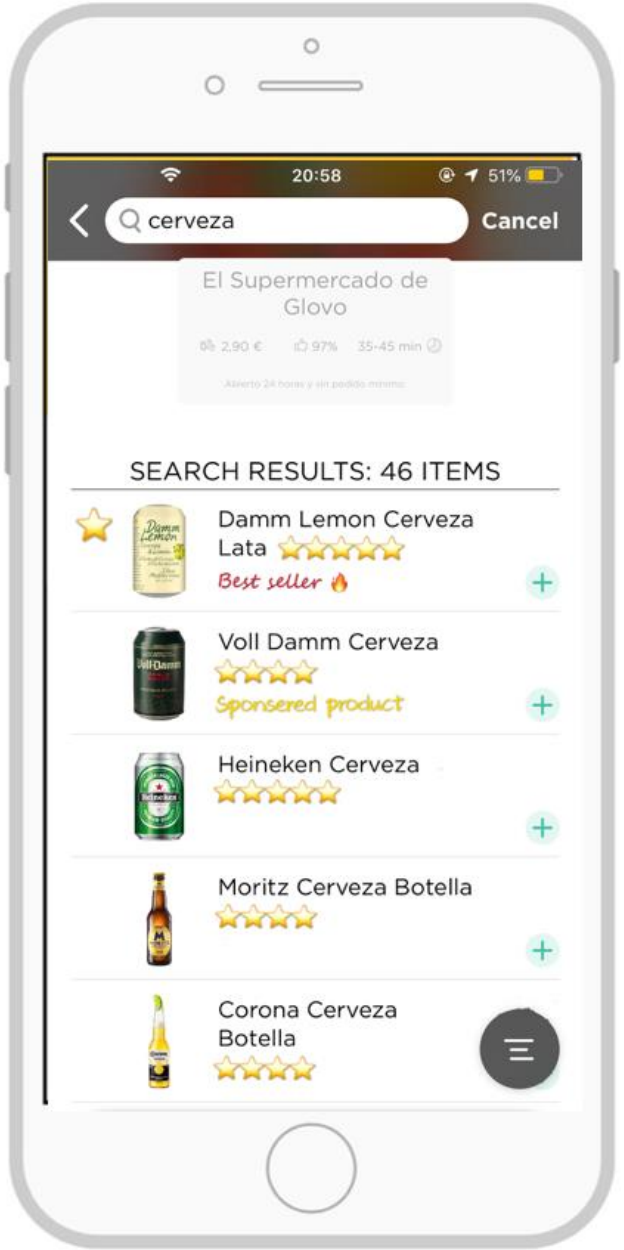
- Allowed for misspelling
- Include variations

