Innovative SMEs Collaborating with Others in Europe

Leogrande, Angelo and Costantiello, Alberto and Laureti, Lucio and Matarrese, Marco Maria

Lum University Giuseppe Degennaro, Lum University Giuseppe Degennaro, Lum University Giuseppe Degennaro, Lum University Giuseppe Degennaro

9 May 2022

Angelo Leogrande[1], Alberto Costantiello[2], Lucio Laureti[3], Marco Maria Matarrese[4]

Innovative SMEs Collaborating with Others in Europe

Abstract

The following article investigates the determinants that lead innovative SMEs to collaborate. Data from 36 European countries is analyzed using Panel Data with Fixed Effects, Panel Data with Random Effects, Pooled OLS, WLS and Dynamic Panel models. The analysis shows that the ability of innovative SMEs to collaborate is positively associated with the following variables: "*Linkages*", "*Share High and Medium high-tech manufacturing*", "*Finance and Support*", "*Broadband Penetration*", "*Non-R&D Innovation Expenditure*" and negatively to the following variables: "*New Doctorate graduates*", "*Venture Capital*", "*Foreign Controlled Enterprises Share of Value Added*", "*Public-Private Co-Publications*", "*Population Size*", "*Private co-funding of Public R&D expenditures*". A clustering with k-Means algorithm optimized by the Silhouette coefficient was then performed and four clusters were found. A network analysis was then carried out and the result shows the presence of three composite structures of links between some European countries. Furthermore, a comparison was made between eight different predictive machine learning algorithms and the result shows that the Random Forest Regression algorithm performs better and predicts a reduction in the ability of innovative SMEs to collaborate equal to an average of 4.4%. Later a further comparison is made with augmented data. The results confirm that the best predictive algorithm is Random Forest Regression, the statistical errors of the prediction decrease on average by 73.5%, and the ability of innovative SMEs to collaborate is predicted to growth by 9.2%.

Keywords: Innovation, and Invention: Processes and Incentives; Management of Technological Innovation and R&D; Diffusion Processes; Open Innovation

JEL Classification: O30; O31, O32; O33; O36

1. Introduction-Research Question

This article analyzes the issue of collaboration between innovative small and medium-sized enterprises in Europe. The issue is relevant for its impact on technological innovation and overall investment in research and development in European national systems. In fact, if on the one hand it is certain that technological innovation has a positive effect in terms of gross domestic product, it is also true that this innovation must be carried out by companies. And this is where the question arises. In fact, most businesses in Europe are small and medium-sized. The size of European SMEs does not allow them to create company departments specialized in product innovation, process innovation, digital transformation, nor does it allow them to create real departments for research and development. And yet it is necessary to find methodologies to ensure that SMEs collaborate in innovation as their contribution is necessary to raise the degree of attractiveness of innovation systems

[1] Assistant Professor of Economics at LUM University Giuseppe Degennaro, researcher at LUM Enterprise s.r.l., Casamassima, Puglia, Italia, European Union, email: leogrande.cultore@lum.it

[2] Professor of Economics at LUM University Giuseppe Degennaro, Casamassima, Puglia, Italia, European Union, email: costantiello@lum.it

[3] Professor of Economics at LUM University Giuseppe Degennaro, Casamassima, Puglia, Italia, European Union, email: laureti@lum.it

[4] Assistant Professor of Economics at LUM University Giuseppe Degennaro, Casamassima, Puglia, Italia, European Union, e-mail: matarrese.phdstudent@lum.it

at national level. It follows therefore that it is necessary to find methodologies to ensure that SMEs systematize their resources to increase the degree of technological innovation. One of the methodologies to ensure that European SMEs are more collaborative consists precisely in the analysis of the indicator that has been chosen in the following discussion: "Innovative SMEs Collaborating with Others in Europe".

There are various models in the literature that have been proposed to allow SMEs to collaborate in technological innovation, such as models of open innovation and models based on the idea of external knowledge. While it is certainly possible to find a dimension of positive externality in knowledge, it is also necessary that this dimension is not random or left to chance, but rather it is contractualised in the productive relations between SMEs. Furthermore, it should be considered that in general the markets where many SMEs operate are characterized by a very marked competitive orientation and this fact prevents companies from investing in research and development and technological innovation. As a result, investment in technological innovation and research and development is often relegated to companies operating under an oligopoly and monopoly regime.

Finally, there is an international condition that makes it necessary to unlock the contribution to technological innovation and research and development that can be produced through small and medium-sized enterprises. Indeed, in the context of the technological conflict between China and the US, the role of Europe can only be better defined by allowing SMEs to fully unleash their innovative potential. In fact, Europe lags both the US and China in the sense of technological innovation and the big companies alone are not enough to close the gap in terms of knowledge production, patents, and intangible assets. Thus arises the need for Europe to create networks between innovative SMEs to ensure that investment in R&D increases globally, and to develop in SMEs an aptitude for both collaboration and innovation capable of generating profits. relevant outputs both from the point of view of the market and from the point of view of the production of social capital.

The article continues as indicated below: the second paragraph contains a brief analysis of the reference literature, the third paragraph illustrates the econometric model, the fourth paragraph contains the clustering analysis with the k-Means algorithm, the fifth paragraph presents the network analysis model applied to European countries, the sixth paragraph shows the results in terms of machine learning for the prediction with original data, the seventh paragraph refers to the prediction with machine learning algorithms applied to the augmented data, the eighth paragraph concludes , the ninth paragraph contains the bibliography, the tenth paragraph represents in the form of an appendix the metric results obtained with the research methods used.

2. Literature Review

An analysis of the literature that refers to the role of collaboration in the sense of technological innovation among European SMEs is briefly presented below. It should be considered that the role of technological innovation has been recognized within economic science already starting from the classical theory of economic growth (Solow, 1956). Furthermore, the role of technological innovation, research and development and the knowledge economy in general has also been recognized in the approach of the theory of endogenous growth (Romer, 1994). Furthermore, the role of technological innovation, also in the distinction between the very idea of innovation and invention, was developed by the Austrian economist Joseph Alois Schumpeter (Schumpeter, 1934). Furthermore, the issue of innovation also significantly affects the issue of legislation due to the controversial role that patents, and industrial property has in allowing or limiting the dissemination of scientific and technological discoveries (Boldrin & Levine, 2002). In this general theoretical context, the scientific literature must be inserted which deals more specifically with the role of collaborative SMEs in the sense of technological innovation.

(Benhayoun, et al., 2020) consider small and medium-sized enterprises that actively participate in collaborative innovation networks. The ability of the individual companies participating in these networks to innovate and acquire the information necessary for growth essentially depends on the specific condition in which the company finds itself, its competitive context, and the ability to translate the innovations acquired within the corporate organizational structure through collaboration. (Ioanid, et al., 2018) analyze the role of social networks in determining the innovative capacity of small and medium-sized enterprises in Romania. The results of the analysis show that Romanian companies can derive benefits from social networks. However, these advantages are relegated to marketing and market positioning aspects. Romanian companies have little ability to use social networks in the sense of technological innovation.

