Capstone Project

Connected Bar

Context

Connected Bar is the name of an ambitious project that will kick off in the Italian market and consists of having telemetry data of the taps of several bars of the country. An IoT device was installed in the bar taps and a team of data scientists, data engineers, designers and more, are working together to have this data available and displayed in the best way for their customers.

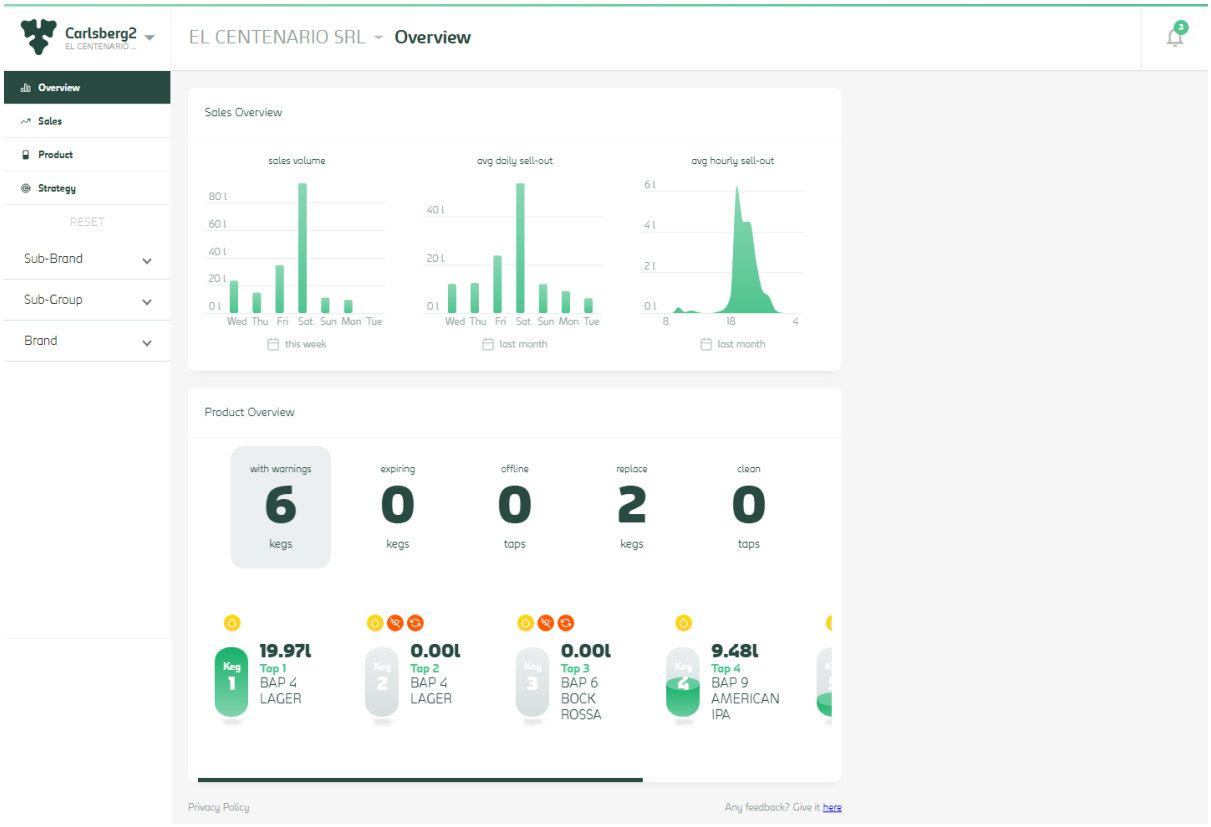
There are 2 key stakeholders in this project:

- The bar owners
- Carlsberg

Due to this project, the bar owners will have an unprecedent control over their sales and have a new tool to analyse their customers habits and eventually tailor their sales to match their customer's needs.

Carlsberg has an underlying goal with this project. On top of offering this tool to the bar owners and gain access to the telemetry data, Carlsberg would like to increase the sales of Craft and Specialities (C&S) type of beers. In order to achieve this goal, a team was put together.

An app will be developed to display the telemetry data. The developers are in charge of developing the app as it is expected, and they will work closely with designers. Your role in the team is to find correlations in the data, share insights with the team (both developers and design team) and create a recommendation system.