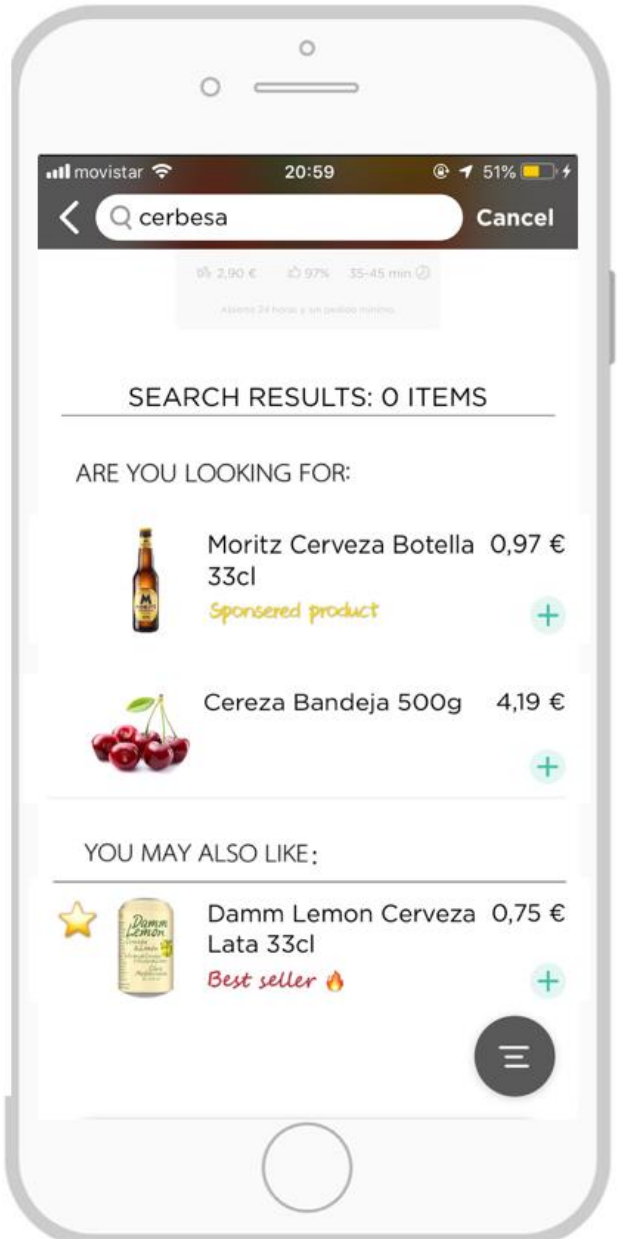
RANKING - NEW USER

Male user, downloaded Glovo 3 months ago but haven't made his first purchase yet. On June 20, he clicked into SuperGlovo for the first time in thea fternoon and searched for Cerveza.

Ranking Strategy:

- Best match + Best seller comes first;
- High rating is important for new users;
- Sponsored products have higher ranking even with lower ratings;
- Allowance for misspelling;
- Best sale product for cold start when lack of users' order history;
- Festival package: make Sant Joan package for uesers which including beers, bottles of water and a box of cigarettes.



