

# DIGITALES ARCHIV

ZBW – Leibniz-Informationszentrum Wirtschaft  
ZBW – Leibniz Information Centre for Economics

Leogrande, Angelo; Massaro, Alessandro; Galiano, Maurizio

## Article

### The location of superstores in Italy : a metric approach

#### Provided in Cooperation with:

International Institute of Social and Economic Sciences, Prague

**Reference:** Leogrande, Angelo/Massaro, Alessandro et. al. (2020). The location of superstores in Italy : a metric approach. In: International journal of business & management 8 (1), S. 35 - 64. <https://www.iises.net/international-journal-of-business-management/publication-detail-25583?download=3>.  
doi:10.20472/BM.2020.8.1.003.

This Version is available at:  
<http://hdl.handle.net/11159/4624>

#### Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [rights\[at\]zbw.eu](mailto:rights[at]zbw.eu)  
<https://www.zbw.eu/econis-archiv/>

#### Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.  
<https://zbw.eu/econis-archiv/terms-of-use>

#### Terms of use:

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.*

DOI: [10.20472/BM.2020.8.1.003](https://doi.org/10.20472/BM.2020.8.1.003)

# THE LOCATION OF SUPERSTORES IN ITALY: A METRIC APPROACH

*ANGELO LEOGRANDE, ALESSANDRO MASSARO, MAURIZIO GALIANO*

## Abstract:

The role of socio-economic determinants for superstore location is analyzed. Data are collected by Italian Minister of Economic Development and ISTAT. Results show a positive effect of superstore location in respect to GDP, Employment and population. Panel data analysis shows positive relations among superstore location and Instruction and Formation, Economic Wellness, Social Relationships.

## Keywords:

Cluster Analysis, Panel Data, Branding, Location

**JEL Classification:** C38, C33, R30

## Authors:

ANGELO LEOGRANDE, Dyrecta Lab s.r.l. , Italy, Email: [angelo.leogrande@dyrecta.com](mailto:angelo.leogrande@dyrecta.com)

ALESSANDRO MASSARO, Dyrecta Lab s.r.l. , Italy, Email: [alessandro.massaro@dyrecta.com](mailto:alessandro.massaro@dyrecta.com)

MAURIZIO GALIANO , Dyrecta Lab s.r.l. , Italy, Email: [maurizio.galiano@dyrecta.com](mailto:maurizio.galiano@dyrecta.com)

## Citation:

ANGELO LEOGRANDE, ALESSANDRO MASSARO, MAURIZIO GALIANO (2020). The Location of Superstores in Italy: A Metric Approach . International Journal of Business and Management, Vol. VIII(1), pp. 35-64., [10.20472/BM.2020.8.1.003](https://doi.org/10.20472/BM.2020.8.1.003)

---

**Design/methodology/approach:** *Cluster Analysis and Panel Data Analysis are used to verify the presence of relations among variables able to explain the superstore location.*

**Research Findings:** *Panel data analysis shows positive relations among superstore location and Instruction and Formation, Economic Wellness, Social Relationships. Data shows the absence of a positive relation among the number of superstore and Work and Life Conditions.*

**Theoretical Contribution/Originality:** *the originality of the research is either in the database either in the matching between two different methodology of analysis.*

**Practitioner/Policy Implication:** *superstore manager can be informed about alternatives in location able to increase the level of profits.*

**Research limitation/Implications:** *the research is geographically limited and needs a deeper analysis able to illuminate meaningful determinants of superstore location.*

---

## 1. INTRODUCTION

The idea of the paper is to find some socio-economic determinants able to solve the question of business location in the sector of superstore. The choice of the location of a superstore may depend on a number of factors. In this article, a set of determinants are investigated to shed light on the variables that can solve the problem of the efficient location of a superstore. The paper presents two different techniques: clusterization and panel data metrics. Superstore are defined as a shop with a surface with more than 2.500 square meters (Development-MISE, s.d.). Data are collected with the objective to analyse the role of some of the main macro-groups in ISTAT-BES (Istat, s.d.) reports. Reported literature review analyze either the question of business location in general, either specifically the question of superstore and shopping centre location. Business location is generally based on a series of elements based on political economy. The level of taxes, the presence incentives, and the regulation of labour, and either the presence of infrastructure as highway can have an effect on the choice of business location. But there are also positive externalities able to increase the probability of entrepreneurs to choose a certain location. In the literature review is analyzed also the question of the distance by the city centre as a determinant for business location. But our approach tries to innovate the literature considering not only classical variables able to solve the question of optimal business allocation, but also taking care of new aspects regarding the socio-economic conditions of the population living in the Italian regions and macro-regions during the period 2004- 2016. We consider some determinants as Health, Instruction and Formation, Work and Life Conditions, Economic Wellness, Social Relations, Quality of Services.

The paper is organized as follows. The first paragraph is devoted to analyze the economic literature based on the economics of location with attention to the location of business. The second paragraph presents the results of clusterized matrix able to describe the location of superstore based on GDP, Employment and Population. The third paragraph presents the econometric model to estimate the number of Italian superstore based on socio-economic determinants. The fourth paragraph discusses the results of the econometric models. The fifth paragraph presents conclusion. Results suggests that the number of superstores in Italian regions and macro-regions positively associated with Health, Instruction and Formation, Economic Wellness, Social Relationships. Data shows that the number of superstores per regions is associated negatively with the level of Work and Life Conditions.

## 2. SUPERSTORE LOCATION: A LITERATURE REVIEW

In this part of the paper we analyze economic and geographic literature able to explain the determinants of business localization to introduce the cultural and the scientific environment of our analysis. The topic of business location is multi-disciplinary and it can take in consideration either urban and cultural studies, network analysis and the economics of incentives. Political economics have an impact in the field, too. But in our attempt to present a little anthology of the topic we concentrate our analysis of geographical economics and business economics, trying to settle the right problem of the location of business.

Superstore location is a part of the economics of location. The question of the optimal location of an enterprise or a firm of a productive organization can shape the entire panorama of cities and urban areas. The subsequent analysis of the literature synthesizes two elements: the business location and the superstore location. In the building process of our literature review we have focused the attention on the relation between the question of business location and the question of superstore location, considering that superstore location is just a specification of the more general problem of business location. But using an inductive method we can reconnect the question of business location to the general question of organizational location that is one of the more common questions in economic geography. It is more important to consider that the same idea of economic geography has been developed to solve the question of location of economic activity in the context of urban city development. In fact, at the origin economic geographies faced the question of the location of the cities and the development of a series of economic conditions in the context of economic geography. In the sequent analysis we consider a series of economist and geographers that have studied the question of location either referred to the city as a whole either referred to categories of economic organization such as industry. But before to analyze their contributions to the development of economic geography is necessary to consider the fact that economic geography is a complex field that puts together a series of disciplines and sub-fields such as for example anthropology, urban studies, and cultural studies, even if the field of economic geography is also scientific in the sense of usage of mathematical and statistical methods, there are always features that requires non-conventional methods, and the application of some kind of heterodox procedures

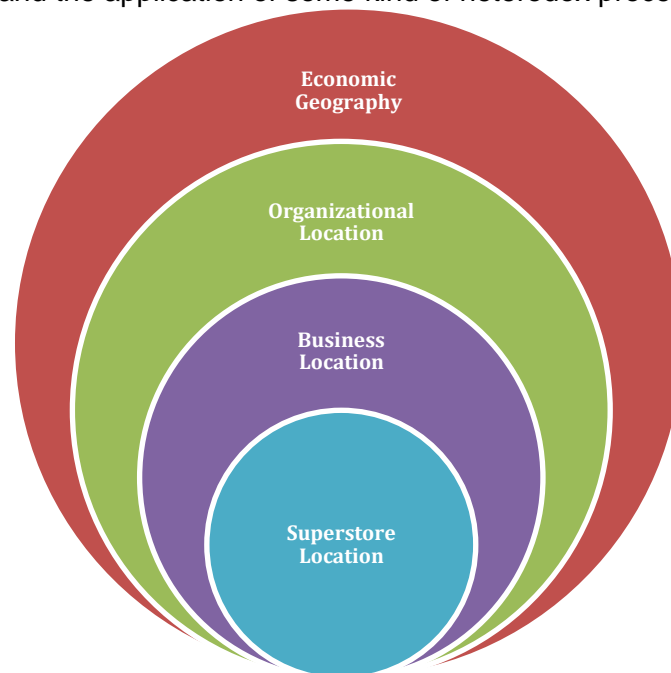


Figure 1. The relations among superstore location, business location, organizational location and economic geography.

Economic geography is a complex field, and the economic geographer need to use various tools and KPIs able to represent a heterogeneous knowledge and diversified scientific approaches. The mix of approaches, the alternance between quantitative and qualitative research, and the contemporaneous usage of analytical and narrative tools is able to shed lights, not only on the research questions, but also on the connections, and linkages among different aspects that are necessary to investigate the specific epistemic characteristics of economic geography.

- **Von Thunen:** the economist Von Thunen has developed a model to solve the question of the value of land. Von Thunen is recognized as one of the founding economists of the field of economic geography. He wrote a book in 1826 entitled "*The Isolated State*" in which he proposed a formula to determine the level of land rent. The formula is as follows:  $R = Y(p - c) - YFm$  in which  $R$  is equal to land rent,  $Y$  is equal to yield per unit of land,  $c$ = production expenses per unit of commodity,  $p$ = market price per unit of commodity,  $F$ = freight rate,  $m$ = distance to market. The model was able to determine the price of the land as a

function of the production, market price and transportation costs. In the theory of Von Thunen the question of the distance and the transportation costs was of great importance and in fact he realized a model based on concentric rings in which the central ring represent the city, and the sequent concentric rings are as follow: the second is the forest for fuel, the third is the grains and field crops, the fourth is ranching, and the fifth is an area of wilderness where the possibility to gain profits with agricultural activity is zero.

*Some critique at the definition of Von Thunen.* The ideas of Von Thunen of concentric rings, that is in the definition of the value of the land as inversely related to the increasing level of distance from the city centers can be considered valid also in the contemporary application of the economic geography. But there are also many limitations in the idea of Von Thunen that is especially the fact that the role of land rent is really marginal in the process of cost analysis.

- **Alfred Weber:** was the brother of the well-known sociologist Max Weber. Alfred Weber was professor at the University of Heidelberg. He was interested in the question of the maximization of transportation function in the least cost theory. The question that has been introduced by Weber consists in the finding a point in the plane that minimizes the sum of the transportation costs from the central point to n destination point, in presence of different destination points and different cost per unit distance.

**The business location.** (Mody & Srinivasan, 1998) analyses the role of political economy in the decision making of firms in respect to location. Authors consider the existence of positive externalities in the business location processes. There is also an agglomerated effect: firms tend to locate near other firms in the temptation to acquire new positive externalities. Fiscal incentives can induce firms to prefer location in governmental predetermined areas.

(Lin, et al., 2016, July) analyses the question of business location by the mean of the principle of proximity and vicinity. The authors sustain the crucial role of business location for the economic development of a certain area. The role of business location develops its effects either in national, regional and local level.

(Bartik, 1988) analyses the effects of environmental regulation on the location of manufacturing by Fortune 500 companies. For most industries, authors have found a statistically significance effect of environmental regulation on business location. Especially the effect of environmental regulation is higher for polluting industries. Policy makers in the process of environmental regulation can have a relevant effect for firms having a higher impact on pollution. (Bartik, 1988) shows the role of policy makers in the complex process of shaping business location preference setting regulation.

(Rohlin, et al., 2014) afford the question of the relation between taxation and business activity. (Rohlin, et al., 2014) analyzed changes in taxation coming at a certain time in cross-border. Changes can be realized to understand what the optimal taxation strategy is able to increase business location. Authors analyze the effect of income taxation on business location. (Rohlin, et al., 2014) have showed that higher personal income taxes are associated to a higher probability that the firm is crossing the border, while the increasing of income tax and sales tax have an opposite effect. (Rohlin, et al., 2014) suggest that a role is played by the presence of agreement among sharing borders State. Results suggests that state level tax policies can affect the entrepreneur's decision to realize a certain level of business location.