(Jespersen, et al., 2018) consider the reasons that push SMEs to collaborate in the direction of innovation using proximity. In fact, companies looking for partners in innovation try to find companies that are similar. The authors analyze four different types of proximity, namely: geographical, cognitive, organizational, and social. In general, the partners of technological innovation are chosen based on geographical proximity. However, companies change their choice criteria also based on the technological and organizational complexity of the collaboration in the sense of innovation. (Akinwale, 2018) considers the positive relationship between open innovation and the financial performance of companies operating in the oil and gas sector in Nigeria. (Zahoor & Al-Tabbaa, 2020) presents an analysis of the literature on the impact of inter-organizational collaboration on the ability of SMEs to innovate. (De Noni, et al., 2018) analyze the role of collaboration between geographical areas with different levels of development in terms of technological innovation in Europe. The results of the panel analysis show that regions that are more technologically backward gain significant advantages from collaborating with regions that are more technologically advanced. (Jasimuddin & Naqshbandi, 2019) they analyze the role of "Knowledge infrastructure Capability" in determining the conditions of efficiency of open innovation they analyze the case of 125 companies operating in France. The data show that the presence of "Knowledge Infrastructure Capability" increases the ability of companies to operate through open innovation models. (Del Vecchio, et al., 2018) investigate the relationship between big data and open innovation in small and medium-sized enterprises and in big corporations. (Radziwon & Bogers, 2019) they consider the complex system of relationships that binds a small and medium-sized enterprise to the external environment in the context of open innovation. In fact, on the one hand, SMEs are not able to complete the innovation processes autonomously, on the other hand, they may have difficulties in aligning themselves with the objectives of open innovation systems. (Grimsdottir & Edvardsson, 2018) consider the case of two companies that practice open innovation models in Iceland. The two companies considered have diversified approaches to open innovation due to their characteristics in terms of technological advancement. The authors tend to demonstrate that while technologically advanced firms prefer inside-out open innovation models, less technologically advanced firms prefer outside-in models. (Saastamoinen, et al., 2018) analyze the positive relationship between the ability of small and medium-sized enterprises to create forms of inter-organizational collaboration and their ability to be more innovative. (Grimaldi, et al., 2021) They analyze the relationship between the ability of small and medium-sized enterprises to collaborate in open innovation and the protection of intellectual property rights. The findings show that firms that have a collaborative intellectual property rights strategy perform better than firms that have a defensive strategy. (Zaridis, et al., 2021) certify the existence of a positive relationship between collaboration in the sense of innovation and the results in terms of business performance in a case study of 504 companies operating in the agricultural sector. (D'Angelo & Baroncelli, 2020) apply a Tobit / Probit model to the analysis of 2591 Italian small and medium-sized enterprises to investigate the existence of a relationship between open innovation and performance. The results show that companies that can create forms of horizontal collaboration in the sense of research and development produce greater outputs in terms of innovation. Collaboration between business and university enhances product innovation. (Anderson & Hardwick, 2017) propose a sociological analysis of the relationship between open innovation and knowledge management in small and medium-sized enterprises, underlining the role of trust in creating the conditions for collaboration between enterprises in the sense of innovation. (Ueasangkomsate & Jangkot, 2019) empirically verify the positive effect of collaboration between small and medium-sized enterprises and the industrial sector in increasing the performance in the sense of innovation in Thailand. The ability to produce collaboration between small and medium-sized enterprises is therefore an essential element for increasing innovation at the national level. To this end, it is also necessary to consider the relationships that technological innovation and collaboration among SMEs have in connection with the following factors:

- The general level of innovation among countries (Leogrande, et al., 2022);
- The ability of firm to improve sales (Costantiello, et al., 2022);
- The capability to innovate at firm level (Costantiello, et al., 2022);
- The production of creative intangibles for industrial purposes (Laureti, et al., 2022);
- The broadband penetration (Leogrande, et al., 2021);
- The role of venture capital in financing corporate innovation (Leogrande, et al., 2021);
- The relevance of human capital (Leogrande & Costantiello, 2021);
- The presence of foreign doctorate students (Laureti, et al., 2022);
- ICT training (Laureti, et al., 2022).

As is evident, the possibility of implementing profitable technological innovation models in small and medium-sized enterprises necessarily requires forms of collaboration. However, the ability of small and medium-sized enterprises to collaborate is also connected to the presence of a set of factors concerning human capital, the state of digitization at the country level, and the presence of an orientation that tends to make the construction of social goods prevail. compared to the predatory tendencies of individual companies.

3. The Econometric Model

We have estimated the following econometric model:

$$\begin{aligned}
\boldsymbol{InnovativeSMEsCollaborating_{it}} &= \boldsymbol{a_1 + b_1(Linkages)_{it}} \\
&+ \boldsymbol{b_2(EmploymentShareHighAndMediumHighTechManufacturing)_{it}} \\
&+ \boldsymbol{b_3(FinanceAndSupport)_{it} + b_4(BroadbandPenetration)_{it}} \\
&+ \boldsymbol{b_5(NonRDInnovationExpenditure)_{it}} \\
&+ \boldsymbol{b_6(NewDoctorateGraduates)_{it} + b_7(VentureCapital)_{it}} \\
&+ \boldsymbol{b_8(ForeingControlledEnterprisesShareOfValueAdded)_{it}} \\
&+ \boldsymbol{b_9(PublicPrivateCoPublications)_{it}} \\
&+ \boldsymbol{b_{10}(PrivateCoFundingOfPublicRDExpenditures)_{it}}
\end{aligned}$$

Where $i = 36$ [5]and $t = [2010; 2019]$

We find that the Innovative SMEs Collaborating are positive associated with the following variables:

[5] Countries are: Austria, Belgium, Bulgaria, Croazia, Cyprus, Czechia, Denmark, Estonia, Finlandia, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, Norway, Poland, Portogallo, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, UK.

- *Broadband Penetration:* is the percentage of the number of companies with a maximum speed of at least 100 mb / s out of the total number of companies. This indicator is very relevant as the possibility for the European Union countries to activate the benefits deriving from technological innovation and digitization depend above all on the possibility of being connected to the internet. The internet connection, especially if fast, allows you to activate e-commerce and to develop services connected to the digital economy both in public administrations and in private companies. There is therefore a positive relationship between the presence of companies that have fast internet and the presence of small and medium-sized companies that can cooperate. This positive relationship derives from the fact that a very significant part of the collaboration between companies takes place precisely using the Internet and information technologies. In fact, the collaboration between companies that takes place on the issues of technological innovation requires as a minimum requirement the presence of an Internet network infrastructure that can support the technological innovations implemented in collaboration between companies.

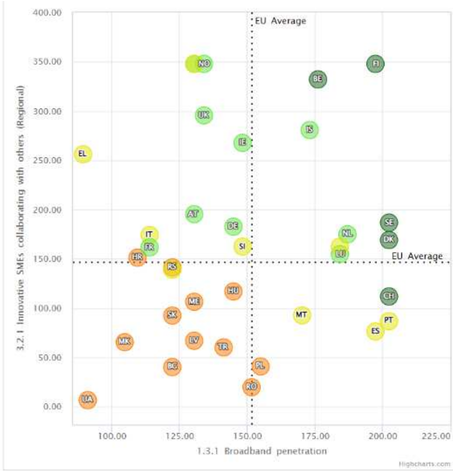


Figure 1. Relationship between Broadband Penetration and Innovative SMEs Collaborating. Source: European Innovation Scoreboard.

