Data driven offer creation

BAKERY & CO.

Pilot

1.- Business problem and business goal

Bakery & Co. is a **national bakery company** that started as a family business twenty years ago in Madrid. In the last ten years the company has experienced a **large growth** with the opening of more than **50 new stores** at a national level. Each store operates individually from each other and reports to the central offices in Madrid.

Bakery & Co. wants to develop a **new data driven approach for managing the offers** that it applies in each store and had request a proposal for applying offers in each store based on each one sells. The main offices of Bakery & Co. are located in Madrid. From there they receive:

- Monthly incomes from each of the different stores
- Materials requests
- Incidents

At this moment, each responsible store decides what offers to apply to each product/group of products and this information is difficult to be tracked from the central office of the company. Bakery & Co. wants to establish a centralized revenue model and a centralized-managed offers model.

In order to achieve this, the company has started to install a centralize ERP system and wants to establish a centralize model based on sells information for the definition of different offers.

The stores are managed by managers depending on its location and are separated in North (12) – Centre (24) – South (14) and Mediterranean divisions (21). Each store has a responsible store. Bakery & Co. needs to centralize the management process of offers, in order to increase the effect of these offers, the number of sells and to facilitate the logistics between its shops. Each shop operates individually, but there is a project aiming to centralize all stores in a single ERP. Based on our previous experience, usual effects of this approach offer the following ROI:

- Increase in sales around 10-30%
- Reduce in the logistics cost by 15%
- Increase the information available to manage the stores.

We propose a three-stage project in order to achieve the goal of applying a model to each store and a global model for the company.

Stage 1 – Individual Bakery shop model (Pilot):

- Establish a pilot model in one bakery in Madrid.
- Obtain records of transactions in the new ERP model format.
- Create a model and monitor it in order to apply it.
- The data will be processed offline, one first model will be applied to historical data

Stage 2 – Implementation of the model:

- Implementation of the model by geography.
- Analysis of the results.

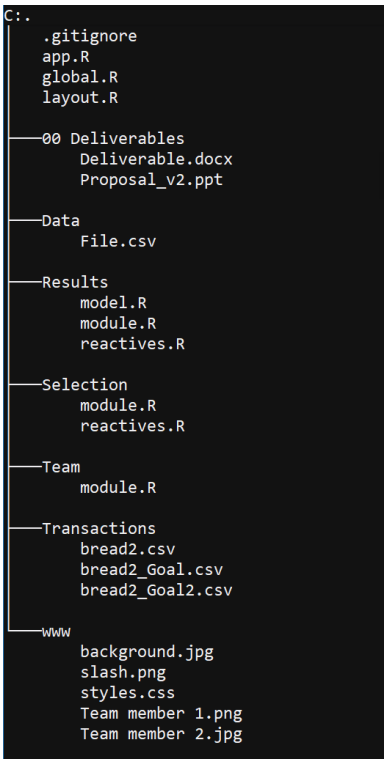
Stage 3 – Centralized model:

- Establish an aggregate company model with each individual model.
- Establish a second model, which will include new data generated and which offers are already established.
- Manage logistics and offers.

1.1 Document goal

The goal of this document is to explain the results of the pilot, detail the work performed and set the basis for the following stages of the project in case Bakery & Co. is interested.

The structure of the project developed is the following:



The more relevant files are:

- **App.R**, file needed to execute the application.
- **Data\File.csv**, file that contains the transaction extracted from the system.
- In the transactions folder there are some alternative data sets to use.

2.- Business understanding and input data

We have performed four meetings with company personnel previously to gather the data necessary for the project:

- **Central offices sales responsible,** in order to acquire the knowledge about the actual revenue model and the one the company wants to achieve.
- **Company CIO,** where we saw the ERP the was going to be implemented and defined the format that the data should have and the query to extract the data.
- **Centre geography manager,** we reviewed the actual problems (logistics, sales, …) that he/she used to face in order to take them into account.
- **Shop responsible,** we had an interview with the shop responsible where we reviewed the offers that are currently made and some other facts about the shop.