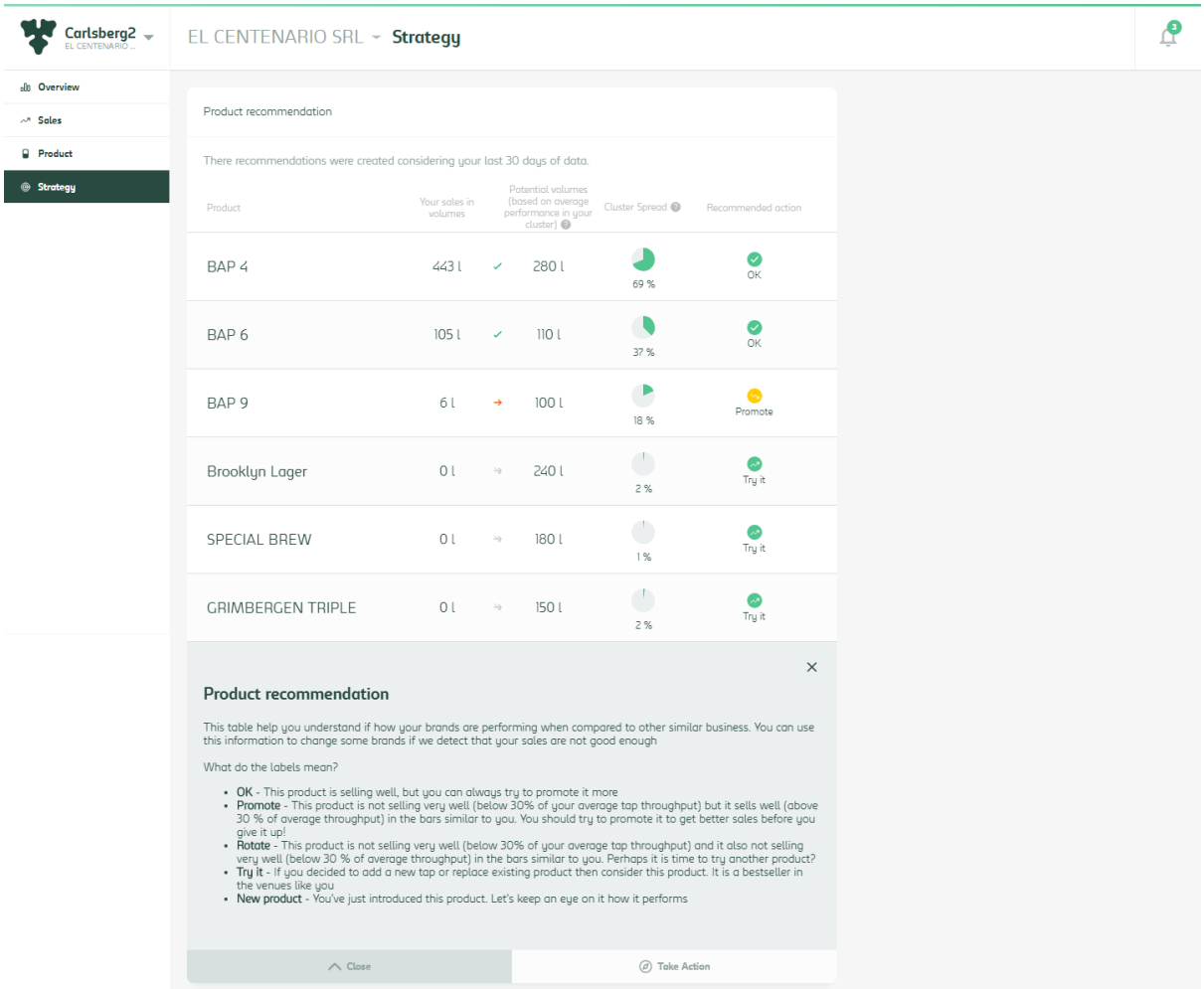


Goal

Your main goal is to design a recommendation system that, taking as input the data that was given to you by the market (described in “Data”), will generate something of use to the app developers: it can be a file or multiple files with your insights.

After having a talk with rest of your team, it is agreed that there will be a “Strategy” page on the app where the bar owners can see the recommendations generated by your model. You will be responsible to “feed” the Strategy page with the required data!

The Strategy page will look like something like this:



The cluster spread should reflect the percentage bars belonging to the cluster that sell that specific material. The potential volumes should give an idea to the user about the average performance of bars similar to theirs (in the same cluster).

Keep in mind:

1. Your goal is to increase C&S sales
2. The data will be refreshed once a week

Hints and tips:

1. Clustering is a good idea to explore in this scenario! Explore different clustering techniques and you may find some good information to help you with this in data itself (geographical data, bar categories, etc
2. Don't forget that you are part of a collaborative team: if you feel the Strategy page is not giving the best insights feel free to add new types of recommendations if they are well justified (layout, type of recommendation, etc)

Bonus:

A project of this magnitude must be well thought before being put into production! Have you taken a second to think of the structure that is required to comply with all the needs of this project? Give a detailed explanation of the different layers explaining the architecture of this project and suggest real tools that you would use for each layer (Data ingestion, Storage, Analytics, etc). There is no right architecture, don't forget that there are several ways to answer this question.

Data

1. Sell-in (sellin.csv) – data containing sales information: amount, quantity, …
2. Customer (customer.csv) – customer information (descriptive)
3. Material (material.csv) – material/product information (descriptive)