(Jensen, et al., 2015) analyses the possibility of using business location as a way to realize a competition between two countries. Authors analyze domestic politics in the United States during the period 1999-2012. (Jensen, et al., 2015) analyze the role of tax deduction and exemption to increase the possibility of a firm to locate a business in a certain locality. (Jensen, et al., 2015) have analyzed data from different municipal programs, examining the way in which electoral competition can shape public domain incentive to business location. Authors analyze the role of incentive in the process of building political and economic programs. Authors suggest the possibility of the existence of a misallocation of information in respect to incentive devoted to business location. Incentives to business location can be simply too high or too low. Municipalities suffer from tax competition devoted to increase the level of business location. But tax competition can have positive effects on the way in which politicians can be chosen by voters. Business location-oriented politician devoted

to reduce taxes or to increase incentives can have greater possibility to be re-elected in respect to politicians having no attention to business location.

(Pajones, et al., 2017) afford the question of business location in connection with the question of sustainability: authors sustain that Corporate Social Responsibility is a crucial topic. The question of sustainability is crucial element even to business location too. (Pajones, et al., 2017) shows that business location should be analyzed also to consider the role of sustainability either in the sense of economic, ecological, and social sustainability business location. (Pajones, et al., 2017) develops a matrix able to show the relation between environment, sustainability and business location. Due to the presence of the factor matrix it is possible to analyze a series of factors that can increase the efficiency of business location.

(Romadona, et al., 2017) analyze data devoted to understand the possibility of business to be located in Taman Siswa Street as a central place theory using a logistic regression with the software E-views 8 as an analytical method employed. The results of the analysis show that business environment can influence business location decision in the area of Taman Siswa Street.

(Jacob Trip, 2007) afford the question of urban quality as a way to develop a long-term urban competitiveness focusing on the Zuidas project in Amsterdam. Zuidas is a process designed to realize the attraction of international business location. But even if Zuidas is based on diversity and on the idea of urban quality of life as a way to increase the attractiveness in respect of business, the project has showed a low level of efficiency. The paper suggests some tools to increase the efficiency of Zuidas project in Amsterdam that is the ability to make the process more flexible and open.

(Xu, et al., 2018, May) consider the complexity of business location selection. Business location is a costly process. (Xu, et al., 2018, May) considers a series of elements able to show the role of business location that are foot traffic, neighbourhood structure, space rent and available workforce. The metric analysis has been conducted using satellite data produced either from satellite images and from night time light data. A comparison between a map generated from satellite and the geography of urban business location has been realized. After collecting data, (Xu, et al., 2018, May) have performed a neural network model oriented to predict the popularity and ranking for each location.

(Chavda, 2004) analyses a set of factors able to guide the process of decision making. Authors consider manufacturing, financial and technological firms. (Chavda, 2004) shows that different industries have different preferences in terms of services offered by municipalities. Manufacturing firms looking for municipalities that offer cost-cutting methodologies, while technological and financial firms prefer to locate business in context characterized by the presence of cultural scenes able to offer well-educated employees.

**Superstore location.** (Somekh, 2012-2013) consider the superstore location as a case of study realized in the context of household location. (Somekh, 2012-2013) has the intend to analyze how location choices can shape a city. (Somekh, 2012-2013) analyze the relation between household choices and shopping behavior. (Somekh, 2012-2013) try to apply a model to understand how superstore can optimize location choices based on household prices. One of the essential assumptions of the cited paper is the presence of income segregation in the analyzed cities. The presence of income segregation shape consumer choices. Consumer can prefer more expensive shop to superstore in the presence of high-level income. (Somekh, 2012-2013) analyze also the level of population in connection to superstore behavior. (Somekh, 2012-2013) finding show that the presence of superstore reduces the level of rental prices incrementing the possibilities for consumer to spend more in consumption. (Somekh, 2012-2013) sustains that the presence of a superstore increases the level of wellness of any kind of consumers. At the same time (Somekh, 2012-2013) show that the location of a superstore generates an increase in the inequality due the presence of income segregation. Rich people seem to benefit from the reduction of the life expenditure due to the presence of the superstore more than poor. In effect the reduction of the costs of the poor's increase the possibility of income segregation creating an increase in social exclusion and in social inequality.

(Poynor, 1984) analyses the question of the optimal location of a shop. In the cited paper the author analyses the results of the conference "*Techniques for Shop Location*" realized in London. (Poynor, 1984) presents counterfactual arguments developed by conference participants. The main ideas developed during the conference regard two alternative juxtaposed idea based on the evaluation of the centres or in contrast the evaluation of the peripheries. (Poynor, 1984) during the debate has

been developed an accusation in respect to local politicians devoted to support the development of superstore. But during the debate as it has reported in (Poynor, 1984) also emerged the possibilities that delocalization of shopping outside the cities in peripheries can be also a way to develop depressed areas.

(Dubé, et al., 2016) analyses the level of location theory. (Dubé, et al., 2016) afford the question of the relation between individual business establishments and the characterization of local economic environment to analyze the presence of economic determinants able to describe the business success (Dubé, et al., 2016) Authors use a micro-spatial perspective able to analyze the distance-based measures to serve in a discrete choice model (DCM). Authors use a 2006 database of individual business establishment in the Lower St-Lawrence region to provide an empirical analysis of the determinants of location decision in the relation between main economic activity and a random utility model (RUM) framework. Results shows the distance to nearby centres, and the co-location i.e. the location in the same area of similar industries, together with the size of establishment are statistically related to location decisions. (Dubé, et al., 2016) show that the presence of colocation is not able to increase location decisions. Also manufacturing distance from market has a negative effect for business location decisions.

(Zünd, et al., 2016) analyses the question of urbanization and its ability to shape either the residential choices either the business locations choices. Even if is difficult to analyze all the complex elements able to shape the planning and design option of urban systems (Zünd, et al., 2016) have analyzed a model that is based on the possibility to increase the level of feedbacks associated to a certain intervention. (Zünd, et al., 2016) authors have analyzed the level of residential and business agglomeration in the stage of modelling and simulation process. A model devoted to capture distances among centre of business and residential choice has been introduced to analyze the level of network in area. Each sub-area is represented through the level of maximum likelihood equilibrium model based on residences and firms. Data from Zurich have been analyzed.

(Wood, et al., 2016) analyze the importance of strategic elements in the process of business location. The authors afford the international strategy of Tesco in entering in the Asian market introducing key elements such as for example the question of strategic location, corporate culture and regulatory and institutional determinants. Clearly the question of superstore location acquires a deeper meaning in the case of internationalization for the fact that the firm interested in entering in a new market in a foreign country should respect a series of implicit and explicit norms and covenants.

(Coca-Stefaniak, et al., 2010) afford the question of localization in two different areas that are Spain and Scotland based on three key themes that are place, people and promotion. The authors have isolated some elements that are able to increase the probability of a successful localization that are: customer service, community embeddedness and informal relations between shop owner and customers. The presence of weak relations among consumers and retails is an argument against globalization due to the fact that globalized firms and productive services are not able to replicate or substitute the relational capital of local based activity. Localization considered with this particular characteristic affecting the relationship between customers and retails seem to be strongly resilient in respect to the massified threats of the globalization.

(Rigby & Vishwanath, 2006) afford the relationship between localization and customization of retail services on local bases i.e. the ability of the retailers to offer customized services able to better encounter the necessity of local communities. Retailers face a challenge between standardization and customizations. The challenge is complicated by the necessity to optimize the localization choice. In fact, and this is true especially for franchising (Galiano, et al., 2018), there exist the possibility for firm to standardize some services especially for example for products and brand, but at the same time, firms should realize some kind of customization in the localization process, trying to respect either implicit and explicit norms that regulate the process of urbanization of retailers. This create a twofold tension between the process of standardization and the process of customization of retail services, and effectively reduce the level of standardization creating more diversified local retails that are able to better interpret the necessity of the local community.

(Shenghao, 2009) analyses the methodology of localization in Korea of tow different retails that are Tesco and Carrefour. In particular the author considers the different locational strategies that are realized based on a series of elements such as for example the presence of band culture, the operational mode, the marketing technique, the human resource and operation management.

(Grimaldi, et al., 2019) the necessity to implement urban policies that are able to reduce the abandon and desertification of certain areas. In effect authors consider that after some adverse events such as for example the global financial crisis, many areas are effectively in bad conditions of abandon and reduction of the presence of shops and retails. The authors develop an heuristics able to consider the convergence of three dimension to solve the question of the location of retails in abandoned metropolitan areas that are in the mix among social elements, business and technology. The results is a tool for city managers able to reduce desertification of retails in certain areas and at the same time reducing the mono-business activity.

### 3. DATA MINING

Data are analyzed building clusters. Regions are divided in 4 different cluster. Clusterized data have been put together in the context of correlations analysis. Data have been elaborated by using KMINE. The clusterized analysis shows a positive correlation between the number of superstores, the level of nominal GDP, the level of population and the level of employments. In the sequent passages the methodological approach is analyzed.

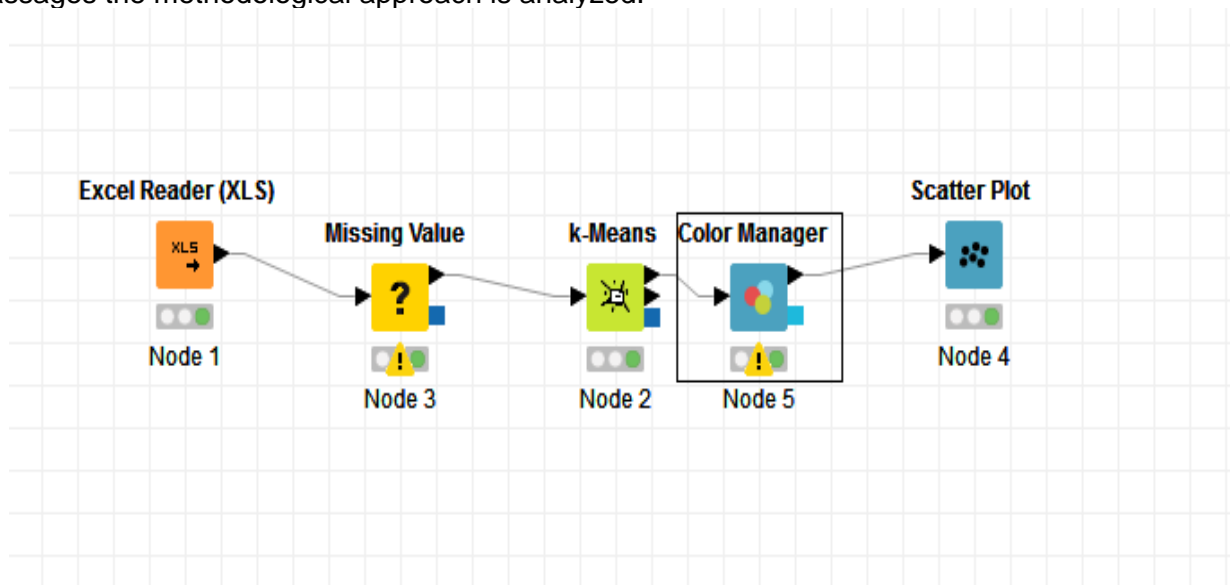


Figure 2. Workflow for clusterization process in KNIME.

In the figure 2 is analyzed the process of data analysis based with KNIME. KNIME is a free software for data analysis and machine learning. The first passage that is the Node 1 read the excel file. Node 3 eliminates the missing values. Node 2 applies the K-Means (Wagstaff, et al., 2001) algorithm that perform as a tool for clusterization. Clusterization has been realized as a preliminary condition to develop the complex set of data used to test OLS and Panel Data regression with fixed and random effects. Colour manager is a function of KNIME that create coloured label for variables. After having realized a change in clusterized variables using the KNIME colour managers, data are finally plotted to visualization. Visualization is an essential part of the process of data exploitation due to the fact that the visual relations among different variables is able to shed lights on aspects, and new features. For example visualization can suggests new relations among variables and new path to realize the investigation of data explaining the meaning of data driven information.