After performing the interviews, we have decided to establish a generic query based on the new ERP of the company for extracting the data that we need with the following format:

Field	Format
Date	YYYY-MM-DD
Time	HH:mm:ss
Transaction	Numeric
Item	String

3.- Data pre-processing

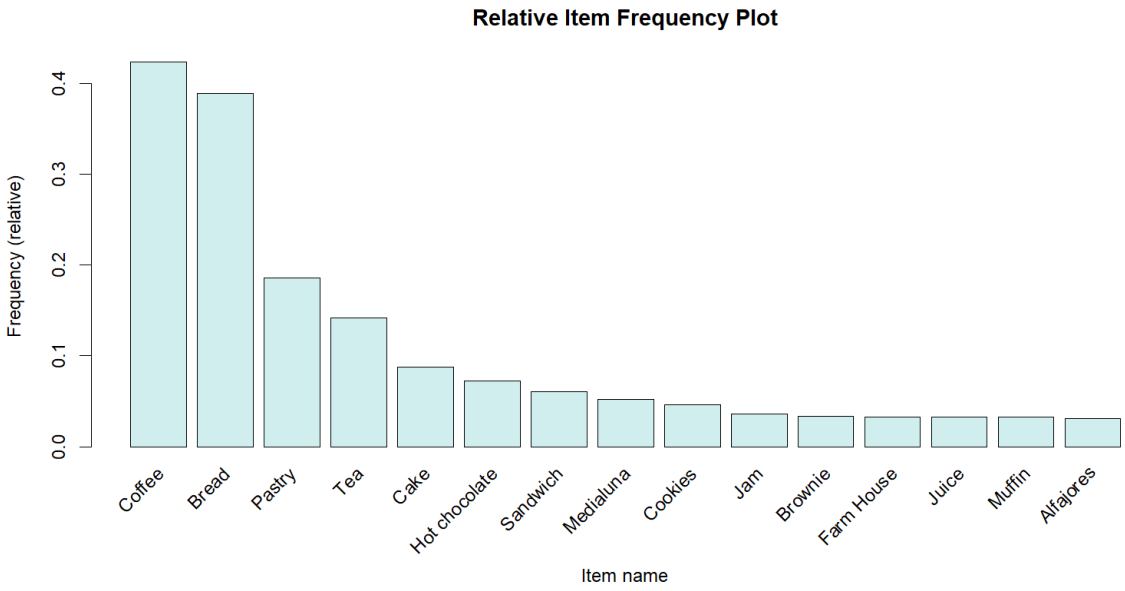
Once the query was calculated, we have received the following dataset:

File.csv

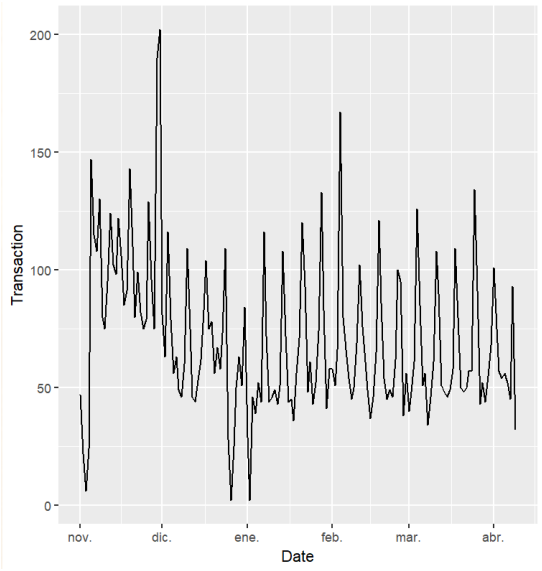
We have applied the following analysis:

- **Summary**,
  - The dataset has 24.547 rows, with 11.245 unique transactions.
  - There is an average of 2,039 items per transaction.