- *Finance and Support:* is a variable made up of three sub-variables, namely: "*R&D expenditures public sector", "Venture capital expenditures*", "*Direct government funding and government tax support for business R&D*". Finance and support therefore represent exactly a set of financial resources that can be invested to guarantee support for research and development from both the private and public sectors. It should be considered that there is a positive relationship between the value of the ability of small and medium-sized enterprises to collaborate and the financial support offered to enterprises for research and development. This relationship is also positive since a large part of the funding, especially public, which is assigned to small and medium-sized enterprises for research and development requires enterprises to collaborate through the construction of networks, consortia, and associations among enterprises. It therefore follows that the companies that manage to collaborate through the creation of cooperation contracts aimed at technological innovation and research and development are also able to obtain more financial resources to support the processes of economic growth and an improvement in the positioning of market.

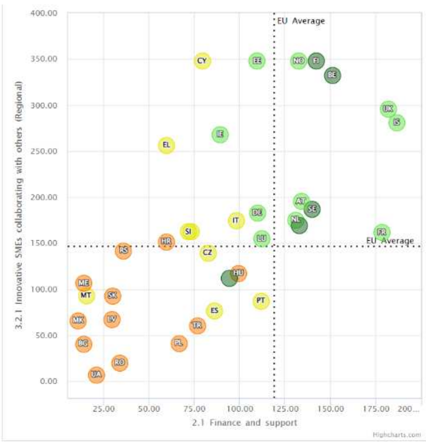


Figure 2. Relationship between Innovative SMEs collaborating with others and Finance and Support. Source: European Innovation Scoreboard.

- *Linkages:* is a variable consisting of three sub-variables namely "*Innovative SMEs collaborating with Others", "Public Private Co-Publications for Million Population*" and "*Job to Job Mobility of Human Resources in Science and Technology*". The presence of linkages is therefore positively associated with the ability of small and medium-sized enterprises to collaborate. Companies that have greater ability to create relationships are also able to generate collaborations that can translate into new products, new services and increase both customers and turnover. It therefore follows that the possibility of creating a context of collaboration between companies in the context of national innovation systems depends on the ability to offer incentives to companies to create relationships that can be transformed into new products and services. In this sense, the presence of forms of systemic collaboration in the context of open innovation, the dissemination of management models of external innovation, and also the reference to models of the "Fab Lab" type may be necessary to develop cooperation between SMEs capable of generating economic value.

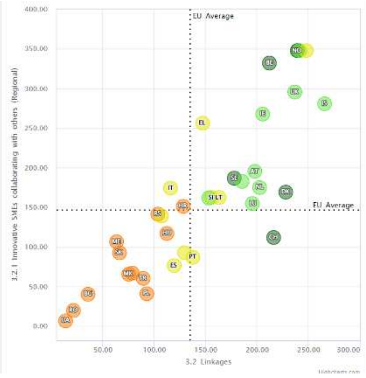


Figure 3. Relationship between Innovative SMEs Collaborating and Linkages. Source: European Innovation Linkages.

- *Non R&D Innovation Expenditure:* is a variable that considers the sum of the total expenditure for innovation with the exclusion of both intramural and extra-mural investments in research and development as a percentage of the total turnover value for all companies. This item also includes investments in equipment, machinery, and intangible assets such as patents and licenses. There is therefore a positive relationship between the value of investment in innovation net of R&D expenditure and the ability of small and medium-sized enterprises to collaborate in technological innovation. Positive relationship that can be understood considering that collaborating companies in general must also invest financial resources in the tangible and intangible infrastructures that allow the company to undertake collaborations. In fact, in the context of the knowledge economy it is necessary that collaboration takes place between companies that have levels of knowledge capital and the possibility of using intangible and material resources that are shared with the aim of producing new services and new products.

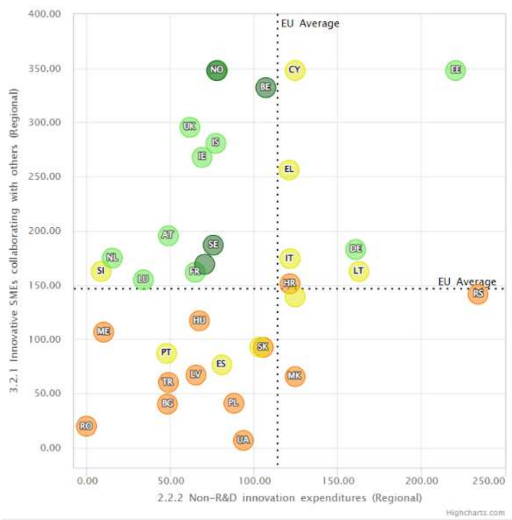


Figure 4. Relationship between Innovative SMEs collaborating with others and non-R&D innovation expenditures. Source: European Innovation Scoreboard.

- *Employment Share High and Medium High-Tech Manufacturing:* is the percentage of employees in the medium and high technology manufacturing sectors. There is therefore a positive relationship between the ability of small and medium-sized enterprises to collaborate and the value of employees in the medium and high-tech manufacturing sectors. This relationship derives from the fact that the presence of qualified human capital in enterprises, especially manufacturing, increases the ability to create the conditions for the construction of collaboration systems between small and medium-sized enterprises. Human capital is therefore necessary to create collaboration structures between companies. In particular, the fact that these operators are present within the manufacturing sector means that the manufacturing sector is more attentive both to the dimension of technological innovation and to the dimension of collaboration between companies. In fact, although technological innovation is transversal between the manufacturing sector and the service sector, it appears that the structure of manufacturing-industrial enterprises is more sensitive to technological innovation and research and development thanks also to the size of human capital.

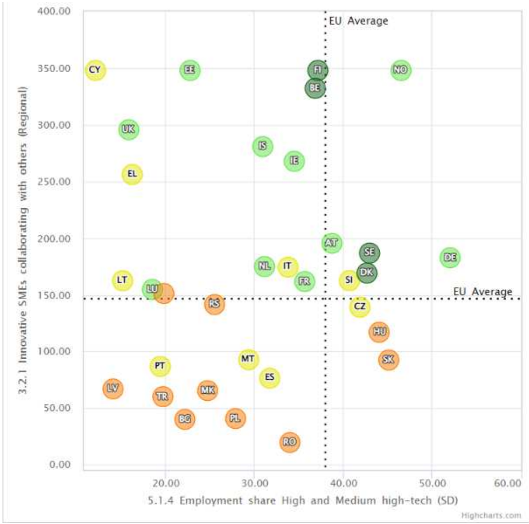


Figure 5. Relationship between innovative SMEs collaborating with others and Employment share in High and Medium High Tech. EIS: European Innovation Scoreboard.

We also find that the level of innovative SMEs Collaborating is negatively associated with the following variables:

- *Venture Capital:* It is an indicator that considers the risk capital expenditures in companies. The investment in risk capital represents a tool to support companies that face significant risks to introduce significant technological innovations into the market. It should be considered that there is a negative relationship between the presence of venture capital and the value of innovative SMEs collaborating with others. This negative relationship is since venture capitalism is generally a very aggressive structure towards the ownership of innovative SME companies and that therefore companies are oriented to seek their top-down market positioning regardless of the horizontalistic collaborations that may be generated because of collaboration between companies. In fact, venture capital tends to ensure that the company develops products, services, patents in an exclusive, private way, with the aim of selling the outputs and not sharing them with third-party companies that are considered as threatening and potential competitors.
- *Foreign Controlled Enterprises Share of Value Added:* considers the added value of companies that are under the control of a foreign parent. There is a negative relationship between the ability of small and medium-sized enterprises to collaborate in technological innovation and the percentage of enterprises that are controlled by a foreign holding. This relationship may be since a foreign-controlled company may have little aptitude for investing in collaboration with local small and medium-sized enterprises and may have little interest in creating a structure of innovation and widespread knowledge among companies in a certain geographical area or of a certain sector. Furthermore, it must be considered that companies that are acquired by foreign counterparties tend to grow significantly in the market and it is not certain that they will remain in the category of small and medium-sized enterprises. In fact, the companies that operate in a certain country being controlled by foreign companies have the possibility of accessing economies of scale and international supply chains.

