

Measuring Wellbeing by extracting Social Indicators from Big Data

R.Campagni¹, L. Gabrielli¹, L. Gabrielli³, F. Giannotti², F. Maggino¹

¹University of Firenze, ²ISTI-CNR Pisa, ³University of Pisa

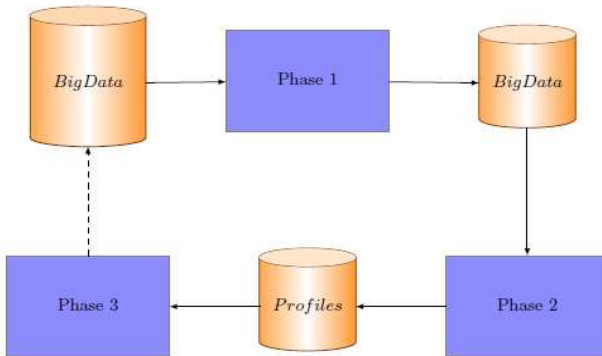
October, 20-21 2016 Budapest



Aims and phases of project

- *Extracting* from big data concerning the purchases useful information to construct indicators describing social phenomena.
- *Analyzing* the behavior of different families in a crucial period, by paying attention to possible changes in the lifestyle of the people also referring to the period of crisis.
- *Defining* new social indicators to describe customer purchase behaviors, by changing the classical methodological approach by considering data collected for other purposes.

The analytical process



- Phase 1: extract from data useful information for analysis
- Phase 2: perform analysis
- **Phase 3: discover some sentry products**

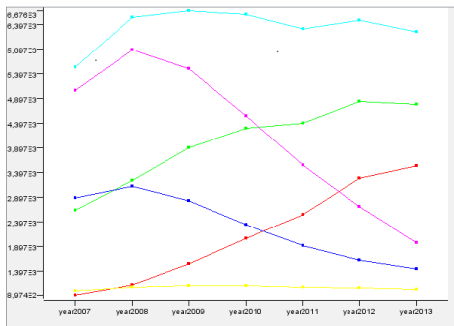
Data for analysis

- We observed purchases of about 13000 customers during 2007-2013 by analyzing several attributes describing the way in which they have been shopping in a store of a big supermarket.
- We classified customers in groups depending on how much they spent in the years under analysis.
- We looked for important factors helping us defining new social indicators related to wellbeing.
- We refer to our previous study where amounts, quantities and times were analyzed.

Previous analysis results

We found some interesting behaviours-groups respect to the annual amounts:

- LC Low Constant (yellow line, 2485 customers)
- LG Low Growing (red line, 1580 customers)
- MG Medium Growing (green line, 1527 customers)



New indicators - the aim

- Obtain timely information to discover important signals related to particular behaviors.
- Predict changes in the macroeconomic context.

By using clustering techniques we grouped customers and observed changes in shopping cart to understand if when *amounts, quantities and number of expenses significantly change* ⇒ *typologies of products purchased also change*.

During crisis, a group of customers has reduced purchase of niche products, to the benefit of lower-end products.

New indicators

- We deepen the analysis concerning the products categories purchased by customers of the three segments LC, LG and MG.
- We analyze how customer behaviours change for what concern the type of purchased products.
- The goal is to find which are the products that can be considered *sentry products*.
- Keep under control these products can help us to identify important signs of change in people's lifestyle.

Products classification

The commodity classification available for our data is up to the category level:

Area	MacroSector	Sector	Department	Category
food	various kinds	grocery	liquid	water
food	fresch	more fresch	bread	internal prod. bread
⋮	⋮	⋮	⋮	⋮
no food	various kinds	chimical	environ. hygiene	environ. deodorant
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

We analyze data from the perspective of this products categories by considering only products belonging to **food area**.

Analyzing products categories

We start from aggregated data, by selecting products being to 75^o percentile; for each year and category we have the purchased quantity.

category	2007	2008	...	2013
bread	5460	6745	...	18271
dried fruit	2900	3036	...	4194
⋮	⋮	⋮	⋮	⋮
potatoes	5971	5910	...	5553
⋮	⋮	⋮	⋮	⋮

- We perform a clustering step of products data of the first year,
- then use the model to group data of the others years.

Clustering analysis

We obtain for each product the list of clusters that they have passed through over the years.

category	2007	2008	...	2013
bread	cluster3	cluster3	...	cluster5
dried fruit	cluster1	cluster1	...	cluster2
⋮	⋮	⋮	⋮	⋮
potatoes	cluster2	cluster2	...	cluster2
⋮	⋮	⋮	⋮	⋮

LC customer categories

We analyzed the trend of products categories purchased from LC customer.

Many products quantities remain constant in the period, but we observe a particular behaviour of some **sentry** products:

- *elaborate red meat* and *slice salumi takeaway* decrease
- *internal production bread* increases

LG customer categories

The trend of products categories purchased from LG customer put in evidence the same **sentry** products, but with some differences:

- *slice salumi takeaway* decrease
- *elaborate red meat* decrease *lightly*
- *internal production bread* increases

MG customer categories

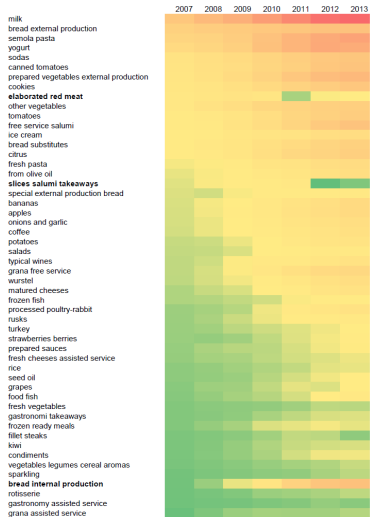
The trend of products categories purchased from MG customer shows a different behaviour:

- *elaborate red meat* decrease
- *internal production bread* increases
- *slice salumi takeaway* stay constant
- *savory snacks* decrease

LC colormap



LG colormap



MG colormap



Conclusions and future works

- We explored how customers change their buying patterns and we found out important signals putting in evidence a crisis that is also reflected in purchasing of essential goods.
- Customers sometimes opt to buy cheaper products, in other cases someone decides to reduce the purchase of certain products for the benefit of others.
- Understanding the reason for which customers behaviours change: it can be because the shops network change or because people generally start to eat less a food, for example the meat.

This study has to be seen as a phase in the definition of indicators that can measure the wellbeing.

This work is supported by the European Community's H2020 Program under the scheme '*INFRAIA – 1 – 2014 – 2015: Research Infrastructures*', grant agreement #654024 '*SoBigData: Social Mining & Big Data Ecosystem*'. (<http://www.sobigdata.eu>).

Thanks for attention!