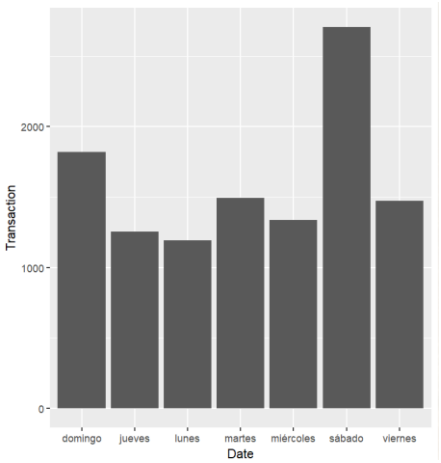
- The following are the **products** with higher sells. We can see that coffee and bread are the most relevant ones:



- **Sales**, we can appreciate that they decreased until January, where they were stabilized.



- **Sales by weekday**, there is a peak on Saturday



Based on this analysis and the interviews with the company personnel, we have decided to apply the algorithm to the following subsets:

- **Weekends**
- **Weekdays at lunch time**
- **All data**

4.- Algorithm application

After the first analysis, we have decided to apply an *apriori* algorithm. This algorithm proceeds by identifying the frequent individual items and extending them to larger and larger item sets, which will allow us to see which items are sold together and target those items with appropriate offers.

This algorithm needs two parameters to be set:

- Support of an item, is the proportion of transactions in which an item appears:

$$suport(X) = \frac{\#\ of\ tansactions\ in\ which\ an\ item\ appears}{\#\ of\ total\ transactions}$$

- Confidence of a rule, this says how likely is that an item Y is purchased when item X is also purchased, expressed as $\{X -> Y\}$

$$Confidence(\{X -> Y\}) = \frac{support(\{X, Y\})}{support(X)}$$

Afterwards, the model will be evaluated using the following measure:

- Lift, this says how likely is that an item Y is purchased when item X is also purchased, while controlling for how popular item Y is.

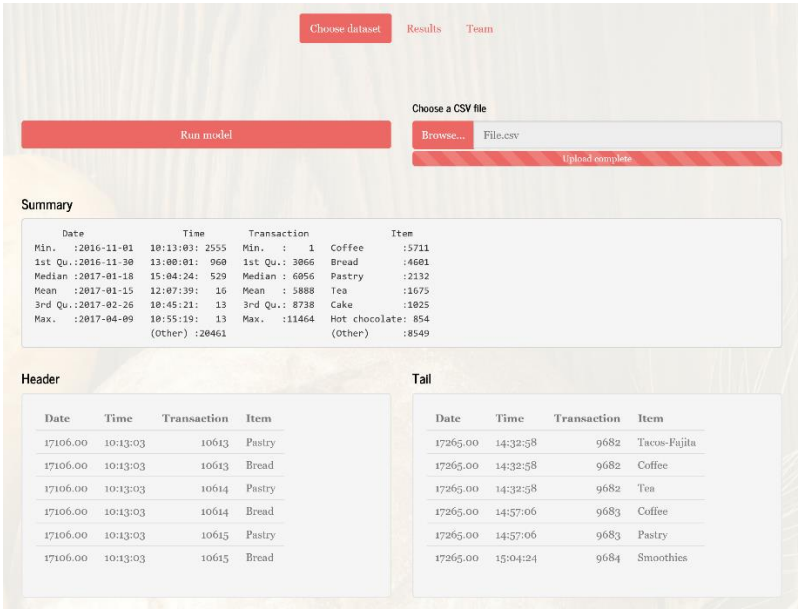
$$Lift(\{X -> Y\}) = \frac{support(\{X, Y\})}{support(X) \times support(Y)}$$

For a detailed explanation of the algorithm, please follow the link below:

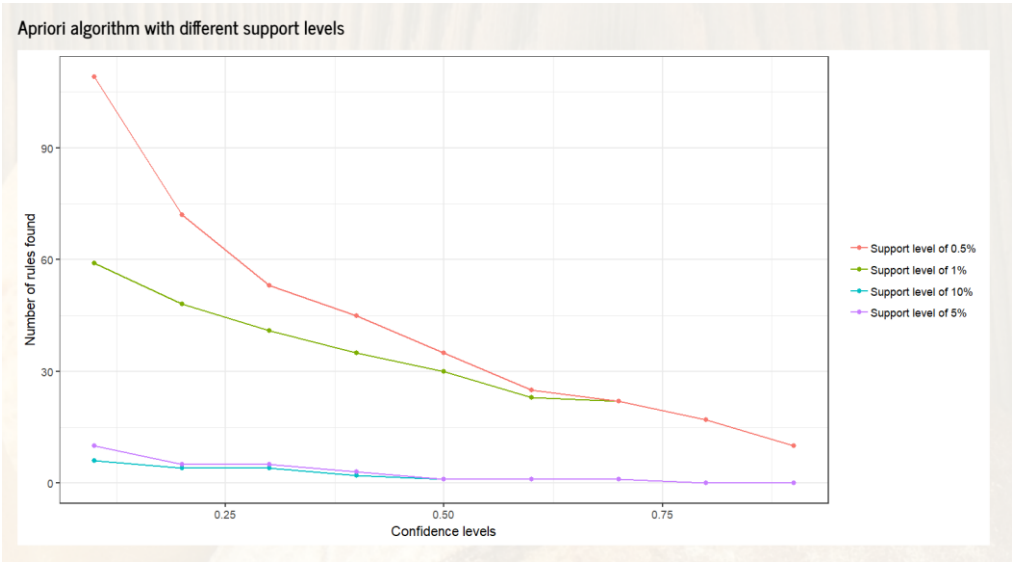
- https://en.wikipedia.org/wiki/Apriori_algorithm

In order to make the process replicable, we have created an app for performing the analysis. This app is separated in three tabs:

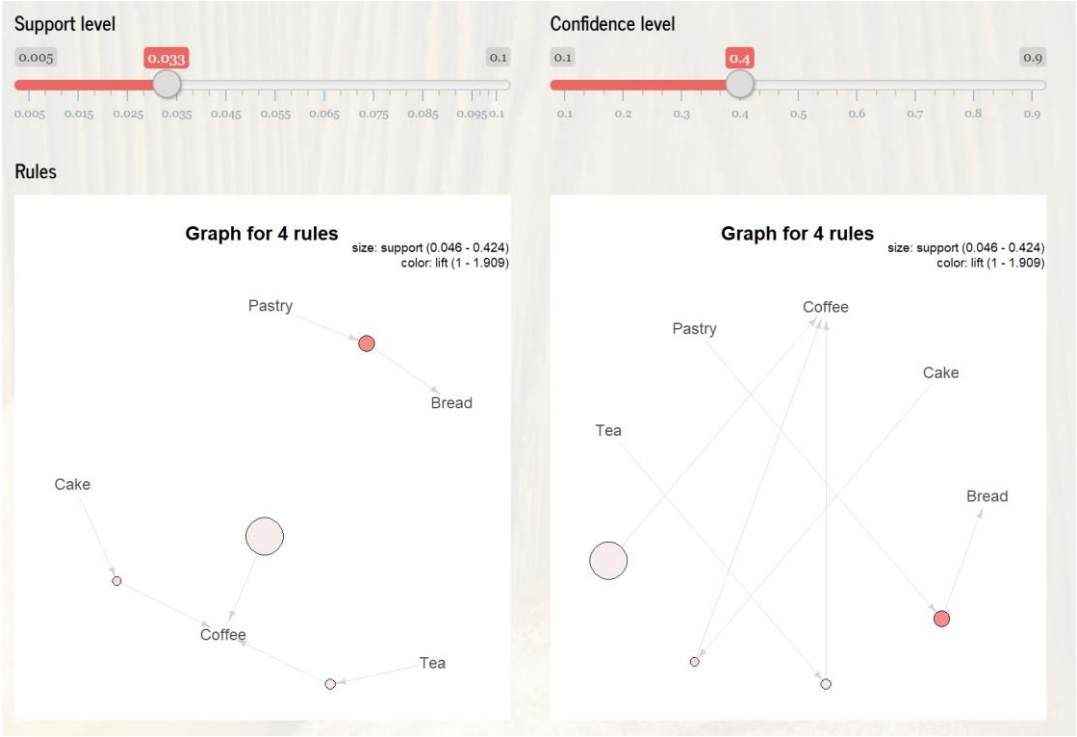
- **Dataset load**, in the first tab it is possible to select a file with the appropriate format an upload it. The app will show a summary of the dataset:



- **Results**, once the user has clicked in “Run model”, the results will appear in the results tab, there are three subtabs (Every day, weekdays and weekends).
  First of all, a summary of rules based on confidence and support levels will be plotted:



Then, it is possible to set up confidence and support in order to show the different rules calculated:



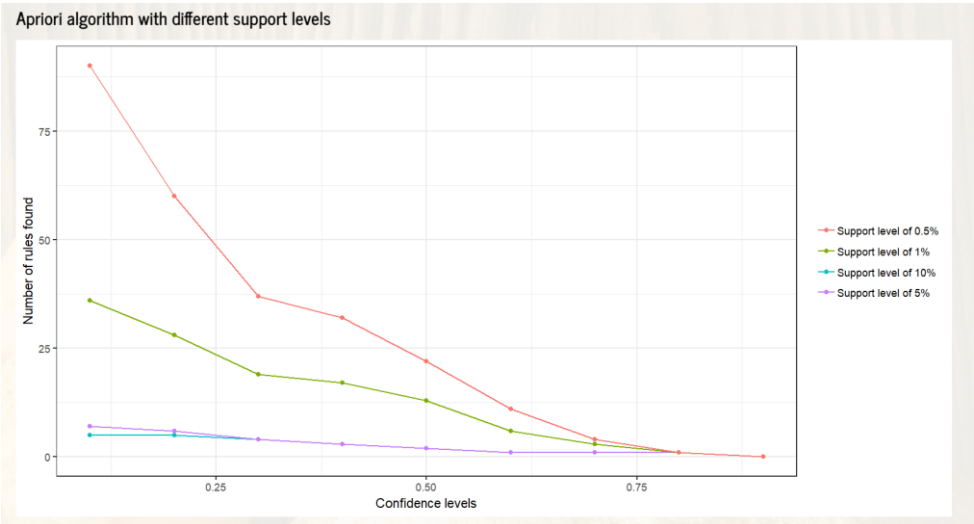
Afterwards some summary graphs and information are shown.

- **Team**, this tab presents the team in charge on maintenance.

5.- Algorithm results

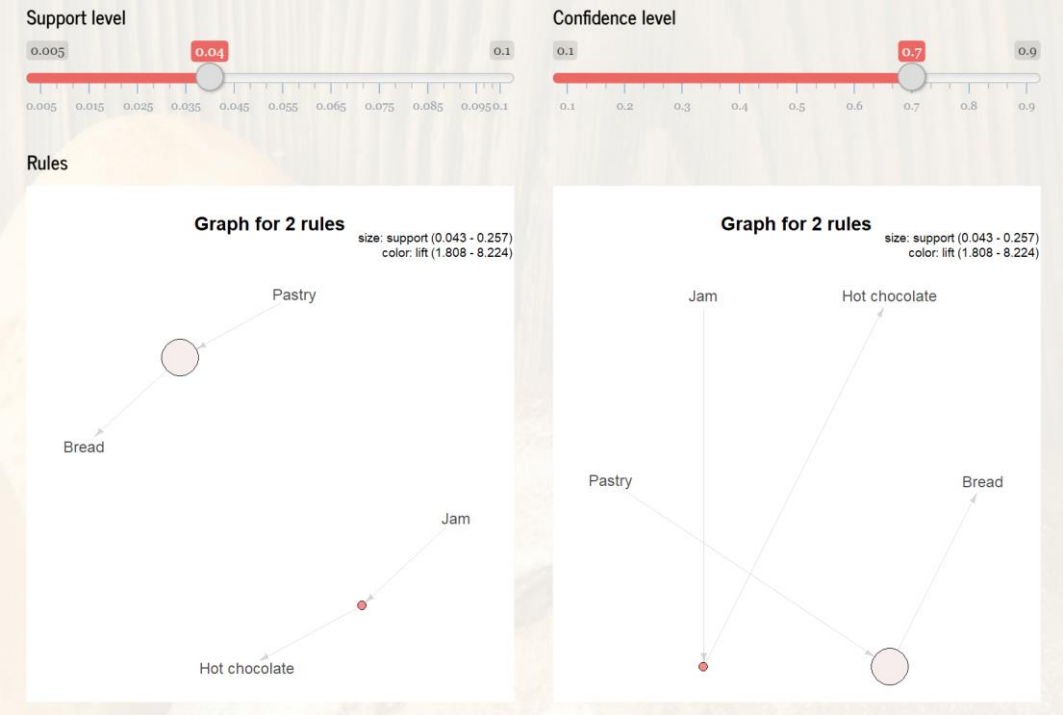
Based on the results of the algorithm, we can conclude the following:

- **Weekends,** in order to analyse the weekends, we have extracted all the transactions that have occurred during Saturday and Sunday. After applying an apriori algorithm with different confidences and support levels, we have obtained the following results:

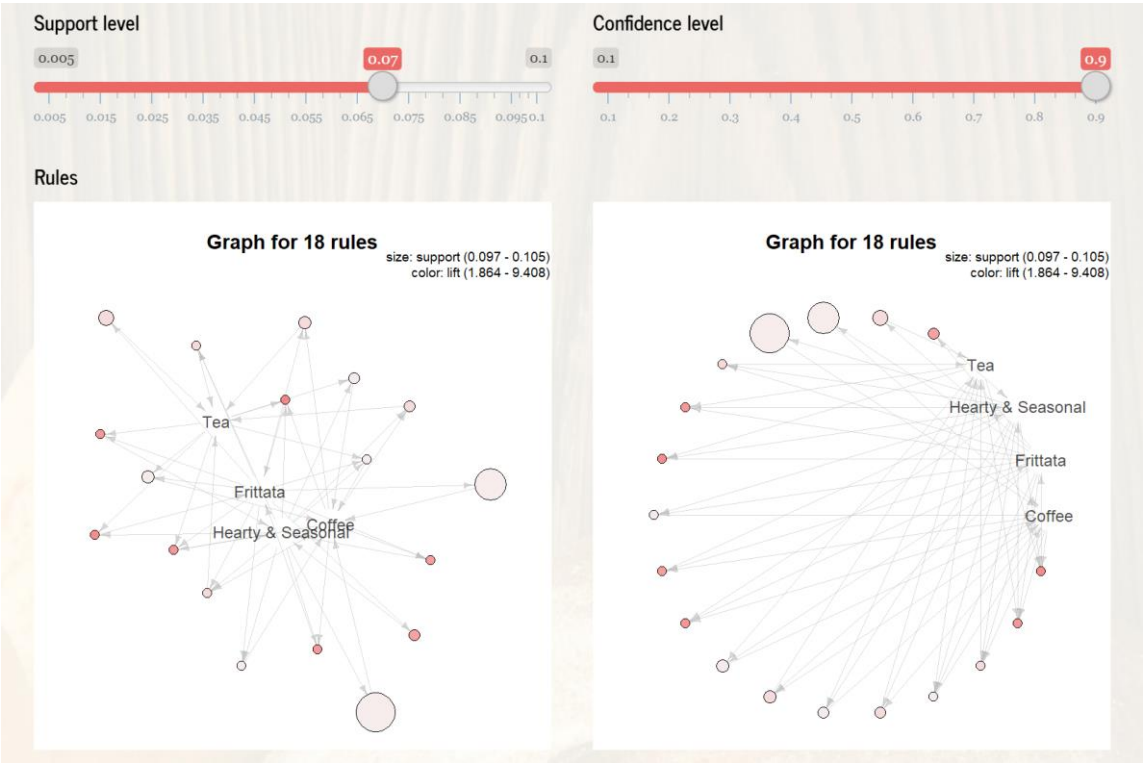


We obtained two rules with confidence level 0.7 and support level 0.04:

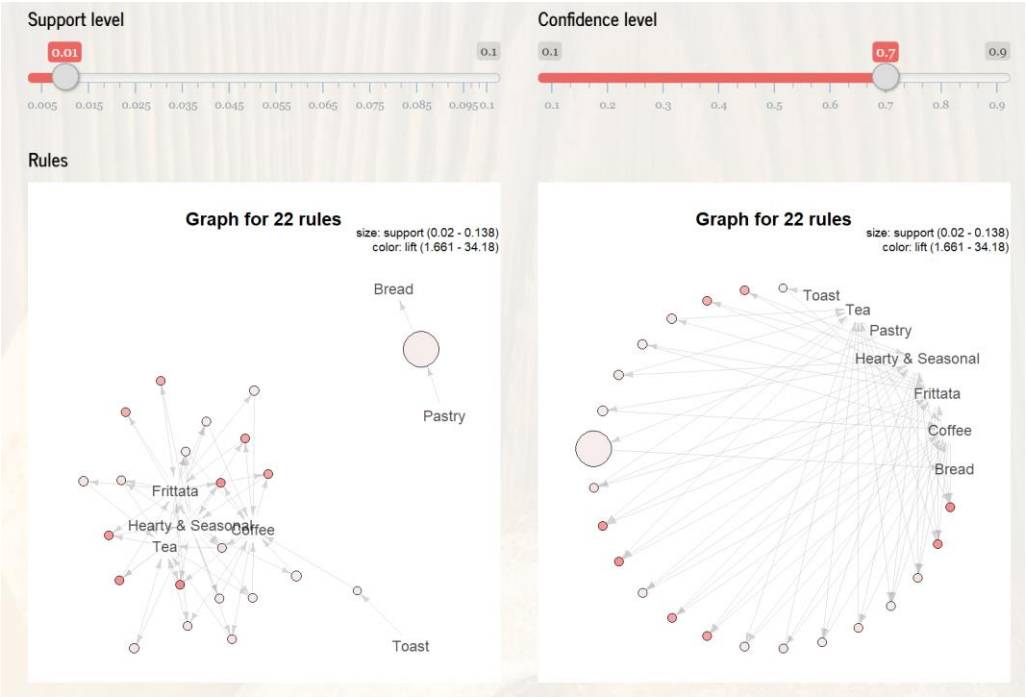
- Pastry → Bread
- Jam → Hot chocolate



- **Weekdays at lunch time,** repeating the process for weekdays and lunch time we see different rules regarding tea, coffee, frittata and "hearty & seasonal":



- **All data**, if we apply the algorithm to all data, we do not obtain very good results. If we decrease the support level to 1%, we start to see the rules that we already have seen in the previous analysis, so there is nothing new to add regarding the dataset as a whole. Nevertheless, we included this option for future applications and in case the company wants to explore further.



All the results have been validated with the company personnel and we have decided together the parameters to use.

In summary, we recommend that the company **targets the following products** during the next three months of monitoring:

- **Tea – coffee – frittata – "hearty & Seasonal"** during weekdays, based on the cost/benefit of these products.
- **Pastry and Bread, Jam and Hot chocolate** during weekends, again based on the cost/benefit of these products

For future analysis, we are considering including the costs of the products in order to be able to calculate an approximation of the profits/losses of each offer.

6.- KPIs analysis & next steps

In order to perform a follow – up during the next months, we propose some KPIs that can be completed with the ones that the company could consider:

- Most sold products by day of the week.
- $\frac{\#\ products\ (or\ groups\ of\ products)\ sold\ before\ offer}{\#\ products\ sold\ after\ offer}$ (in stablished periods, 1 week, 1 month,…), this number should be less than 1.
- $\frac{\#\ transactions\ containing\ an\ offer}{\#\ of\ total\ transactions}$, in order to know if the offers are being selected.

The follow up will be done by scheduling a monthly meeting during the next three months in order to check KPIs and propose action plans according to its evolution.

If the application is succeeded, we will start with its implementation in the different geographies, and with the development of the global model that will include more variables to the analysis.