- *New Doctorate Graduates:* is an indicator that considers the presence of doctorates in STEM disciplines per 1000 inhabitants between the ages of 25 and 34 out of the total number of graduates in STEM disciplines. Graduates in STEM disciplines generally have a very high capacity to actively participate in the development of technological innovation in enterprises and public institutions. However, the analysis shows that there is a negative relationship between the value of STEM PhDs and the ability of small and medium-sized enterprises to collaborate in technological innovation. This negative relationship can be better understood considering that companies that hire STEM doctorates generally can produce innovative products and services internally by building research and development departments that can be autonomous with respect to the collaboration of others.
- *Population Size:* represents the resident population in a certain country. It is necessary to consider that there is a negative relationship between the value of the resident population and the value of companies able to collaborate in technological innovation. This condition indicates that with the growth of the population the number of companies that are willing to collaborate decreases. To better understand this relationship, it is necessary to consider that most of the countries where companies have a high capacity for innovation are sparsely populated countries.
- *Private Co-Funding of Public R&D Expenditures:* represents the aggregate of R&D expenditure in the public sector and in the education sector which is financed by the private sector. It is an indicator that measures the cooperation between the public and the public. Private investment in the research and development of the university system should increase the orientation of technological innovation in the short term to serve the interests of the industry. There is a negative relationship between the value of private investment in public institutions and the ability of small and medium-sized enterprises to collaborate in technological innovation. This negative relationship is since private financing of public research and development does not necessarily involve collaboration between small and medium-sized enterprises. In fact, this financing is generally carried out by a single company that seeks a relationship with the university system to develop its own products or services without involving other companies.
- *Public-Private Co-Publications:* is the number of scientific publications characterized by public-private collaboration. This indicator considers the links between the public sector and the private sector in terms of publication production. However, this collaboration between the public and the private sector is negatively associated with the ability of small and medium-sized enterprises to collaborate in technological innovation. This negative relationship means that the collaboration between the public and the private sector does not necessarily also involve small and medium-sized enterprises. In fact, it is much easier for a large company that may already have a research and development department to collaborate with public bodies to produce outputs that can be evaluated in terms of research and development. Small and medium-sized enterprises, on the other hand, may have greater difficulties in collaborating with the public and certainly it may be even more difficult for consortia of small and medium-sized enterprises to establish partnerships with the public for the related bureaucratic-administrative issues and for the scarce economic and financial incentives.

4. Clusterization with k-Means Algorithm Optimized with the Silhouette Coefficient

A clustering was then carried out using the k-Means algorithm to verify the presence of groupings among European countries by value of innovative collaborative companies. The choice of the optimal number of clusters was achieved through the application of the following criteria, namely:

- maximization of the value of the Silhouette coefficient for clusters;

- value of the Silhouette coefficient of each element of every cluster greater than 0.

By applying these two rules, four different clusters have been identified as follows:

- *Cluster 1*: Poland, Latvia, Malta, Bulgaria, Hungary, Italy, Spain, Ukraine, Romania, North Macedonia, Turkey, Serbia, Slovakia, Portugal, Croatia;
- *Cluster 2:* Ireland, Czechia, France, Sweden, Slovenia, Cyprus, Denmark, Germany, Montenegro, Luxembourg, Netherlands, Lithuania, Switzerland;
- *Cluster 3:* Belgium, United Kingdom, Iceland;
- *Cluster 4:* Finland, Greece, Estonia, Norway, Austria

It is possible to establish an analysis of the ordering of the clusters considering the value of the median of the observed variable or the ability of European small and medium-sized enterprises to collaborate with others. The analysis shows that the median value of the countries of cluster 4-C4 is equal to 347,832, the median value of the countries of cluster 3-C3 is equal to 295,538, the median value of the countries of cluster 2-C2 is equal to 162,477, the median value of cluster 1-C1 is equal to 76,397. Therefore, the following ordering of the clusters in terms of median value derives, that is: C4 = 347,832> C3 = 295,538> C2 = 162,477> C1 = 76,397. In other words, it follows that the ability of companies in cluster 4 countries to collaborate in technological innovation is approximately 1.17 times higher than the corresponding value of companies in cluster 3-C3 countries, 2.14 times higher than those of cluster 2-C2 countries, and 4.55 times higher than those of cluster 1-C1 countries. If we look at the geographical distribution of the countries belonging to the cluster, it appears that the countries of Northern Europe have much higher levels of collaboration between companies than those of the countries of Southern Europe. This contrast between Northern and Southern Europe is not so much to be traced back to capitalist factors, although these are certainly relevant, but rather to sociological and cultural supporters in a broad sense. In fact, in the countries of northern Europe human capital and social capital tend to be much more widespread than in the countries of southern Europe which are characterized by phenomena of individualism and zero-sum play between public and private goods. It follows that the inability of small and medium-sized enterprises in Southern Europe to collaborate in building strong networks and links aimed at technological innovation ends up having a negative effect on the entire technological innovation capacity of the countries considered. These factors, being of a sociological-cultural nature, can hardly be changed in the short term through economic policies, although there are still possibilities for action. For example, European policy makers could make incentive plans aimed at small and medium-sized enterprises that are able to collaborate and create networks, consortia, associations, and partnerships. Through these incentives it could be possible, through interventions external to the socio-cultural and institutional dimension of the companies, to introduce virtuous phenomena of collaboration between the companies in technological innovation.

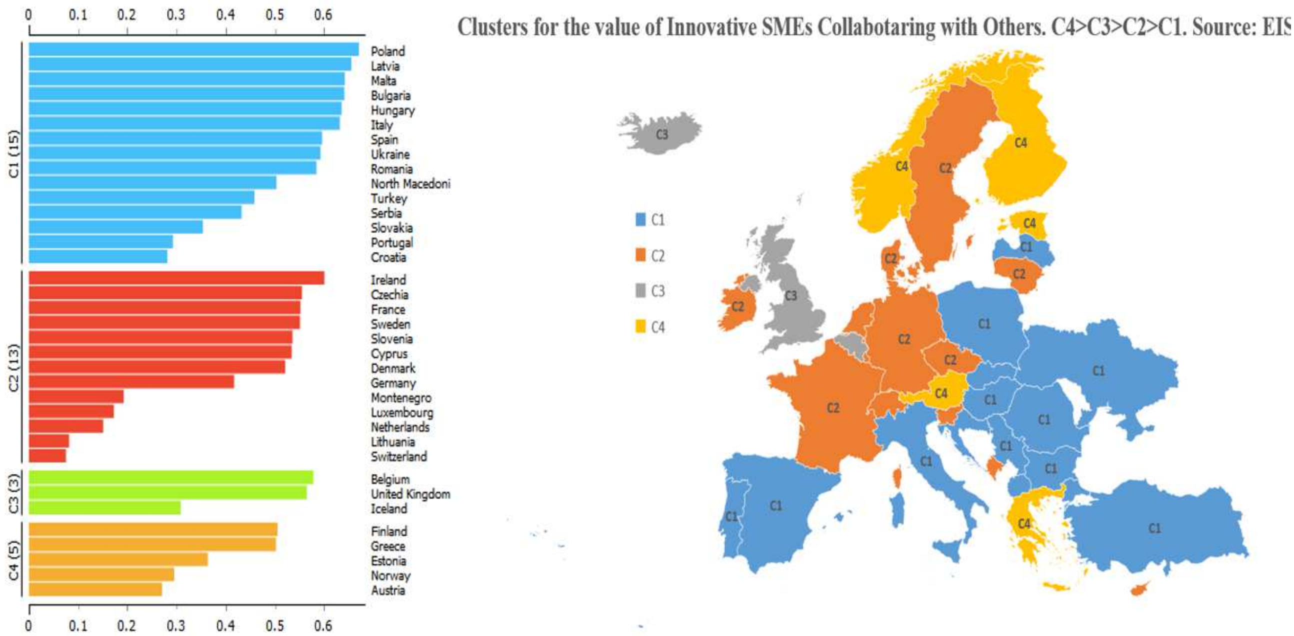


Figure 6. Clusterization with k-Means algorithm optimized with the Silhouette coefficient.