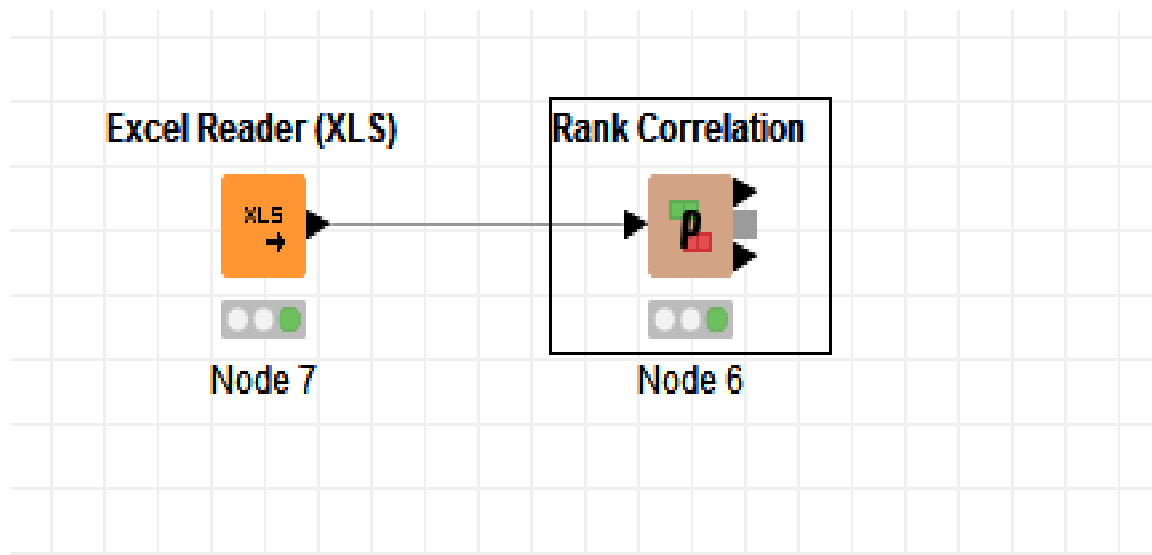


Figure 3. Workflow used in KNIME to generate ranked clusters based on excel reading.

In the figure 3 is represented the workflow uses in KNIME to realized correlation. The node 7 works as and Excel Reader while the node 6 operates effectively the rank correlation.

**The role of visualization of data as a tool for thinking.** Data visualization is not only a way to give representation to data. Data visualization is a way to extract meaning from data (Elouni, et al., 2016). Data visualization affect the cognitive process of the researcher, analysts or decision maker to suggest the tools and the analytical strategy to extract meaning from data. The process of data visualization is a process of extract meaning. Knowledge is embodied in the context of data and using visualization process can increase the ability of analytics to generating sense and meaning (Molcho & Schneor, 2015). The quantitative determination of data has a new sense in the context of data visualization. In effect data visualization creates the conditions to transform the obscure quantitative determination of data in a qualitative meaning able to increase the cognitive ability of the observer in creating deeper meaningful analyses. The passage between the quantitative skinny structure of data to a qualitative analysis is realized using through the implementation of data visualization techniques.

We analyze data to investigate the impact of some variables such as:

**Surface based clusters:** is a measure of the square meters of superstore. The dimension of superstore is a quantitative essential prerequisite of a store to be considered as a superstore. The Italian Minister of Economic Development-MISE, has created an observatory that is the “*National Observatory on Commerce*” (MISE, 2019) to analyze the degree, diffusion and square meters of the stores. The “*National Observatory of the Commerce*” reports data from italian regions and for each region the observatory present the category of the superstore.

**Purchasing group based clusters:** are clusters of firms that operate in italian GDO that is the acronym of “*Grande Distribuzione Organizzata*”. These firms create purchasing groups to buy product to optimize the supply chain to obtain lower prices and to increase the level of efficiency (Zimon, 2019). Many firms try to reduce the cost of products and services and create aggregation able to increase the level of the economy of scale. Purchasing groups are associations of firms based that maximizes their commercial power to reduce the costs of inputs and to increase the level of market size. Firms that participate in the context of purchasing group have the possibility to increase the value of their own brand and to improve the supply of products and services in respect to the final customer. Firms that participate in purchasing groups show higher level of sales and profits. The strategy to create purchasing groups operates as a multi-objective long run strategy for the firms especially for the fact that it consents to improve market share, to maximize brand value, to improve the supply of product and increase also the market power. But at the same time, for the point of view of the market as a whole, it is also necessary to consider the risk of creation of oligopolistic and monopolistic markets, that are realized through the reduction of the competition in the market. In effect the creation of purchasing groups in the context of distribution and logistics has the ability to transform meddle size firms and corporations in big corporation, with an increase in the dimension

of the players that can effectively reduce market competition (Zimon, 2017). The reduction in market competition can have negative effects especially for the customer that are manifested in the substitution of brand product with commercial product that are product realized by the purchasing group. The creation of purchasing group has positive for the profits and sales of the participating firms but it has negative effects for the market as a whole for the fact that the creation of purchasing group can reduce the level of competition creating conditions for monopolistic and oligarchic markets.

**Nominal GDP based clusters:** nominal GDP cluster are realized with data from ISTAT to give a representation of the value of nominal income among Italian regions. Data are clusterized. These data are reported due to the fact that the higher value of nominal GDP is associated with higher presence of superstores. In effect superstore are more diffused in regions that have higher nominal GDP such as the regions of the North Italy. The positive relation between the presence of superstores and the level of nominal GDP is counterfactually confirmed by analysing the presence of superstore in south of Italy, where a lower level of nominal GDP is associated to a lower presence of superstores. Superstores are positively associated to the level of nominal GDP.

**Population based clusters:** these are data that represent the value of population distribution in Italian regions in the considered period, based on clusterized data. The increasing level of population is positively associated with an increase in the level of superstore locations. This is due to the fact that superstores offer goods and services for populations especially for population that live in the suburban areas of the cities. The greater the population the greater the number of superstores located. This positive relation is also counterfactual confirmed by the fact that regions that have low level of population have also a low presence of superstore. This can be considered as a derivation of the fact that superstores are oriented to offer their product to urban masses. In effect in the regions of North of Italy, where cities are more developed and urbanized in respect to that of South of Italy, the presence of superstore is higher than in respect to the South of Italy. This means that there is a positive relation between population and the presence of superstore at a regional degree.

**Occupation based clusters or Employment:** Occupational based clusters represent clusters based on the distribution of population in the considered period in Italian regions. The level of employment is able to explain the presence of superstore location in Italian regions. In effect we can say that superstores are more common in areas that are characterized by the presence of high level of employment. But this relation should be better understood considering the fact that employment levels are very high in North of Italy, while at the contrary they are very low in the regions of South of Italy. It is possible that the macro-economic conditions of the Italian regions has an impact in the distribution of the superstore in the national territory. The level of employment is positively associated with the level of the presence of superstore location for the fact that the increasing level of employment generated also an increasing level of consumption and the level of consumption can be considered as a driven tool for the choice of superstore location. The level of income is essential in the context of the determination of superstore location. It is necessary to consider that the higher level of employment is associated to a higher level of consumption, and the choice of superstore location is determined on the basis of financial context.

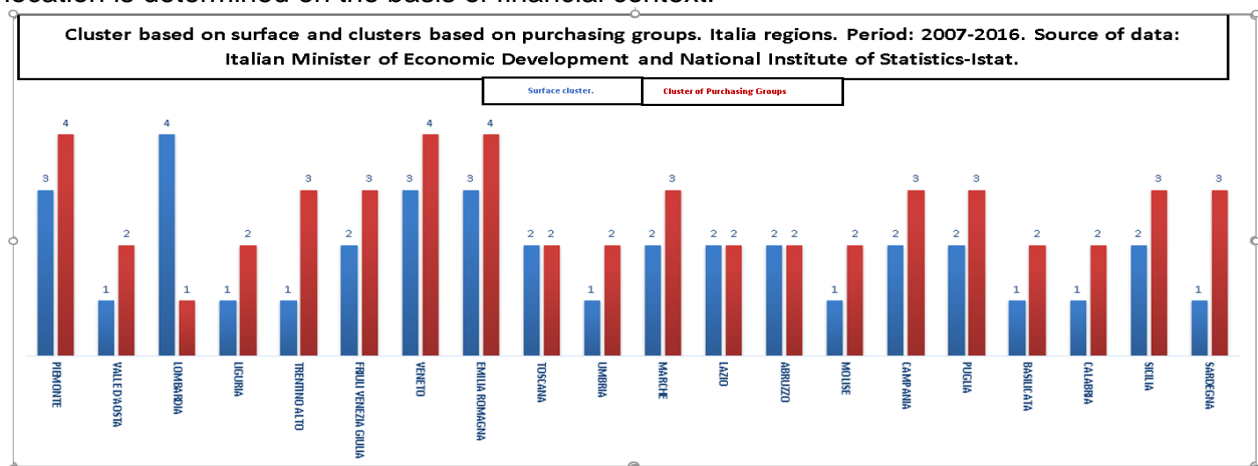


Figure 4. A comparison among clusterized between surface cluster and clusters of purchasing group. Clusterization is realized in KNIME.

In particular we can see that there is a large inequality in the distribution of cluster in the three Italian macro-regions that are Northern Italy, Central Italy and Southern Italy. In fact if we consider the number of surface cluster for superstores we found that they are 18 or equivalent to 50,00% of the national market in the Northern Italy, 7 in the Central Italy equivalent to the 19,4% of the national market, and 11 in the Southern Italy equivalent to the 30,6% of the national market. Similar differences are also in the number of clusters for purchasing group. The number of clusters for purchasing group in the Northern Italy is equal to 23 that is the 46,00% of the national market, while in the Central Italy the number is equal to 9 that is the 18,00% of the national market, and in the Southern Italy the number of cluster for purchasing group is equal to 18 that is the 36,00% of the national market. As we can see from the clusterization process it is clear that the number of superstores is greater in the context of North of Italy, characterized by the presence of regions with greater income, more populated cities and a more developed labour market.

The geographical distribution of the presence of superstore is the representation of the “*italian divario*” (Felice, 2011) that is the divergence of the regional economies of the North of Italy, more oriented in respect to the richer and opulent center Europe, and the economic condition of the South of Italy, that is characterized by a stagnant economy characterized by low level of employment and low level of Gdp growth. Piedmont and Lombardy have the highest number of superstores clusterized in the sense of surface, while purchasing group are diffuse not only in Piedmont but also in Veneto and Emilia Romagna. At the contrary the figure 4 shows that the Italian regions of the South of Italy are characterized by a less developed market with a low number of superstore and also middle and low level of purchasing groups, even if also in the South of Italy is possible to consider the presence of regions that are able to have a good perform that is the case of Campania and Sicily. In this sense clusterized data shows the presence of a the “*Questione meridionale*” (Gramsci, 2008).

<b>KMINE BASE CLUSTERIZATION. Source of data: Italian Minister of Economic Development and Italian National Institute of Statistics. Period of analysis: 2007-2014.</b>					
<b>Regions</b>	<b>Surface based clusters</b>	<b>Purchasing group-based clusters</b>	<b>Nominal GDP based clusters</b>	<b>Population based clusters</b>	<b>Occupation based clusters</b>
<b>PIEMONTE</b>	3	4	3	3	3
<b>VALLE D'AOSTA</b>	1	2	1	1	1
<b>LOMBARDIA</b>	4	1	4	4	4
<b>LIGURIA</b>	1	2	2	2	2
<b>TRENTINO ALTO</b>	1	3	2	2	2
<b>FRIULI VENEZIA GIULIA</b>	2	3	2	2	2
<b>VENETO</b>	3	4	3	3	3
<b>EMILIA ROMAGNA</b>	3	4	3	3	3
<b>TOSCANA</b>	2	2	3	3	3
<b>UMBRIA</b>	1	2	1	2	1
<b>MARCHE</b>	2	3	2	2	2
<b>LAZIO</b>	2	2	3	3	3
<b>ABRUZZO</b>	2	2	2	2	2
<b>MOLISE</b>	1	2	1	1	1
<b>CAMPANIA</b>	2	3	3	3	3
<b>PUGLIA</b>	2	3	2	3	3
<b>BASILICATA</b>	1	2	1	1	1
<b>CALABRIA</b>	1	2	2	2	2
<b>SICILIA</b>	2	3	3	3	3
<b>SARDEGNA</b>	1	3	2	2	2

Figure 5. KNIME based clusterization. Source of data: Italian Minister of Economic Development and Italian National Institute of Statistics. Period of analysis: 2007-2016.