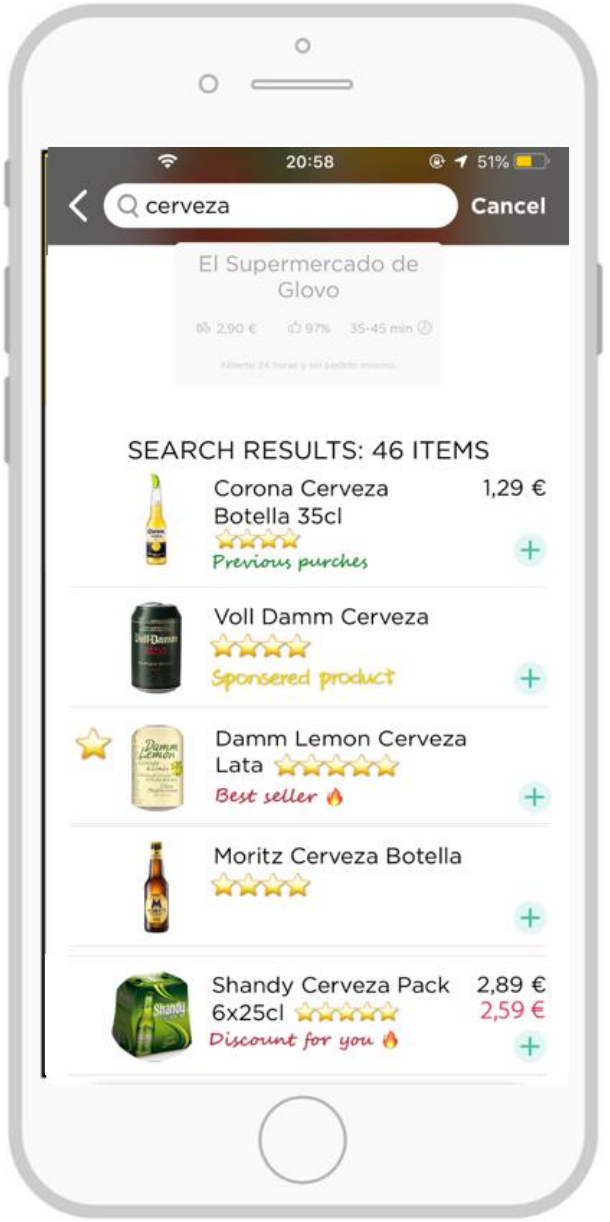


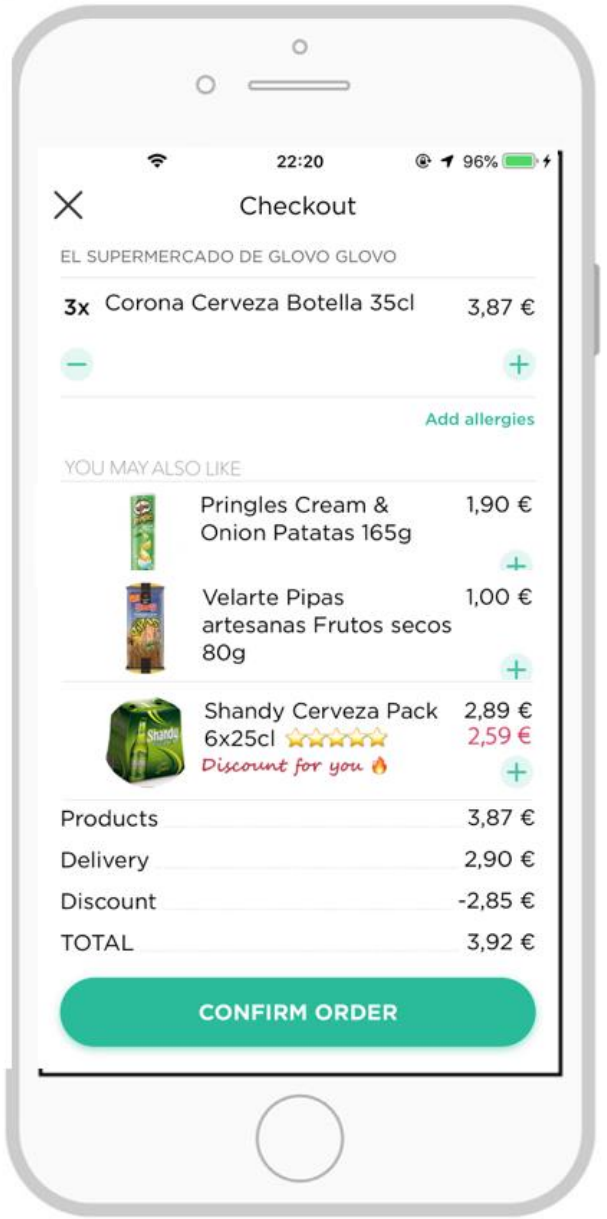
SIX MONTHS LATER

Made serveral purches during past 6 months. This afernoon he opened Glovo and searched for Cerveza again.
He is also a N26 user. Last time he made a 5,7€ order, which gave him 2,85€ discount this time.

Ranking Strategy:

- Best match + previous purchase comes first;
- Based on his order history applying User-based and Item-based recommendation algoritm to give out similier product recommendations and offer discounts as promotion;
- Related products and in Checkout session to increase bucket size.





METRICS

CHECKOUT TIME

One of the objective of the categorization is to reduce the checkout time.
It could be measured by the time spend by the user on the app. (missing data regarding the consumers' access from our dataset in order to calculate it)

BASKET SIZE

Calculate the size of the basket of each consumers.
- Products per order
- Total expenses (Price of the basket)


```
Order_sizeEU=EU.groupby(['order_id']).agg({'order_id':'count'}).sort_values(by='order_id',ascending=False)
```

RETENTION RATE

To check whether the consumers buy multiple times. Calculated by counting the orders from the same consumer ID.


```
consumer=EU[['customer_id','order_id']].groupby(['customer_id']).count()
```

CUSTOMER EXPERIENCE

The consumer experience can be assessed in different ways. The rating post the order is easy but most of the time is a missing information (because people do not like to complete surveys after the order)