5. Network Analysis

A network analysis was carried out below to verify whether there are any network structures and particular two-way links between the historical series dynamics of the countries considered in terms of the ability of innovative small and medium-sized enterprises to collaborate. The analysis was carried out using the Manhattan distance. The analysis shows the presence of 36 nodes, 44 edges, with an average degree value of 2.44 and an amount of density equal to 0.06984. Three different network structures were found, two of which are bi-univocal and one with three elements. The three-element mesh structure is shown below:

- *Switzerland* is connected with *Luxembourg* through a link which has a value of 0.22;
- *Luxembourg* is linked to *Switzerland* through a link of 0.22 and to Montenegro with a value of 0.31;
- *Montenegro* is connected to *Luxembourg* with a link with a value of 0.31.

The two one-to-one relationships are indicated below:

- *Denmark* is connected to *Slovenia* with a link with a value of 0.34;
- *Romania* is connected to *Ukraine* with a link having a value of 0.22.

Obviously, these links are simply relations that exist in the internal dynamics of the development of countries' data and do not indicate any kind of cause-and-effect relationship among countries in terms of improving the ability of small and medium-sized enterprises to collaborate in technological innovation.

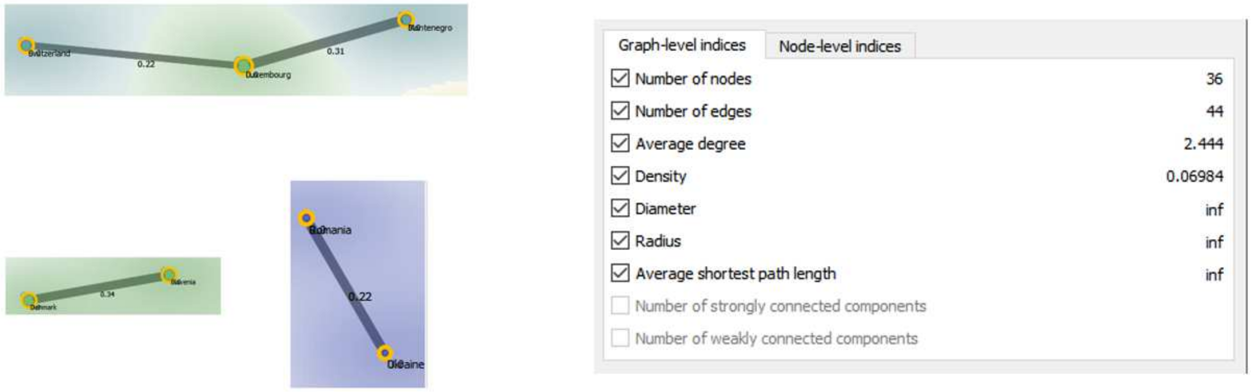


Figure 7. Results of Network Analysis with Distance of Manhattan.

6. Machine Learning and Predictions

A comparison was then made with eight different machine learning algorithms for the edition of the future value of the observed variable, i.e. the collaboration between European small and medium-sized enterprises in technological innovation. The algorithms have been analyzed in their ability to maximize R-squared and minimize the following statistical errors, namely: "Mean absolute error", "Mean squared error", "Root mean squared error". 70% of the data was used for learning the algorithms while the remaining 30% was used for the actual prediction. Based on the application of these criteria, the following ranking of the algorithms was obtained based on their predictive capacity, that is:

- *Random Forest Regression* with a payoff value of 5;
- *Simple Regression Tree* with a payoff value of 9;
- *ANN-Artificial Neural Network* with a payoff value of 10;
- *Tree Ensemble Regression* with a payoff value of 16;
- *Linear Regression* with a payoff value of 23;
- *Polynomial Regression* and *Gradient Boosted Trees Regression* with a payoff value of 26;
- *PNN-Probabilistic Neural Network* with a payoff value of 29.

Based on the application of the Random Forest Regression algorithm it was therefore possible to make the following predictions:

- Austria with an increase from a value of 195.368 up to a value of 327.021 or a variation equal to a total of 131.653 units equal to a value of 67.38%;
- Belgium with a decrease from an amount of 332.09 units up to a value of 279.49 units or equal to a value of -52 units equal to a value of -15.83%;
- Bulgaria with a decrease from an amount of 40.13 units up to a value of 36.72 units or equal to a variation of -3.41 units equal to a value of -8.49%;
- Switzerland with an increase from an amount of 111.799 units up to a value of 135.08 units or equal to a variation of 23.28 units equal to a value of 20.83%;
- Cyprus: with a decrease from an amount of 347.83 units up to a value of 192.342 units or equal to a value of -155.34 units equal to a variation of -44.70%;
- Czechia with an increase from an amount of 139.009 units up to a value of 220.356 units or equal to a value of 81.34 units equal to an amount of 58.51%;
- Germany with a decrease from an amount of 182.65 units up to a value of 154.69 units or equal to a value of -27.96 units equal to an amount of -15.31%.