The clusterization seems to confirm the substantial differences existing among the different Italian regions. The presence of the Italian divide and the accumulation of financial, and economic power in the context of northern Italian regions. The presence of a differentiated economic situation between North of Italy and South of Italy has an impact also for the strategic definition of marketing orientation in the context of investment planning and business development programming. In effect firms interested in the investment in the sector have to diversify their offer, creating at least two plans one for the North of Italy and the other for the South of Italy, paying attention for the differences in the socio-economic conditions and perspective of the two differentiated areas.

As we can see in the context of figure 6, that is a graphical representation of the clustered data, there are positive relations among the studied variables that tend to grow together even in the clusterization process. The correlation matrix that has been realized with the connected variables. All the variables considered are positively correlated but the main positive correlation is that between surface based clusters and nominal GDP (0.825) and surface cluster and total population (0.825) and surface cluster and total employment (0.838). Purchasing cluster are also positively associated to surface cluster (0.397) and also to nominal GDP (0.3) total population (0.297) and total employment (0.322). The correlation matrix shows the presence of a positive relation between the level of Cluster surface and the level of cluster purchase group that is weak and equal to 0.397. This means that are in general middle size superstore that realize purchasing groups while the superstore with high surfaces are not interested in the building of purchasing groups for the fact that they have the sufficient extension to organized a better offer in respect to customers and clients.

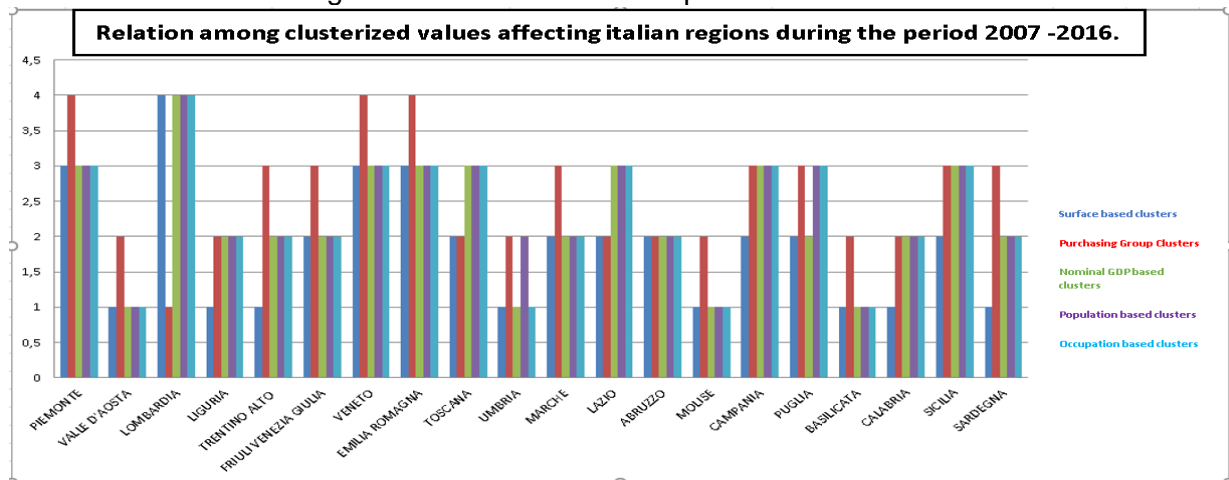


Figure 6. Relation among clusterized values affecting Italian regions during the period 2007-2016. Data Source Italian Minister of Economic Development and Italian National Institute of Statistics.

Correlation Matrix - 0:6 - Rank Correlation					
File	View				
Row ID	Cluster Superficie	Cluster Gruppi Acquisto	PIL (Milioni di Euro)	Totale Popolazione	Totale occupazione
Cluster Superficie	1	0.397	0.825	0.825	0.838
Cluster Gruppi A...	0.397	1	0.3	0.297	0.322
PIL (Milioni di Euro)	0.825	0.3	1	0.929	0.956
Totale Popolazione	0.825	0.297	0.929	1	0.977
Totale occupazione	0.838	0.322	0.956	0.977	1

Figure 7. Correlation matrix based on clusterized data. Source: Italian Minister of Economic Development and Italian National Institute of Statistics. Data shows the presence of positive relations among variables. Clusterization process has been realized in a period between 2007 and 2016.

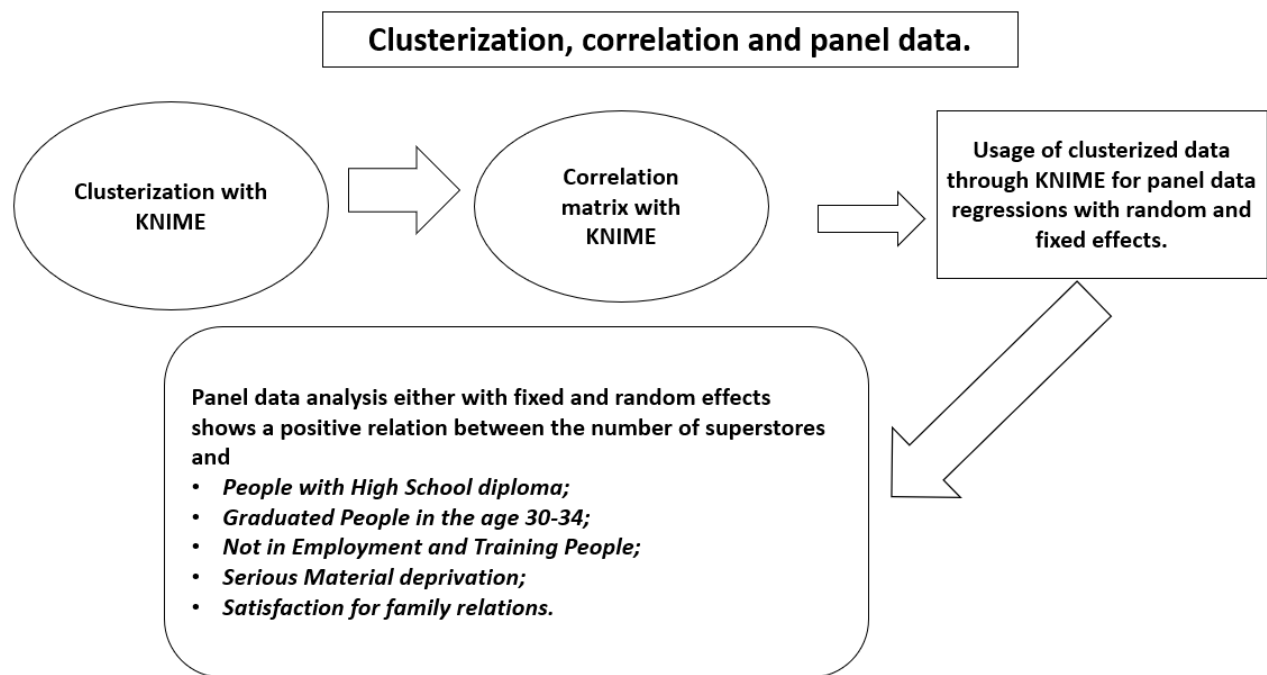


Figure 8. Flow chart able to describe the analytical process started with clusterization in KNIME, through correlation matrix, arriving to econometric analysis

In conclusion the data based on the clusterized model suggests to realize a superstore location in regions characterized by the presence of increasing nominal GDP, an increasing level of employment and an increasing level of population.

#### 4. RESULTS AND DISCUSSION

The model is devoted to estimate the number of superstores in Italian regions and macro-regions. The model is devoted to estimate the level of number superstore based on a series of dependent variables. All the variables that are positively and negatively associated to the location of superstores are statistically significant with a p-value equal to 0.000, with the exception of the model estimated with WLS. The localization choice can be driven by different motivations. Our estimation is based on findings that have found some positive elements that are able to positively determine the choice of superstore location and some elements that are negatively associated with superstore location.

<b>Socio-economic determinants that positively affect the choice of superstore location.</b>	
<b><i>People with High School Diploma</i></b>	People with high school diploma are more common in peripheries. Since superstores generally are located in peripheries there is a positive relation between superstore location and people with high school diploma.
<b><i>Graduated People in the age 30-34</i></b>	Graduated People in the age 30-34 is generally associated with low income, generally present in peripheries and low-income urban areas.
<b><i>Not in Employment and Training people</i></b>	The presence of Neet is increasing in low income areas, such as suburban peripheries that are generally selected to install superstores.
<b><i>Serious material deprivation</i></b>	Serious material deprivation is more common in peripheries and marginal urban areas that are generally the choice for the superstore location.
<b><i>Satisfaction for family relations</i></b>	Satisfaction for family relations is higher in the area that are characterized by the presence of suburban areas. The increasing in

	the income of individuals is associated with low fertility rates, and low satisfaction for family relation due to the absence of life-work balance. So the presence of satisfaction for family relations is a characteristic of low income families in peripheries.
<b><i>Social Participation</i></b>	Social participation is positively associated with the presence of superstore location. Social participation tends, that is the shape of a communitarian life, is positively associated with low income areas. In effect in high level income area there is low participation to communities. Low income urban areas are more prone to develop social participation.

In particular we found that the location of superstore is associated to a series of variables that are characterized by the presence of poor human capital and by a certain degree of social participation. To explain this relation, it is important to consider the fact that superstores and poor human capital are in general located in the same areas of the cities. In particular superstores are located in peripheries that are generally associated with poor human capital. For these reasons the location of superstore is positively and significantly associated with the number of people having high school Diploma, the number of people graduated in the age 30-34, the presence of NEET, Serious material deprivation. But at the same time peripheries are also the domain of households and for this reason the presence of superstore is associated positively with the satisfaction for family relations and social participation. It can be sound strange that peripheries, that are generally associated with low income and low levels of consumption, are also the urban scenario for the adfirmation of superstore, but it is easier to understand the locational choice considering the fact that peripheries offer also generally adequate space and transportation facilities that can sustain the migration of urban population in the direction center-periphery either in centripetal and centrifugal sense.

As we can see the possibility to realize an efficient superstore location is associated with the presence of socio-economic and demographic characteristics that are more developed in urban and peripheral areas.

<b>Socio-economic determinants that negatively affect the choice of superstore location.</b>	
<b><i>Lifelong Learning Participation</i></b>	The presence of people that are involved in lifelong learning participation is negatively associated with the presence of superstore location. People that can have access to lifelong learning participation is generally characterized by middle-high income and is located in central areas of the cities. While people living in suburbs or peripheries has low probability to have access to Lifelong Learning participation. For these reasons there is a negative association between the presence of lifelong learning and the superstore localization.
<b><i>Cultural Participation</i></b>	Cultural participation is a characteristic of the middle high-income class. People with medio and high-level income can spend money and time to acquire some kind of cultural experience and participation. But in the case of cultural participation in the peripheries and suburban areas the cultural participation is low, and at the limit equal to zero. So in the places in which superstore are located the presence of cultural participation is low.
<b><i>Employment</i></b>	The level of employment rate is negatively associated with the presence of superstore location. Superstore location are associated to low income areas in which there is low employment. Due to this element the presence of employment is negatively associated with the presence of superstore location.
<b><i>Unemployment rate</i></b>	Unemployment rate is negatively associated with the presence of Superstore location. The reason can seem counterfactual, but it is possible o argue in peripheries the number of discourage workers tends to be increasing. So, the fall of unemployment rate is determined by the relative increase in discouraged workers.