- Denmark with an increase from an amount of 169.12 units up to a value of 207.74 units or equal to a value of 38.62 units equal to an amount of 22.83%;
- Estonia with a decrease from an amount of 347.832 up to a value of 292.699 units equal to a variation of -55.133 units equal to a variation of -15.85%;
- Greece with a decrease from an amount of 256.109 units up to an amount of 251.547 units or equal to a value of -4.56 units equal to a variation of -1.78%;
- Spain with an increase from an amount of 76.38 units up to a value of 106.62 units or equal to a variation of 30.22 units equal to a variation of 39.55%;
- Finland with a decrease from an amount of 347,832 units up to a value of 188,446 units or equal to a value of -159,386 units equal to a variation of -45.82%;
- France with an increase from an amount of 161,656 units up to a value of 171, 553 units or equal to a value of 9,897 units equal to an amount of 6.12%;
- Croatia with a decrease from an amount of 151.156 units up to a value of 121.987 units or equal to a variation of -29.269 units equal to a value of -19.29%;
- Hungary with a decrease from a change of 116.92 units up to a value of 96.13 units or equal to a value of -20.789 units equal to an amount of -17.78%;
- Ireland with a decrease from an amount of 267.710 units up to a value of 172.953 units or equal to a variation of -94.75 units equal to an amount of 35.39%;
- Iceland with an increase from an amount of 280.69 units up to a value of 294.42 units or equal to a variation of 13.73 units equal to a variation of 4.89%;
- Italy with a decrease from an amount of 174.26 units up to a value of 83.76 units or equal to an amount of -90.50 units equal to a value of -51.93%;
- Lithuania with an increase from a value of 162,404 units up to a value of 173.45 units or equal to a variation of 11.05 units equal to an amount of 6.80%;
- Luxembourg with an increase from a value of 162,404 units up to a value of 173,456 units equal to an amount of 11,052 units or equivalent to a value of 6.80%;
- Latvia with a variation from 66.93 units up to a value of 103.551 units or equal to a variation of 36.62 units equal to 54.71%;
- Montenegro with an increase from an amount of 106,482 units up to a value of 125,950 units or equal to a value of 19,468 units equal to a value of 18.28%;
- North Macedonia with an increase from an amount of 65.45 units up to a value of 94.06 units or equal to a variation of 28.61 units equal to an amount of 43.70%;
- Malta with an increase from a value of 92.70 units up to a value of 110.63 units or equal to a value of 17.92 units equal to a value of 19.34%;
- Netherlands with an increase from an amount of 174.832 units up to a value of 215.74 units or equal to an amount of 40.90 units equal to 23.39%;
- Norway with a decrease from an amount of 347.83 units up to a value of 195.89 units or equal to an amount of -151 units equal to a value of -43.68%;
- Poland with an increase from an amount of 40.83 units up to a value of 67.28 units or equal to a value of 26.44 units equal to a value of 64.76%;
- Portugal with an increase from an amount of 86.80 units up to a value of 94.56 units or equal to a variation of 7.76 units equal to an amount of 8.94%;
- Romania with an increase from an amount of 19.68 units up to a value of 21.28 units equal to a value of 1.60 units equal to 8.14%;
- Serbia with a value from an amount of 141,351 units up to a value of 111,392 units or equal to a variation of -29.59 units equal to a variation of -21.19%;
- Sweden with a decrease from an amount of 186.76 units up to a value of 177.78 units or equal to a variation of -8.97 units equal to a variation of -4.80%;

- Slovenia with an increase from a value of 162.47 units up to a value of 229.85 units or equal to a variation of 67.37 units equal to a value of 41.46%;
- Slovakia with an increase from an amount of 92,450 units up to a value of 95.24 units or equal to a variation of 2.79 units equal to a variation of 3.02%;
- Turkey with a variation of 60.13 units up to a variation of 88.98 units or equal to a variation of 28.85 units equal to a variation of 47.97%;
- Ukraine with an increase from an amount of 6.59 units up to a variation of 31.59 units or equal to a variation of 24.99 units equal to a variation of 378.86%;
- United Kingdom with a decrease from an amount of 295.53 units up to a value of 294.52 units or equal to a value of -1.01 units equal to -0.34%.

On average, using the Random Forest Regression algorithm it is possible to predict a reduction from 165.62 units up to 158.28 units or equal to -7.3 units equivalent to -4.3% of the trend of the variable considered, i.e., the capacity of small and medium-sized enterprises to collaborate in innovation,

Prediction with the original data using the Random Forest Regression algorithm					
Country	**2021**	**Prediction**	**Country**	**2021**	**Prediction**
Austria	195,368	327,021	*Lithuania*	162,404	173,456
Belgium	332,092	279,495	*Luxembourg*	154,751	132,809
Bulgaria	40,133	36,722	*Latvia*	66,930	103,551
Switzerland	111,799	135,087	*Montenegro*	106,482	125,950
Cyprus	347,832	192,342	*North Macedonia*	65,455	94,065
Czechia	139,009	220,356	*Malta*	92,702	110,631
Germany	182,658	154,690	*Netherlands*	174,832	215,741
Denmark	169,124	207,744	*Norway*	347,832	195,897
Estonia	347,832	292,699	*Poland*	40,835	67,281
Greece	256,109	251,547	*Portugal*	86,802	94,565
Spain	76,398	106,621	*Romania*	19,684	21,288
Finland	347,832	188,446	*Serbia*	141,351	111,392
France	161,656	171,553	*Sweden*	186,761	177,784
Croatia	151,156	121,987	*Slovenia*	162,477	229,854
Hungary	116,921	96,132	*Slovakia*	92,450	95,245
Ireland	267,710	172,953	*Turkey*	60,133	88,984
Iceland	280,691	294,423	*Ukraine*	6,597	31,591
Italy	174,263	83,761	*United Kingdom*	295,538	294,525

Figure 8. Prediction with original data using Random Forest Regression.

7. Prediction Using Machine Learning Algorithms with Augmented Data

A further prediction was then made through augmented data. In other words, the results of the prediction analyzed in the previous paragraph have been added to the time series. Increased data were thus obtained. Eight different machine learning algorithms aimed at prediction were used to analyze the augmented data. The algorithms were analyzed based on the ability to maximize R-squared and essential statistical errors i.e. "Mean absolute error", "Mean squared error", and "Root mean squared error". The algorithms were trained with 70% of the available data while the remaining 30% were used for the actual prediction. Therefore, the following ranking of the algorithms in terms of performance was derived, that is:

- Random Forest Regression with a payoff value of 4;
- Linear Regression with a payoff of 8;
- Tree Ensemble Regression with a payoff equal to 12;
- Polynomial Regression with a payoff value of 18;
- ANN-Artificial Neural Network with a payoff value of 20;
- Gradient Boosted Tree Regression with a payoff value of 24;
- PNN-Probabilistic Neural Network with a payoff value of 26;
- Simple Regression Tree with a payoff value of 32.

Prediction Using Augmented Data Obtained Using the Random Forest Regression Algorithm				
Country	**Prediction 2021**	**Prediction Augmented Data**	**Absolute Variation**	**Percentage Variation**
Ciprus	192,342	227,778	35,436	18,423
Finland	188,446	254,688	66,242	35,152
France	171,553	181,084	9,531	5,556
Italy	83,761	105,555	21,794	26,019
Lithuania	173,456	157,855	-15,601	-8,994
Norway	195,897	184,822	-11,075	-5,653
Romania	21,288	54,363	33,075	155,369
Slovenia	229,854	216,674	-13,18	-5,734
Slovakia	95,245	106,423	11,178	11,736
Ukraine	31,591	54,363	22,772	72,084
United Kingdom	294,525	289,262	-5,263	-1,787
Mean	152,542	166,624	14,083	9,232

Figure 9. Prediction using Augmented Data.

Therefore, by applying the best performer algorithm or the Random Forest Regression to the augmented data, the following predictions were made, namely:

- Cyprus with an increase from an amount of 192.34 units up to a value of 227.77 units or equal to a variation of 35.43 units equal to a value of 18.42%;
- Finland with an increase from a value of 188.44 units up to a value of 254.68 units or equal to a variation of 66.24 units equal to a value of 18.42%;
- France with an increase from an amount of 171.553 units up to a value of 181.08 units equal to a value of 9.53 units equivalent of 5.55%;
- Italy with an increase from an amount of 83.76 units up to a value of 105.55 units or equal to a value of 21.79 units equal to a variation of 26.019%;
- Lithuania with an increase from an amount of 173,456 units up to a value of 157,855 units or equal to a variation of -15.6 units equal to a value of -8.99%;
- Norway with a decrease from an amount of 195.897 units up to a value of 184.82 units or equal to a variation of -11.07 units equal to a variation of -5.65%;
- Romania with an increase from a value of 21.288 units up to a value of 54.36 units or equal to a variation of 33.07 units equal to an amount of 155.36%;
- Slovenia with a decrease from an amount of 229.85 units up to a value of 216.67 units or equal to a value of 33.07 units equal to a variation of 155.36%;
- Slovakia with an increase from an amount of 229,854 units up to a value of 216,674 units or equal to a variation of -13.18 units equal to -5.73%;
- Ukraine with an increase from an amount of 31.59 units up to a value of 54.36 units or equal to a value of 22.77 units equal to 72.08%;
- United Kingdom with a decrease from an amount of 294.525 units up to a value of 289.262 units or equal to a value of -5.26 units equal to a value of -1.78%.

On average, the predicted value with the increased data and the Random Forest Regression for the countries considered is expected to grow from an amount of 152.542 units up to a value of 166.62 units or equal to a value of 14.08 units equal to a value of 9.23%.

Comparison among statistical errors and R-squared between prediction with original data and prediction with augmented data				
Statistical Metrics	Original Data	Augmented Data	Absolute Variation	Percentage Variation
	Random Forest Regression	Random Forest Regression		
R^2	0,033274166	0,918671301	0,885	2660,915
Mean absolute error	0,186156067	0,054200754	-0,132	-70,884
Mean squared error	0,074290063	0,00642922	-0,068	-91,346
Root mean squared error	0,272562035	0,080182414	-0,192	-70,582
Mean of Errors	0,177669388	0,046937463	-0,131	-73,582

Figure 10. Comparison among R-Squared and Statistical Errors between prediction with original data and prediction with augmented data.

Finally, by comparing the value of the R-square and of the statistical errors of the prediction made with the original data and the prediction made with the increased data, there is a significant improvement in the prediction with the increased data. In fact, it is possible to verify both an increase in the R-squared value and a reduction in statistical errors. Specifically, it is possible to check the following improvements:

- R-squared: growth from 0.03 to 0.91 with an absolute change of 0.88 units and a percentage change of 2660.9%;
- Mean Absolute Error with a reduction from 0.18 to 0.054 units or equal to -0.132 units equal to -70.884%;
- Mean Squared Error with a reduction from 0.074 to 0.0064 units or equal to a variation of -0.068 units equal to -91.34%;
- Root Mean Squared Error with a reduction from 0.27 units down to 0.08 units or equal to -0.19 units equal to -70.58%.

Overall, taking the average of the three statistical errors alone, there is a reduction from 0.177 units up to 0.046 units or equal to -0.131 units equal to -73.58%. It follows that the prediction with augmented data is more efficient than the prediction with the original data thanks to a reduction of statistical errors considered equal to an average of 73.58%.

8. Conclusions

Collaboration between small and medium-sized enterprises in the sense of technological innovation is a strategic factor for the success of innovation systems at national level. Europe appears to be lagging in the production of intangible goods in the context of the tech war between the US and China. The development of economic policies capable of allowing greater collaboration between small and medium-sized enterprises in the sense of technological innovation can allow unlocking an innovation potential that could generate a positive effect in terms of economic growth and convergence between European regions. However, the ability of companies to collaborate depends not only on the presence of adequate human capital, financial capital, networks, and technologies, but also on the presence of human capital. In fact, as the data show, collaboration between SMEs in the sense of innovation is easier in countries where there is more social capital, while it is less widespread in countries that show a more individualistic orientation. In this sense, economic policies should also offer incentives to remove the individualistic culture, present above all in Southern Europe, and orient companies towards a more prosperous and confident collaboration in the sense of innovation.

9. Bibliography

Akinwale, Y. O., 2018. Empirical analysis of inbound open innovation and small and medium-sized enterprises' performance: Evidence from oil and gas industry. *South African Journal of Economic and Management Sciences,* 21(1), pp. 1-9.

Anderson, A. R. & Hardwick, J., 2017. Collaborating for innovation: the socialised management of knowledge. *International Entrepreneurship and Management Journal ,* 13(4), pp. 1181-1197.

Benhayoun, L., Le Dain, M. A., Dominguez-Péry, C. & Lyons, A. C., 2020. SMEs embedded in collaborative innovation networks: how to measure their absorptive capacity?. *Technological Forecasting and Social Change,* 159(120196).

Boldrin, M. & Levine, D., 2002. The case against intellectual property. *American Economic Review,* 92(2), pp. 209-212.

Costantiello, A., Laureti, L., Matarrese, M. & Leogrande, A., 2022. The innovation-sales growth nexusin Europe.. *Academy of Accounting and Financial Studies Journal,* 26(2), pp. 1-14.

Costantiello, A., Leogrande, A. & Laureti, L., 2022. The corporate innovation in Europe. *International Journal of Entrepreneurship,* 26(S4), pp. 1-9.

D'Angelo, A. & Baroncelli, A., 2020. An investigation over inbound open innovation in SMEs: Insights from an Italian manufacturing sample. *Technology Analysis & Strategic Management,* 32(5), pp. 542-560.

De Noni, I., Orsi, L. & Belussi, F., 2018. The role of collaborative networks in supporting the innovation performances of lagging-behind European regions. *Research Policy,* 47(1), pp. 1-13.

Del Vecchio, P. et al., 2018. Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management,* 27(1), pp. 6-22.

Grimaldi, M., Greco, M. & Cricelli, L., 2021. A framework of intellectual property protection strategies and open innovation. *Journal of Business Research,* Volume 123, pp. 156-164.

Grimsdottir, E. & Edvardsson, I. R., 2018. Knowledge management, knowledge creation, and open innovation in Icelandic SMEs. *Sage Open,* 4(2158244018807320), p. 8.

Ioanid, A., Deselnicu, D. C. & Militaru, G., 2018. The impact of social networks on SMEs' innovation potential. *Procedia Manufacturing,* Volume 22, pp. 936-941.

Jasimuddin, S. M. & Naqshbandi, M. M., 2019. Knowledge infrastructure capability, absorptive capacity and inbound open innovation: evidence from SMEs in France.. *Production Planning & Control,* 10-12(30), pp. 893-906.

Jespersen, K., Rigamonti, D., Jensen, M. B. & Bysted, R., 2018. Analysis of SMEs partner proximity preferences for process innovation. *Small Business Economics,* 51(4), pp. 879-904.

Laureti, L., Costantiello, A., Leogrande, A. & Leogrande, D., 2022. Creative Intangible Products and Services for Industry in Europe. *Academy of Entrepreneurship Journal, ,* 28(55), pp. 1-9.

Laureti, L., Costantiello, A., Matarrese, M. & Leogrande, A., 2022. Enterprises Providing ICT Training in Europe. *University Library of Munich, Germany,* Issue 111756.

Laureti, L., Costantiello, A., Matarrese, M. & Leogrande, A., 2022. Foreign Doctorate Students in Europe. *University Library of Munich, Germany.*

Leogrande, A. & Costantiello, A., 2021. Human Resources in Europe. Estimation, Clusterization, Machine Learning and Prediction. *American Journal of Humanities and Social Sciences Research (AJHSSR)* , 9(5), pp. 240-259.

Leogrande, A., Costantiello, A. & Laureti, L., 2021. The Broadband Penetration in Europe. *Journal of Applied Economic Sciences,* 3(73), p. 324– 349.

Leogrande, A., Costantiello, A. & Laureti, L., 2021. The Impact of Venture Capital Expenditures on Innovation in Europe. *American Journal of Humanities and Social Sciences Research (AJHSSR),* 10(5), pp. 85-102.