<b><i>The number of employees having short term contract in the last five years</i></b>	The number of people having short term contracts in the last five years is a measure of the dynamics of the labour markets. Due to the fact that labour markets are more efficient in the center of the cities and have a decreasing efficiency in suburban areas there is a negative relation between the presence of employees having short terms contract and the presence of superstore location.
<b><i>Employed with low income</i></b>	The presence of employed with low income is negatively associated with the presence of superstore. This can be considered difficult to understand but it is necessary to consider that in low income suburban areas there is an abundance of discouraged workers in respect to low income employed.
<b><i>Overeducated workers</i></b>	Overeducated workers are clearly associated negatively in respect to superstore location for the fact that overeducated workers are typically characteristics of city centers. Overeducated workers are poorly represented in suburban areas as for example in the context of peripheries. For these reasons there is a negative relationship between superstore location and the presence of overeducated workers.
<b><i>Non-regular employed</i></b>	Non regular employers are negatively associated with the presence of superstore location. This relation can seem counterfactual. But in reality, non-regular workers, as a form of flexible workers are considered as an element of complexity of the labour market that is characteristic of central cities.
<b><i>Involuntary part time</i></b>	The presence of a negative relation between involuntary part time and superstore location is the manifestation of the low level of labour market development in the peripheries. In effect in city centers, and more development urban areas the labour markets present an higher degree of complexity admitting also flexible contracts such as for example involuntary part- time.
<b><i>Poverty risk</i></b>	The risk of poverty is negatively associated to the presence of superstore. This can be considered as a controversial relation in respect to the presence of superstore due to the fact that in our hypothesis superstores are located in suburban areas that are by definition characterized by the presence of lower income classes.
<b><i>Serious Economic Difficulty</i></b>	Serious Economic Difficulty is negatively associated with the presence of superstore location. Also, this element is counterfactual, due to the fact that generally people with serious economic difficulties live in suburban areas. But again, as in the previous points, this analysis seems to depict a social stratification that is more complex in the context of city centers. While suburban cities seem oriented to a homogeneity of social class based on low level income but equally characterized by the absence of poverty, starvation and social exclusion.
<b><i>Low Working Intensity</i></b>	There is a negative relation between low working intensity and the presence of superstore location. Low working intensity has a low distribution in suburban areas while it seems more diffused in the context of city centers.
<b><i>Volunteering</i></b>	Volunteering is associated negatively with the presence of superstore. In effect volunteering seems more distributed in city centers that in respect to suburban areas. The presence of volunteering is low in suburban areas that are characterized by low degree of volunteering.
<b><i>Association financing</i></b>	Association financing consists in the fact that people, firms, organizations finance associations to pursue social goals. The level of association financing seems to be a condition of city centers, while

	suburban areas are characterized by low level of association financing.
<b>Generalized Trust</b>	The level of generalized trust is negatively associated with the presence of superstore location. Generalized trust is able to increase the level of economic activity, of entrepreneurship and of firm organization. The reduction of generalized trust is characterized of suburban areas that generally have low level of trust and are oriented to develop some kind of economic organization such as for example superstore.

In our analysis superstores are located in suburban areas. Suburban areas are characterized by the presence of low income and less educated people. These kinds of social classes are more distributed in the context of suburban areas and contribute specifically to the formation of the conditions preliminary to the diffusion of superstores. But while the model seems to be coherent with this idea of superstore location there at least some association of the superstore location that are counterfactual such as for example the variables that are related to the labour market. The variables that are characterized by a counterfactual relation in respect to superstore location are indicated subsequently:

- *Employment;*
- *Unemployment rate;*
- *The number of employees having short term contract in the last five years;*
- *Employed with low income;*
- *Non-regular employed;*
- *Involuntary part time;*
- *Low Working Intensity;*

Due to the presence of low skilled people, and low-income workers, in the context of suburban areas we will expect a positive relation at least in the case of unemployment rate, employed with low income, overeducated workers, non-regular employed, involuntary part time, low working intensity. Instead these variables are negatively associated to superstore location. The reason of this negative association with superstore location can be founded in the presence of a high level of discouraged workers in the suburban areas that reduce either the level of employment, either the level of unemployment. The result is that in suburban areas a low level of social stratification and labour contracts diversification is determined due to the low labour force participation. But the presence of employees having short term contract in the last five years, employed with low income, non-regular employed, involuntary part-time, low working intensity, in reality can be not only considered as characteristic of a low-income labour market, but also can be considered as a more complex labour market. In effect it is more probable that flexibilization of labour force is greater in a more complex labour market, such as that of city centers, than in poorly organized labour market such as that developed in the suburban areas. For this reason, we found a negative relation between form flexibility and precariousness in the labour market and the presence of superstore location: because flexibility and precariousness are associated to more complex labour market conditions that are deeper in the city center and absent in the context of suburban area.

At the same time the presence of superstore is negatively associated with the presence of a series of elements that are able to design a deeper cultural and educational environment such as for example Lifelong Learning participation, Cultural participation, Overeducated workers. At the same time the superstore location is negatively associated to the presence of a variables that are able to describe a deeper structured job market such as for example: employment, unemployment rate, the number of employees having short term contracts in the last five years, employed with low income, non-regular employed, involuntary part time, low working intensity. Thirdly we found that also elements that are able to capture poverty and material deprivation are negatively associated with the presence of superstore such as for example in the case of poverty risk, serious economic difficulty, low working intensity. Finally, there is a negative association also with variable that are able to define the level of trust in the society such as for example: Volunteering, association



financing, generalized trust. Syntetically we can find that the presence of a deeper cultural environment, the development of job market, conditions of poverty and deprivation, and the general level of the trust in the society, are negatively associated with the presence of superstore. Not all these elements are characteristics of the city center, for example the presence of conditions of poverty and material deprivation are characteristics of some peripheries. But these relations let us conclude that even if superstore are generally allocated in peripheries and not in the city centers, at the same time, not all peripheries can host a superstore, but only some kind of peripheries that are at a certain mean between the poverty of suburban areas at the wealth and culture of city centers. So, to better understand the role of relations in the context of our econometric analysis we have to distinguish:

- **Labour market conditions in the city centers:** labour market conditions in the city center are characterized by the presence of complexity. The complexity in the labour market conditions is a characteristic that increase the typology and diversity of labour contract and can be characterized by the presence of more flexible and precarious contracts. The presence of part time workers, an employee that have short term contracts in the last five years, and employed with low income, and low working intensity are characteristics of more complex labour markets and deeper labour specialization. Labour market conditions in the city centers can increase the diversity of labour contract distributed, and this can increase the level of part-time workers, low payed workers, working-poor, and low intensity workers. These areas all consequences of a deep structured labour market in which the presence of differentiated labour contracts is a sign not only of the complexity of the supply side of the labour market, but also of the diversity of demand side of labour market. City centers labour market have almost a precise relation between income payed for workers and output generated of the workers. For these reasons the relation between superstore location-that are located in suburban areas- and flexible and precarious labour contracts- that are located in the city centers- is negative.
- **Labour market condition in the suburban areas:** Labour market conditions in suburban cities are less developed in respect to city centers labour markets. This means that there is more homogeneity of labour contracts, and less presence of precarious and flexible labour contracts. In effect even if the presence of labor contracts characterized by flexibility and precariousness can be associated with low income and low income are more common in that suburban areas in which superstore generally locate. But not all the low-income labour contracts are associated with low income areas, since, especially for the case of part-time, these can be more efficiently associated with city centers and more complex labour markets. For this reason, our data suggest that labour market conditions in the suburban areas where superstore are located are also associated with low diversity of labour contracts with the low presence of precarious and flexibility.

So, the characteristics of the superstore location is based on the presence of a series of socio-economics determinants that are associated with low income, low skilled workers, less educated people. Suburban areas are accumulated by low level of human capital, that has low income, but due to the presence of low level of workforce labour participation there is also a negative relation between superstore location and flexible and precarious works.

<b>The differences between labour market in city centers and urban centers that have an impact on superstore location.</b>	
<b>City Centers</b>	City centers have deeper, interconnected and complex labour markets. The increasing in labour markets complexity determine also the presence of heterogenous and precarious labour contracts due to the increase in labour force participation. At the same time the increase in labour force participation determine an increase either in the level of employment either in the level of unemployment. In fact the reduction of the number of discouraged workers determine an increase in the level of labour force increasing either the level of

	employment either the level of unemployment. The complexity of the labour market in the city centers explain the reason why there is a negative relation between superstore location and some of the labour market variables that describe flexible and precarious labour markets.
<b>Suburban areas</b>	Suburban areas are characterized by the presence of less developed labour markets in respect to city centers. A lower level of complexity determines the reduction of the complexity of labour contracts, and the reduction in number of flexible and precarious workers. In this sense the level of flexible and precarious in suburban areas is low. For the same reason also the level of employment rate and unemployment rate is negatively associated with superstore location that are located in the suburban areas. The negative relation between employment and unemployment rate with respect to superstore location is due to the strong presence of discouraged workers in suburban cities. The increasing level of discouraged workers decreases either the level of employment and the level of the unemployment rate. In fact the increasing level of discouraged workers, that are workers that have renounced to find a job, reduces also the level of employment and the unemployment rate. For these reasons there is the counterfactual result that suburban areas that should be characterized by the presence of more flexible and precarious workers, are in reality, negatively associated with the presence of weak protected jobs.

Even if we have analyzed many socio-economic elements of the location of superstores, it is important to note that there are others factors that effectively managers should consider in the location of superstores. We have considered only 4 dimensions of the location choice that are instruction and formation, work and life conditions, economics wellness and social relations. But other important determinants are for example:

- **Institutional condition:** the presence of local and communal institutions that are interested in developing local political economy to improve commerce can act as an incentive to invest in a certain area. The localization of superstore should consider also the presence of local political economy that either can finance new locations either can design fiscal tools to improve the offer of commercial services for the population in a certain area. In effect the presence of a development local commerce system can promote some public goods that are valuable for the local municipality such as for example increasing the level of security, promote the economic growth, and the development of firms.
- **Access to credit for entrepreneurs:** the presence of banks and financial services in a certain area can incentivize the location of superstores. The presence of banks and financial services is the signal of the presence of a more active market that can sustain costumers in their consumption behavior. The development of consumer-based credit systems can promote compulsive and impulsive buying behavior that can act as a tool to improve extra-sales and extra-profits for superstores.
- **Development of commerce on a local base:** the development of an urban environment dedicated to customers is a tool to improve commerce on a local basis. In particular shops tend to be located in the same areas, due to the fact that the probability to improve the consumer buying behavior tends to increase with the presence of commercial activity in the retail sector. For example if the local municipality create an area that is devoted to promote commerce, there is an increasing probability of a superstore location in that area.
- **Dynamism of private sector:** the presence of a certain dynamism in the local entrepreneurial system can facilitate the location of superstores, due to the fact that consumers tend to concentrate their buying activity in certain local area. For example, the presence of different shops that are able to serve the interests of consumers in various areas such as for example in the housing sector, or in offering individual services, can improve the

consumer buying behavior and increment the probability of profits in the case of the location of a superstore.

- **Presence of public transportation:** the presence of public transportations services in the area in which the superstore can be located is of particular interest to solve the problem of the efficient location. In effect the profitability and the increase of the sales are positively connected to the presence of public transportation due to the fact that transportation can improve the numerosity of customers in a certain area.
- **Demography:** the number of people that live in a certain area, their age, their individual conditions and marital status is a relevant information to solve the question of the location of superstore. More favorable demographic conditions can improve the probability of an increase in profits and sales associated to a certain superstore location.

Further studies should be realized to analyze these new determinants to solve the question of superstore location.

## 5.CONCLUSION

We have analyzed the socio-economic determinants of superstore location with a comparison between cluster analysis and panel data. Data are collected from Italian Minister of Economic Development and ISTAT. Cluster analysis and panel data suggest the presence of a positive relation between GDP, Employment and Population and the number of superstores. To increase the profundity of the analysis we have used a panel data analysis showing that there is positive relation between the number of superstore and Instruction and Formation, Economic Wellness, Social Relations. Data shows the absence of a positive relation between the number of superstore and the level of Work and Life Conditions.

Our model consider the possibility that the choice of a superstore location is positively associated with the presence of People with High School Diploma, with Graduated People in the age 30-34, with the presence of Not in Employment and Training people, with the presence of Serious Material Deprivation, the value of Satisfaction for family relations, and the presence of Social Participation. All these elements can be considered has coherent with our main hypothesis that consider the superstore location realized in the context of suburban areas.

But at the same time there are different variables that can are negatively associated with the presence of superstores such as: Lifelong learning participation, Cultural Participation, Employment, Unemployment rate, the number of employees having short term contracts in the last five years, Employed with low income, Overeducated workers, Non-regular employed, Involuntary part-time, Poverty risk, Serious Economic Difficulty, Low Working Intensity, Volunteering, Association financing, Generalized Trusts. Among these elements that are negatively associated with the presence of superstore, there are some that seems at least formally in contradiction with the hypothesis of a superstore location in the context of suburban areas, and these are especially the labour variables. As it has been argued, if superstores are located in suburban areas than the relation among between superstore location and the variables that describe poverty, low income workers, and serious economic difficulty should be considered positively associated with the presence of superstore. But, in the estimated model based on BES-ISTAT there is a negative relation among these variables and the superstore location. This result can be considered counterfactual but effectively we have to consider that suburban area are connected with the presence of discouraged workers, and the increasing number of discouraged workers, that are workers that have negative expectations about the possibility to find a job, can reduce either the employment rate either the unemployment rate. But the presence of discouraged workers can only partially solve the large negative relation existing among labour variables and superstore location. To better explain this point it is necessary to understand that the differences between city center labour market and suburban labour market. City center labour markets are more complete in respect to suburban labour market and for this reason in the city center there is a more intense presence of flexible and precarious labour contracts either in the form of involuntary part-time. In effect in the city center there is a more complex labour market either in the sense of offer, especially for precarious and flexible job offer, and, especially in the sense of demand, for the presence of more opportunities to workers. The

reasons for the presence of a negative relation between the presence of low income worker and superstore location can be considered as the manifestation of a less development labour market in the suburban areas, a fact that generate an inverse relation between labour market variables and superstore location.

## 6. LIMITATIONS

The article presents two main limitations:

- *data are exclusively related to italian superstores.* This limitation impedes to apply the propositions to non-italian superstores with the logical consequence that the results are valid only for the italian market;
- *methodology used is traditional.* Other methodologies could be used such as for example neural networks, decision trees or principal factor analysis. The consequence of this methodological choice is the reduction of the knowledge detection process.

## 7. ACKNOWLEDGMENT

The present paper has been financed in application of the project «*Bigdata/B.I. per il volantino dinamico profilato dalla clientela e per il supporto alle decisioni basato su analisi predittive, analisi/utilizzo "Social" e su ricerche di mercato in ambito G.D.O.: "GDO-DSS Dynamic Intelligence"*».

## 8. APPENDIX

### Independent variable:

- **Number of superstores:** Superstores are store that have a surface of at least 2.500 square meters. Superstore can have two sections: food and non-food. Superstore generally are realized by really big corporations having shares quoted in financial markets (Italian Minister of Economic Development, 2018).

### Dependent variable (Istat-BES, 2018):

- **People with high school diploma (25-64 years):** Percentage of people in the age 25-64 that have a high school diploma on the total of people in the age 25-64;
- **Graduated People 30-34 years:** Percentage of people in the age 30-34 that have a university level degree on the total of people in the age 30-34;
- **Not in Employment and Education and Training people:** Percentage of persons in the age 15-29 not in employment and education and training on the total amount of people having an age 15-29;
- **Lifelong Learning participation:** Percentage of persons in the age 15-64 that have been enrolled in a process of instruction and formation in the four weeks before the interview on the total amount of person in the age 25-64;
- **Cultural Partecipation:** Percentage of persons with more than 6 years that in 12 months before the interview have developed more than three or more cultural activity on the total amount of people. Cultural activities are considered as cinema, theatre, concert, and reading newspapers and books;
- **Employment Rate 20-64:** Percentage of employed with 20-64 years on a population of 20-64 years;
- **Unemployment Rate:** Percentage of unemployment in the age 15- 74 years;

- **Employed in short term contracts in the last five years:** Percentage of employed with short renewable work contracts since last five years on the total of employed and collaborators in the short term;
- **Employed with low income:** Percentage of employed with hour income less than 2/3 of the median income of the total of employed;
- **Over-educated workers:** Percentage of workers that have a degree higher than necessary for a certain work on the total of workers;
- **Non-regular employed:** Percentage of worked that fail to respect fiscal, financial, and labour laws on the total amount of workers;
- **Employment rate of woman with 25-49 years with son and daughters in respect to percentage of woman without son and daughters:** Employment rate of mothers in the age 25-49 years in respect of the employment rate of non-mother woman in the same age per 100;
- **Involuntary Part Time:** Percentage of employed that have a part time job due to the absence of full-time job on the total amount of employed;
- **Mean Gdp per capita:** Mean Gdp per capita available per families and total number of resident persons;
- **Poverty Risk:** Percentage of persons that live a poverty risk with an income less of 60% of the equivalent median income of resident persons;
- **Serious material deprivation:** Percentage of persons that live in families with at least 4 of 9 problems considered on the total amount of resident persons.
- **Serious Economic Difficulty:** Part of person in families that summing all the incomes declare to having serious difficulties in affording monthly costs;
- **Low working intensity:** Percentage of persons living in families in which the labour intensity is less than 0,20. Incidence of persons that live in families where persons in working age have worked less than 20% of their working capabilities in the last year;
- **Satisfaction for family relations:** Percentage of persons of 14 years that show an high level of satisfaction for familiar relations;
- **Social Participation:** Persons with more than 14 years that in the last 12 years have participated at least one time to social activity on the total population of the same age.
- **Volunteering:** Person of 14 years and more that in the last 12 years have realized some non-remunerated activity for associations or volunteering groups on the total amount of persons in the same age;
- **Association financing:** Persons of more than 14 years that in the last 12 years have financed association on the total amount of persons of the same age;
- **Generalized Trust:** Percentage of persons with more than 14 years believing that trust people in general on the total amount of people of the same age;
- **Crowded Penitentiary:** Percentage of incarcerated persons in penitentiary on the total of the available number of persons regularly hosted;
- **Homicides:** Number of homicides on the total amount of the population pe 100.000 persons;
- **Theft in residence:** Number of thefts in residence on the total amount of households\*1000;
- **Robbery:** Number of robberies per 1.000 habitants;
- **Satisfaction for free time:** Percentage of persons of 14 years that are high satisfied of the free time on the total amount of persons in the same age;
- **Current Municipal Expenditure for the protection of cultural and monumental assets:** Municipal Expenditure for libraries, museum, and art gallery;
- **Unregulated Building:** Number of unregulated building for 100 authorized buildings from Municipalities;
- **Knowledge workers:** Percentage of workers with a university instruction in technological and scientific professions on the total amount of workers;

- **Children that have used municipal service for infancy:** Percentage of children between the age 0-2 that have used services for infancy based on Municipalities on the total amount of children 0-2;
- **Old Age persons with domiciliary assistance:** Percentage of person in old age that receive domiciliary assistance on the total amount of the person in the old age. Person in old age have more than 65 ages;
- **Difficulty at the access to certain services:** Percentage of families declaring to having difficulties to use essential services as for example on the total amount of families;
- **Irregularity in water distribution:** Percentage of families that suffers for irregularities in water distribution;
- **Irregularity in electric service:** Number for client suffering for irregularity in electric service distribution;

**Number of km offered from the public transportation:** Product of number in kms of public transportation service per mean of the kms divided the total number of resident persons

The model has been estimated by using either a model with fixed effects, a model with random effects.

The econometric models estimated a model with 46 variables that are defined to analyze the socio-economic determinants of superstore location. The model has been realized with a process of approximation. The general the general model is as indicated in the sequent formula:

$$\begin{aligned} \text{NumberOfSuperstore}_{it} &= \beta_{1t} + \beta_2(\text{InstructionAndFormation})_{2t} + \beta_3(\text{WorkAndLifeConditions})_{3t} \\ &+ \beta_4(\text{EconomicWellness})_{4t} + \beta_5(\text{SocialRelations})_{5t} \end{aligned}$$

Each of the variable in the main model has been investigated by using approximated variables.

**Model 1-Estimating the number of superstores using variables describing the level of Instruction and Formation.**

$$\begin{aligned} \text{NumberOfSuperstore}_{it} &= \beta_1 + \beta_2(\text{PeopleWithHighSchoolDiploma2564})_{2t} \\ &+ \beta_3(\text{GraduatedPeople3034})_{3t} + \beta_4(\text{NotInEmploymentAndTrainingPeople})_{4t} \\ &+ \beta_5(\text{LifelongLearningParticipation})_{5t} + \beta_6(\text{CulturalPartecipation})_{6t} \end{aligned}$$

**Results.** The number of superstores is positively associated with People with High School Diploma with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects. The level of Graduated People in the age 30-34 is positively associated with the number of superstores with a level of p-value equal to 0.000 in the case of panel data fixed and random effect. The association between Not in Employment and Training people and the number of superstores is positively associated with a level of p-value equal to 0.000 in panel data with fixed and random effects. The relation between Lifelong Learning Participation and Number of Superstore is negative and with a level of p-value equal to 0.000 in the panel data with fixed and random effects. The level of cultural Partecipation in negatively associated with the number of superstores with a level of p-value equal to 0.000 in the case of panel data with fixed effects.

<b>Panel Data with Fixed Effects 77 observations, 20 cross section units. Time Series: min1-max4. Dependent Variable: Number of Superstore on a set of variables representative of Instruction and Formation.</b>					
<b>Panel data fixed effects</b>	<b>People with high school diploma 2564</b>	<b>Graduate People 3034</b>	<b>Not in employment and training people</b>	<b>Lifelong Learning participation</b>	<b>Cultural participation</b>
<i>Coefficient</i>	1,07361	0,66227	2,53277	-2,50913	-117.296
<i>Standard Errors</i>	0,23235	0,17743	0,35858	0,48519	0,17665
<i>T-Value</i>	4,621	3,733	7,063	-5,171	-6,64
<i>P-Value</i>	0,0007	0,003	<0,0001	0,0003	<0,0001
	***	***	***	***	***
<i>Mean dependent variable</i>	31,25974	<i>Log-verosimiglianza</i>	1,071189	<i>Statistical Test</i>	<i>F(46,11)=12,4819</i>
<i>Residual sum of squares</i>	4,38463	<i>Akaike Criteria</i>	129,8576	<i>With p-value</i>	<i>P(F(46,11)&gt;12,4819)=4,01416e-005</i>
<i>Standard Error of the regression</i>	0,63135	<i>Criterio di Schwarz</i>	284,5488	<i>Test for difference in group intercept</i>	
<i>R quadro LSDV</i>	0,999956	<i>Hannan-Quinn</i>	191,7327	<i>Null Hypothesis</i>	
<i>R-quadro intragruppi</i>	0,981202	<i>rho</i>	-0,39154	<i>Test statistics</i>	<i>F(19,11)=432,48</i>
<i>LSDV F(65,11)</i>	3816,347	<i>Durbin Watson</i>	2,221923	<i>p-value</i>	<i>P(F(19,11)&gt;432,48)=3,92559e-013</i>
<i>P-value(F)</i>	1,18E-18	<i>Regression tests</i>			

Table 1. Panel Data with Fixed Effects.

<b>Panel Data with Fixed Effects 77 observations, 20 cross section units. Time Series: min1-max4. Dependent Variable: Number of Superstore on a set of variables representative of Instruction and Formation.</b>					
<b>Panel data random effects</b>	<b>People with high school diploma 2564</b>	<b>Graduate People 3034</b>	<b>Not in employment and training people</b>	<b>Lifelong Learning participation</b>	<b>Cultural participation</b>
<i>Coefficient</i>	1,0479	0,6512	2,4653	-2,4319	-1,1405
<i>Standard Errors</i>	0,1562	0,1193	0,2391	0,3248	0,1180
<i>T-Value</i>	0,0000	0,0000	0,0000	0,0000	0,0000
<i>P-Value</i>	0,0007	0,0030	<0,0001	0,0003	<0,0001
	***	***	***	***	***
<i>Mean dependent variable</i>	31,260	<i>Hannan-Quinn</i>	979,072	<i>Asymptotic statistical test</i>	<i>Chi-quadro(46)=12</i>
<i>Residual sum of squares dependent variable</i>	36,071	<i>Variance "between"</i>	3559,190	<i>p-value</i>	2,98E-233
<i>Residual sum of squares</i>	249729,500	<i>Variance "Within"</i>	0,057	<i>Test Breusch-Pagan</i>	
<i>Standard Error of the regression</i>	89,754	<i>theta medio</i>	0,998	<i>Null Hypothesis in the variance of the</i>	0
<i>Likelihood function</i>	-420,505	<i>corr(y,chat)^2</i>	0,041	<i>Asymptotic statistical test</i>	<i>Chi-quadro(46)=1,0</i>
<i>Akaike Criteria</i>	935,010	<i>Conjunct test on regressors</i>		<i>p-value</i>	0,298519
<i>Schwarz Criteria</i>	1045,160				

Table 2. Panel Data with Random Effects.

<b>Dependent Variable: Number of Superstore on a set of variables representative of Instruction and Formation.</b>					
<b>Panel data WLS</b>	<b>People with high school diploma 2564</b>	<b>Graduate People 3034</b>	<b>Not in employment and training people</b>	<b>Lifelong Learning participation</b>	<b>Cultural participation</b>
<i>Coefficient</i>	-0,5108	0,8264	1,4098	-3,1743	-2,0425
<i>Standard Errors</i>	1,3692	1,1785	1,9781	2,0903	0,5727
<i>T-Value</i>	-0,3731	0,7012	0,7127	-1,5190	-3,5660
<i>P-Value</i>	0,7117	0,4866	0,4816	0,1393	0,0012
					***
<i>Residual sum of squares</i>	56,597	<i>F(46,30)</i>	25,373	<i>Mean dependent variable</i>	31,25974
<i>Standard Error</i>	1,374	<i>P-Value(F)</i>	0,000	<i>Residual Sum of Squares Dependent variable</i>	3,61E+01
<i>R-squared</i>	0,975	<i>Schwarz Criteria</i>	398,971	<i>Residual Sum of Squares</i>	3770,291
<i>Corrected R-Squared</i>	0,937	<i>Hannan-Quinn</i>	332,874	<i>Square error</i>	11,21055

Table 3. Panel Data WLS.

**Model 3- Estimating the number of superstores using variables approximating the level of Working and Life conditions.**

$$\begin{aligned}
 \text{NumberOfSuperstore}_{it} = & \beta_{1t} + \beta_2(\text{EmploymentRate2064})_{2t} + \\
 & \beta_3(\text{UnemploymentRate})_{3t} + \\
 & \beta_4(\text{EmployedInShortTermContractsInTheLastFiveYears})_{4t} + \\
 & \beta_5(\text{EmployedWithLowIncome})_{5t} + \beta_6(\text{OvereducatedWorkers})_{6t} + \\
 & \beta_7(\text{NonRegularEmployed})_{7t} + \beta_8 \frac{\text{EmploymentRateOfWomanSon}}{\text{EmploymentRateOfWomanWithoutASon}} * 100_{8t} \\
 & + \beta_9(\text{InvoluntaryPartTime})_{9t} + u_{it}
 \end{aligned}$$

**Results.** Employment rate in the age 20-64 is negatively associated with the number of superstores in the panel data with fixed effects with a level of p-value equal to 0.000 in the panel data with fixed effects, with a p-value equal to 0.000 in a panel data with random effects. Unemployment rate is negatively associated with the number of superstores with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects. The number of employed having short term contrast in the last five years is negatively associated with the number of superstores with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects. Employed with low income are negatively associate with the number of superstores with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects. The level of overeducated works is negatively associated with the number of Superstore with a level of p-value equal to 0.000 in the case of panel



data with fixed and random effects. The association between non-regular employed and the number of superstores in negative and with a p-value equal to 0.000 in the case of panel data with fixed and random effects, the association between number of superstore and the percentage rapport between Employment rate of woman with son and employment rate of woman without a son is negative and with a p-value equal to 0.000 in the case of panel data with fixed and random effects. The relation between Involuntary part time and the number of superstores is negative and with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects.

**Panel Data with Fixed Effects 77 observations, 20 cross section units. Time Series: min1-max4. Dependent Variable: Number of Superstore on a set of variables representative of Working Life Conditions.**

	Employment Rate 2064	Unemployment Rate	Employed in short term contracts in the last five years	Employed with low income	Overeducated workers	Non regular employed	(Employment Rate of Woman with Son)/Employment Rate of woman without a Son*100	Involuntary Part Time
Coefficient	-2,4409	-5,1107	-1,2451	-1,4744	-1,0161	-9,3626	-0,2932	-5,4885
Standard Errors	0,4418	0,6584	0,1242	0,2516	0,2860	1,1940	0,0366	0,6867
T-Value	-5,5250	-7,7620	-10,0200	-5,8610	-3,5530	-7,8410	-8,0040	-7,9920
P-Value	0,0002	<0,0001	<0,0001	0,0001	0,0045	<0,0001	<0,0001	<0,0001
	***	***	***	***	***	***	***	***
Mean dependent Variable	31,260	P-value(F)		0,000	Durbin Watson	2,221823	Null Hypothesis	
Residual Sum of Squares	4,385	Log-Verosimiglianzi		1,071	Regression tests	Test statistics	F(19,11)=432,48	
Standard Errors of the regression	0,631	Akaike criteria		129,858	Statistical test	p-value	F(19,11)>432,48=3,92559e-013	
R-quadro LSDV	1,000	Criterio di Schwarz		284,549		F(46,11)=12,4819		
R-quadro ingragruppo	0,981	Hannan-Quinn		191,733	with p-value	P(F(46,111)>12,4819=4,0156e-005		
LSDV F(65,11)	3816,347	rho		-0,392	test for differences in group			

**Table 4. Panel Data With Fixed Effects.**

77 observations, 20 cross section units. Time Series: min1-max4. Dependent Variable: Number of Superstore on a set of variables representative of Working Life Conditions.								
	Employment Rate 2064	Unemployment Rate	Employed in short term contracts in the last five years	Employed with low income	Overeducated workers	Non regular employed	(Employment Rate of Woman with Son)/Employment Rate of woman without a Son*100	Involuntary Part Time
Coefficient	-2,41318	-4,99222	-1,22475	-1,4404	-1,00291	-9,10439	-0,293759	-5,34832
Standard Errors	0,297449	0,440199	0,0831373	0,168532	0,1922778	0,792932	0,0236863	0,456941
z	-8,113	-11,34	-14,73	-8,547	-5,202	-11,48	-11,9	11.70
P-Value	4,94E-16	8,24E-30	4,43E-49	1,27E-17	1,97E-07	1,63E-30	1,19E-32	1,21E-31
	***	***	***	***	***	***	***	***
Mean dependent Variable	31,25974	Akaike Criteria		935,0098	Theta medio	0,9979	Test Breusch Pagan	
Residual sum of squares dependent variable	36,07062	Schwarz Criteria		1045,169	corr(y,yhat)^2	0,041028	Null Hypothesis in the variance of the specified erros	0
Residual sum squares	249729,5	Hannan-Quinn		979,0724	Conjunct test on regressors		Asymptotic Statistical test	Chi-quadro(1)=1,0808
Standard error of the regression	89,75406	Variance "between"		3559,19	Asymptotic statistical test	quadro(46)=125		
Likelihood function	-420,5049	Variance "Within"		0,056943	p-value	7,47	P-value	0,298519
						2,98E-233		

**Table 5. Panel Data with Random Effects.**

#### Model 4-Estimating the number of superstores using variables approximating the level of economic wellness.

*NumberOfSuperstore<sub>it</sub>*

$$= \beta_{1t} + \beta_2(\text{MeanGdPerCapita})_{2t} + \beta_3(\text{PovertyRisk})_{3t} \\ + \beta_4(\text{SeriousMaterialDeprivation})_{4t} + \beta_5(\text{SeriousEconomicDifficoultu})_{5t} \\ + \beta_6(\text{LowWorkingIntensity})_{6t} + u_{it}$$

**Results.** The level of the number of superstores is negatively associated with the level of Mean GPD per capita with a level of p-value equal to 0.000 in the case of panel data with fixed effects, with a level of p-value equal to 0.000 in the case of panel data with random effects. Poverty risk is negatively associated with the level of the number of superstores with a level of p-value equal to 0.000 in the case of panel data with fixed effects, and with a level of p-value equal to 0.000 in the case of panel data with random effects. The association between Serious material deprivation and the number of superstores is positive and with a level of p-value equal to 0.000 in the case of panel data with fixed effects, with a level of p-value equal to 0.000 in the case of pane data with random effects. The association between Serious Economic Difficulty and the number of superstores is negative with a level of p-value equal to 0.000 in the case of panel data with fixed effects, with a level of p-value equal to 0.000 in the case of panel data with random effects. The association between low Working Intensity and the Number of Superstore is positive with a level of p-value equal to 0.000 in the case of panel data with fixed effects, and with a level of p-value equal to 0.000 in the case of panel data with random effects.

<i>Panel Data with Fixed Effects 77 observations, 20 cross section units. Time Series: min1-max4. Dependent Variable: Number of Superstore on a set of variables representative of Economic Wellness</i>					
<b>Fixed Effects</b>	<b>Mean Gdp per capita</b>	<b>Poverty Risk</b>	<b>Serious material Deprivation</b>	<b>Serious Economic Difficultu</b>	<b>Low Working Intensity</b>
<i>Coefficient</i>	-0,00806077	-0,361917	0,73541	-0,693674	1,07006
<i>Standard Errors</i>	0,00112	0,11115	0,08013	0,10935	0,17267
<i>T-value</i>	-7,215	-3,256	9,178	-6,344	6,197
<i>P-Value</i>	<0,0001	0,0077	<0,0001	<0,0001	<0,0001
	***	***	***	***	***
<i>Mean dependent Variable</i>	31,25974	<i>LSDV F(65,11)</i>	3816,347	<i>rho</i>	-0,391540
<i>Residual sum of squares</i>	4,38463	<i>P-value(F)</i>	1,18E-18	<i>Durbin Watson</i>	2,221823
<i>Standard error of the regression</i>	0,63135	<i>Log-Verosimiglianza</i>	1,071189	<i>Regression test</i>	
<i>R-quadro LSDV</i>	0,999956	<i>Criterio di Schwarz</i>	284,5488	<i>With p-value</i>	<i>P(F(46,11)&gt;12,4189)=4,01516e-005</i>
		<i>Hannan-Quinn</i>	191,7327	<i>Test for differences in group</i>	

Table 6. Panel Data with Fixed Effects.

<i>Panel Data With Random Effects. 77 observations, 20 cross section units.</i>					
<i>Time series: min1-max4. Dependent variable: numero of superstore on a set of variables representative of Economic Wellness</i>					
<i>Random Effects</i>	<i>Mean Gdp per capita</i>	<i>Poverty Risk</i>	<i>Serious material deprivation</i>	<i>Serious Economic Difficoulty</i>	<i>Low Working intensity</i>
<i>Coefficients</i>	-0,00788071	-0,352032	0,722886	-0,673289	1,04461
<i>Standard Error</i>	0,00075	0,07461	0,053617	0,0729	0,11539
<i>z</i>	-10,53	-4,718	13,48	-9,236	9,053
<i>p-value</i>	6,14E-26	2,38E-06	1,98E-41	2,57E-20	1,39E-19
	***	***	***	***	***
<i>Mean dependent variable</i>	31,25974	<i>Variance between</i>	3559,19	Null Hypothesisi in the variance of the specified errors	0
<i>Residual Sum of squares dependent variable</i>	36,07062	<i>Variance "within"</i>	0,056943	Asymptotic statistical test	Chi-quadro(1)=1,0808
<i>Residual sum of squares</i>	249729,5	<i>thetat medio</i>	0,9979	p-value	0,298519
<i>Standard Error of the regression</i>	89,75406	<i>corr(y, yhat)^2</i>	0,041028		
<i>Likelihood funcion</i>	-420,5049	<i>Conjunct test on regressors</i>			
<i>Akaike Criteria</i>	935,0098	<i>asymptotic statistical test</i>	Chi-quadro(46)=1257,47		
<i>Schwars Criteria</i>	1045,169	<i>p-value</i>	2,98E-233		

Table 7. Panel Data with Random Effects.

### **Model 5- Estimating the number of superstores using variables approximating social relations.**

#### **NumberOfSuperstore<sub>it</sub>**

$$= \beta_{1t} + \beta_2(\text{SatisfactionForFamilyRelations})_{2t} + \beta_3(\text{SocialPartecipation})_{3t} + \beta_4(\text{Volunteering})_{4t} + \beta_5(\text{AssociationFinancing})_{5t} + \beta_6(\text{GeneralizedTrust})_{6t}$$

**Results.** The association between Satisfaction for family relations and Number of Superstore is positive with a level of p-value equal to 0.000 in the case of panel data with fixed effects, and equal to 0.000 in the case of panel data with random effects. The association between social Partecipation and Number of Superstore is negative and with a level of p-value equal to 0.000 in the case of panel data with fixed effects, equal to 0.000 in the case of panel data with random effects. The association between volunteering and the Number of Superstore is positive with a level of p-value equal to 0.000 in the case of panel data with fixed and random effects. The relation between association financing and the number of superstores is negative and equal to 0.000 in the case of panel data with fixed effects, in the case of panel data with random effects. The association between Generalized Trust and the number of superstores is negative and with a p-value equal to 0.000 in the case of panel data with random effects and with fixed effects.

Panel data with fixed effects. 77 observations, 20 cross section units. Time series: min1-max3. Dependent variable: Number of Superstores on a set of variables representative of Social Relations					
Fixed Effects	Satisfaction for Family relations	Social participation	Volunteering	Association financing	Generalized trust
Coefficients	1,67812	-2,40046	2,70478	-2,02461	-0,665937
Standard errors	0,1569	0,31612	0,48662	0,27437	0,11762
t-value	10,7	-7,593	5,558	-7,379	-5,662
P-value	<0,0001 ***	<0,0001 ***	0,0002 ***	<0,0001 ***	<0,0001 ***
Mean dependent variable	31,25974	LSDV F(65,11)	3816,347	$\rho$	-0,39154
Residual sum of squares	4,38463	P-value	1,18E-18	Durbin-Watson	2,221823
Standard Error of the regression	0,63135	Log-verosimiglianza	1,071189	Regression tests	
R-quadro LSDV	0,999956	Akaike Criteria	129,8576	Statistical test	F(46,11)=12,4819
R-quadro intragruppi	0,981202	Criterio di Schwarz	284,5488	with p-value	P(F(46,11)>12,4819)=4,01516e-005
		Hannan-Quinn	191,7327	Test for differences in group intercept	

Table 8. Panel Data with Fixed Effects.

Panel data with random effect 77 observation, 20 cross section units. Time series min1-max4, dependent variable: number of superstore on a set of variables representative of Social Relations					
Random effects	Satisfactory for family relations	Social participation	Volunteering	Association Financing	Generalized Trust
Coefficients	1,65141	2,34563	2,60716	-1,97384	-0,648208
Standard Errors	0,10483	0,21127	0,3242	0,18313	0,07879
T-Value	15,75	-11,1	8,042	-10,78	-8,227
P-Value	6,54E-56 ***	1,22E-28 ***	8,86e-0,16 ***	4,34E-27 ***	1,93E-16 ***
Mean Dependent Variable	31,25974	Variance between	3559,19	Null Hypothesis in the variance of the specified errors	0
Residual sum of squares dependent variable	36,07062	Variance "within"	0,056943	Asymptotic statistical test	Chi-quadro(1)=1,0808
Residual sum of squares	249729,5	theta medio	0,9979	p-value	0,298519
Standard Error of the regression	89,75406	corr(y,yhat)^2	0,041028		
Likelihood function	-42,5049	Conjunct test on regressors			
Akaike Criteria	935,0098	Asymptotic statistical test	Chi-quadro(46)=1257,47		

Table 9. Panel Data with Random Effects.

## INDEX OF FIGURES

Figure 1. The relations among superstore location, business location, organizational location and economic geography.

Figure 2. Workflow for clusterization process in KNIME.

Figure 3. Workflow used in KNIME to generate ranked clusters based on excel reading.

Figure 4. A comparison among clusterized between surface cluster and clusters of purchasing group. Clusterization is realized in KNIME.

Figure 5. KNIME based clusterization. Source of data: Italian Minister of Economic Development and Italian National Institute of Statistics. Period of analysis: 2007-2016.

Figure 6. Relation among clusterized values affecting italian regions during the period 2007-2016. Data Source Italian Minister of Economic Development and Italian National Institute of Statistics.

Figure 7. Correlation matrix based on clusterized data. Source: Italian Minister of Economic Development and Italian National Institute of Statistics. Data shows the presence of positive relations among variables. Clusterization process has been realized in a period between 2007 and 2016.

Figure 8. Flow chart able to describe the analytical process started with clusterization in KMINE, through correlation matrix, arriving to econometric analysis

## INDEX OF TABLES

- Table 1. Panel Data With Fixed Effects.
- Table 2. Panel Data With Random Effects.
- Table 3. Panel Data WLS.
- Table 4. Panel Data With Fixed Effects.
- Table 5. Panel Data With Random Effects.
- Table 6. Panel Data With Fixed Effects.
- Table 7. Panel Data With Random Effects.
- Table 8. Panel Data With Fixed Effects.
- Table 9. Panel Data With Random Effects.

## BIBLIOGRAPHY

- Bartik, T. J., 1988. The effects of environmental regulation on business location in the United States. *Growth and Change*, 19(3), pp. 22-44. <https://doi.org/10.1111/j.1468-2257.1988.tb00473.x>
- Chavda, R. K., 2004. An analysis of the salient factors in business location decision making. *Academy of Management Proceedings*, Volume 1. <https://doi.org/10.5465/ambpp.2004.13857313>
- Coca-Stefaniak, J. A., Parker, C. & Rees, P., 2010. Localisation as a marketing strategy for small retailers. *International Journal of Retail & Distribution Management*.
- Development-MISE, I. M. o. E., t.thn. *Osservatorio Commercio*. [Online]  
Available at:  
<http://osservatoriocommercio.sviluppoeconomico.gov.it/DEFINIZIONI%20STATISTICHEattuali.pdf>  
[Diakses 23 07 2018].
- Dubé, J., Brunelle, C. & Legros, D., 2016. Location Theories and Business Location Decision: A Micro-Spatial Investigation in Canada. *The Review of Regional Studies*, 46(2), pp. 143-170.
- Elouni, J., Ltifi, H. & Ayed, M. B., 2016. Knowledge visualization model for intelligent dynamic decision-making. *International Conference on Hybrid Intelligent Systems*, pp. 223-235. [https://doi.org/10.1007/978-3-319-27221-4\\_19](https://doi.org/10.1007/978-3-319-27221-4_19)
- Felice, E., 2011. Regional value added in Italy, 1891–2001, and the foundation of a long-term picture. *The Economic History Review*, 64(3), pp. 929-950. <https://doi.org/10.1111/j.1468-0289.2010.00568.x>
- Galiano, A., Leogrande, A. & Massaro, A., 2018. Le determinanti socio-economiche del franchising in Europa nel 2007-2016. *Economia e diritto del terziario-Open Access*, Volume 3. <https://doi.org/10.3280/edt3-2017oa6266>
- Gramsci, A., 2008. La questione meridionale. *La Riflessione*.
- Grimaldi, D., Fernandez, V. & Carrasco, C., 2019. Heuristic for the localization of new shops based on business and social criteria. *Technological Forecasting and Social Change*, Volume 142, pp. 249-257. <https://doi.org/10.1016/j.techfore.2018.07.034>
- Istat, t.thn. *Rapporto BES*. [Online]  
Available at: <https://www.istat.it/it/archivio/207259>  
[Diakses 28 07 2018].

- Jacob Trip, J. A. N., 2007. The role of urban quality in the planning of international business locations: the case of Amsterdam Zuidas. *Journal of Urban Design*, 12(2), pp. 275-293.  
<https://doi.org/10.1080/13574800701306385>
- Jensen, N. M., Malesky, E. J. & Walsh, M., 2015. Competing for global capital or local voters? The politics of business location incentives. *Public Choice*, 164(3-4), pp. 331-356. <https://doi.org/10.1007/s11127-015-0281-8>
- Lin, J. et al., 2016, July. Where is the goldmine?: finding promising business locations through Facebook data analytics. *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, pp. 93-102.  
<https://doi.org/10.1145/2914586.2914588>
- MISE, 2019. *Osservatorio Nazionale del Commercio*. [Online]  
Available at: <http://osservatoriocommercio.sviluppoeconomico.gov.it/>
- Mody, A. & Srinivasan, K., 1998. Agglomeration, regional grants and firm location. *Institute for Fiscal Studies*.
- Molcho, G. & Schneor, R., 2015. MIND—semantic based knowledge visualization. *Procedia CIRP*, pp. 89-94. <https://doi.org/10.1016/j.procir.2015.03.011>
- Pajones, M., Steiner, T. & Hackner-Jaklin, N., 2017. Development of a location-factor-matrix for sustainable business locations. *WIT Transactions on Ecology and the Environment*, Volume 210, pp. 407-415.  
<https://doi.org/10.2495/SDP160341>
- Poynor, M. W., 1984. Getting the shop in the right place: the practical application of locational techniques. *Retail and Distribution Management*, 12(4), pp. 20-24. <https://doi.org/10.1108/eb018236>
- Rigby, D. K. & Vishwanath, V., 2006. Localization--the revolution in consumer markets. *Harvard business review*, 84(4), pp. 82-92.
- Rohlin, S., Rosenthal, S. S. & Ross, A., 2014. Tax avoidance and business location in a state border model. *Journal of Urban Economics*, Volume 83, pp. 34-49. <https://doi.org/10.1016/j.jue.2014.06.003>
- Romadona, I., Azizatunnishak, K. & Monica, A. A., 2017. Disparity in Determining Business Location: A Case Study in Unnes Sekaran Area. *Advanced Science Letters*, 23(8), pp. 7170-7172.  
<https://doi.org/10.1166/asl.2017.9320>
- Shenghao, Z., 2009. The Success and Failure of Multinational Retail Cooperations' Localization Strategy—A Comparative Study on Carrefour's and Tesco's Operating in Korean Market.. *Finance & Economics*, , p. 1..
- Somekh, B., 2012-2013. Commuting And Shopping: Determinants Of City Income Structure. *Working Papers WP2012/3, University of Haifa, Department of Economics*.
- Wagstaff, K., Cardie, C., Rogers, S. & Schrödl, S., 2001. Constrained k-means clustering with background knowledge. *Icml*, pp. 577-584.
- Wood, S., Coe, N. M. & Wrigley, N., 2016. Multi-scalar localization and capability transference: exploring embeddedness in the Asian retail expansion of Tesco.. *Regional Studies*, , , 50(3), pp. 475-495..  
<https://doi.org/10.1080/00343404.2014.926317>
- Xu, Y., Shen, Y., Zhu, Y. & Yu, J., 2018, May. Fusing Satellite Data and Urban Data for Business Location Selection: A Neural Approach. *International Conference on Database Systems for Advanced Applications*, Issue Springer, Cham., pp. 589-605. [https://doi.org/10.1007/978-3-319-91458-9\\_36](https://doi.org/10.1007/978-3-319-91458-9_36)
- Zimon, G., 2017. Procurement logistics in branch purchasing group.. *Research in Logistics & Production*, p. 7. <https://doi.org/10.21008/j.2083-4950.2017.7.2.9>

Zimon, G., 2019. Branch Group Purchasing Organizations vs. Sales Profitability of Commercial Companies. *Eurasian Economic Perspectives* , pp. 263-275. [https://doi.org/10.1007/978-3-030-11833-4\\_17](https://doi.org/10.1007/978-3-030-11833-4_17)

Zünd, D., Woodbury, R. & Schmit, G., 2016. Meso-scale modeling of residential and business locations. *Simulation*, 92(3), pp. 295-306. <https://doi.org/10.1177/0037549716629223>