Leogrande, A., Laureti, L. & Costantiello, A., 2022. The Innovation Index in Europe. *University Library of Munich, Germany,* Issue 112841.

Radziwon, A. & Bogers, M., 2019. Open innovation in SMEs: Exploring inter-organizational relationships in an ecosystem. *Technological Forecasting and Social Change,* Volume 146, pp. 573-587.

Romer, P. M., 1994. The origins of endogenous growth. *Journal of Economic perspectives,* 8(1), pp. 3-22.

Saastamoinen, J., Reijonen, H. & Tammi, T., 2018. Should SMEs pursue public procurement to improve innovative performance?. *Technovation,* Volume 69, pp. 2-14.

Schumpeter, J. A., 1934. *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle.* Transaction Books ed. s.l.: New Jersey.

Solow, R. M., 1956. A contribution to the theory of economic growth. *The quarterly journal of economics,* 70(1), pp. 65-94.

Ueasangkomsate, P. & Jangkot, A., 2019. Enhancing the innovation of small and medium enterprises in food manufacturing through Triple Helix Agents. *Kasetsart Journal of Social Sciences,* 40(2), pp. 380-388.

Zahoor, N. & Al-Tabbaa, O., 2020. Inter-organizational collaboration and SMEs' innovation: A systematic review and future research directions. *Scandinavian Journal of Management,* 2(101109), p. 36.

Zaridis, A., Vlachos, I. & Bourlakis, M., 2021. SMEs strategy and scale constraints impact on agri-food supply chain collaboration and firm performance. *Production Planning & Control,* 32(14), pp. 1165-1178.

10. Appendix

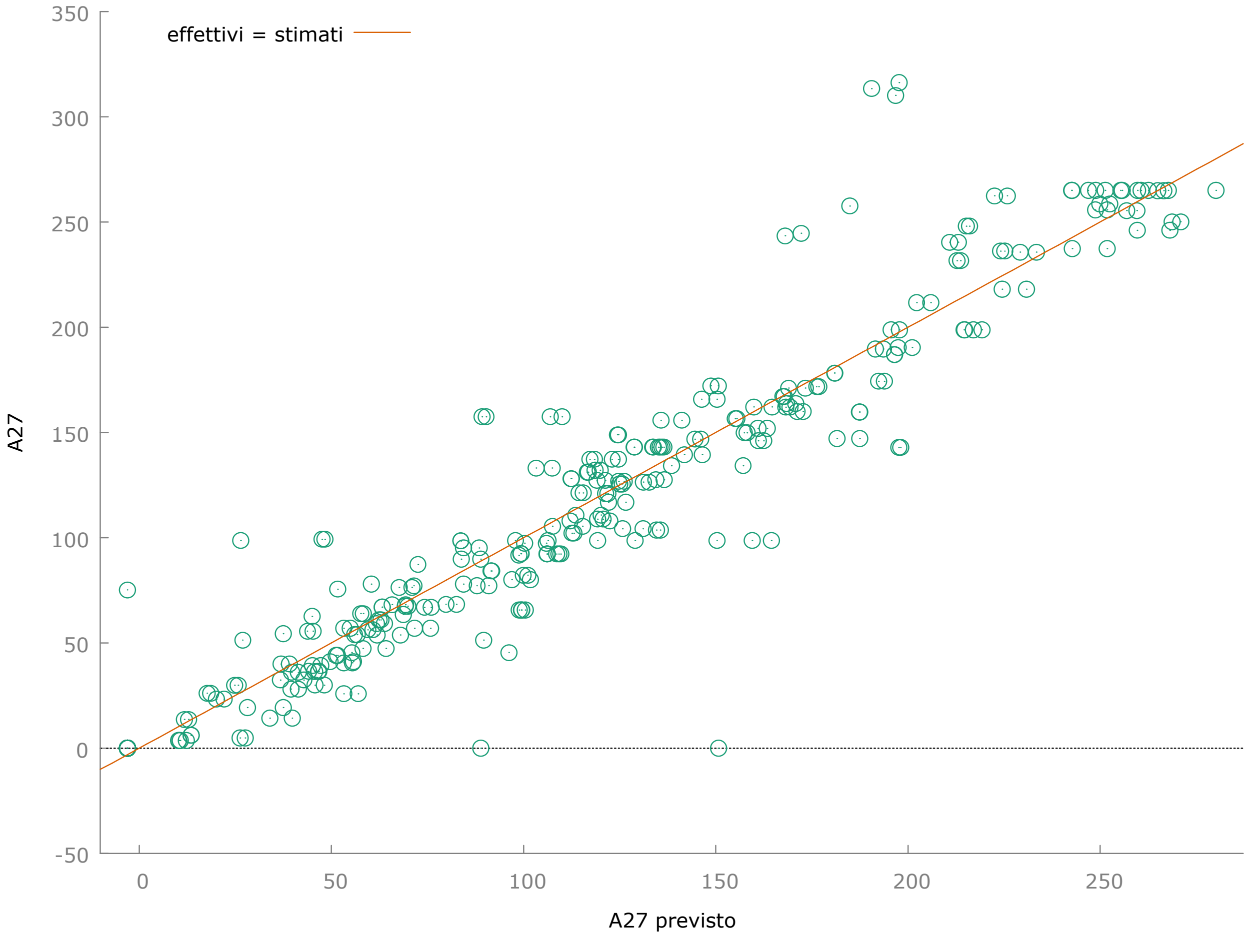
Regression Results

Variable	**Label**	***Meaning***
y	A27	*Innovative SMEs Collaborating*
x_1	A5	*Broadband Penetration*
x_2	A17	*Finance and support*
x_3	A20	*Foreign-controlled enterprises – share of value added (SD)*
x_4	A33	*Linkages*
x_5	A37	*New doctorate graduates*
x_6	A38	*Non-R&D innovation expenditure*
x_7	A42	*Population size (SD)*
x_8	A43	*Private co-funding of public R&D expenditures*
x_9	A45	*Public-private co-publications*
x_{10}	A50	*Share High and Medium high-tech manufacturing (SD)*
x_{11}	A59	*Venture capital*

Modello 821: Pooled OLS, usando 360 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche = 10
Variabile dipendente: A27

	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	−3,01107	2,42873	−1,240	0,2159	
A5	0,0767732	0,0167163	4,593	<0,0001	***
A17	0,432059	0,0817547	5,285	<0,0001	***
A20	−0,370247	0,0937700	−3,948	<0,0001	***
A33	2,74646	0,0731256	37,56	<0,0001	***
A37	−0,104530	0,0351432	−2,974	0,0031	***
A38	0,0549617	0,0210269	2,614	0,0093	***
A42	−0,539015	0,107459	−5,016	<0,0001	***
A43	−1,47634	0,0610346	−24,19	<0,0001	***
A45	−0,510753	0,0273427	−18,68	<0,0001	***
A50	0,449130	0,0780471	5,755	<0,0001	***
A59	−0,160145	0,0335182	−4,778	<0,0001	***

Media var. dipendente	100,1376	SQM var. dipendente	83,72179
Somma quadr. residui	196180,0	E.S. della regressione	23,74312
R-quadro	0,922038	R-quadro corretto	0,919574
F(11, 348)	374,1555	P-value(F)	2,7e-185
Log-verosimiglianza	−1644,941	Criterio di Akaike	3313,882
Criterio di Schwarz	3360,515	Hannan-Quinn	3332,424
rho	0,713321	Durbin-Watson	0,610062

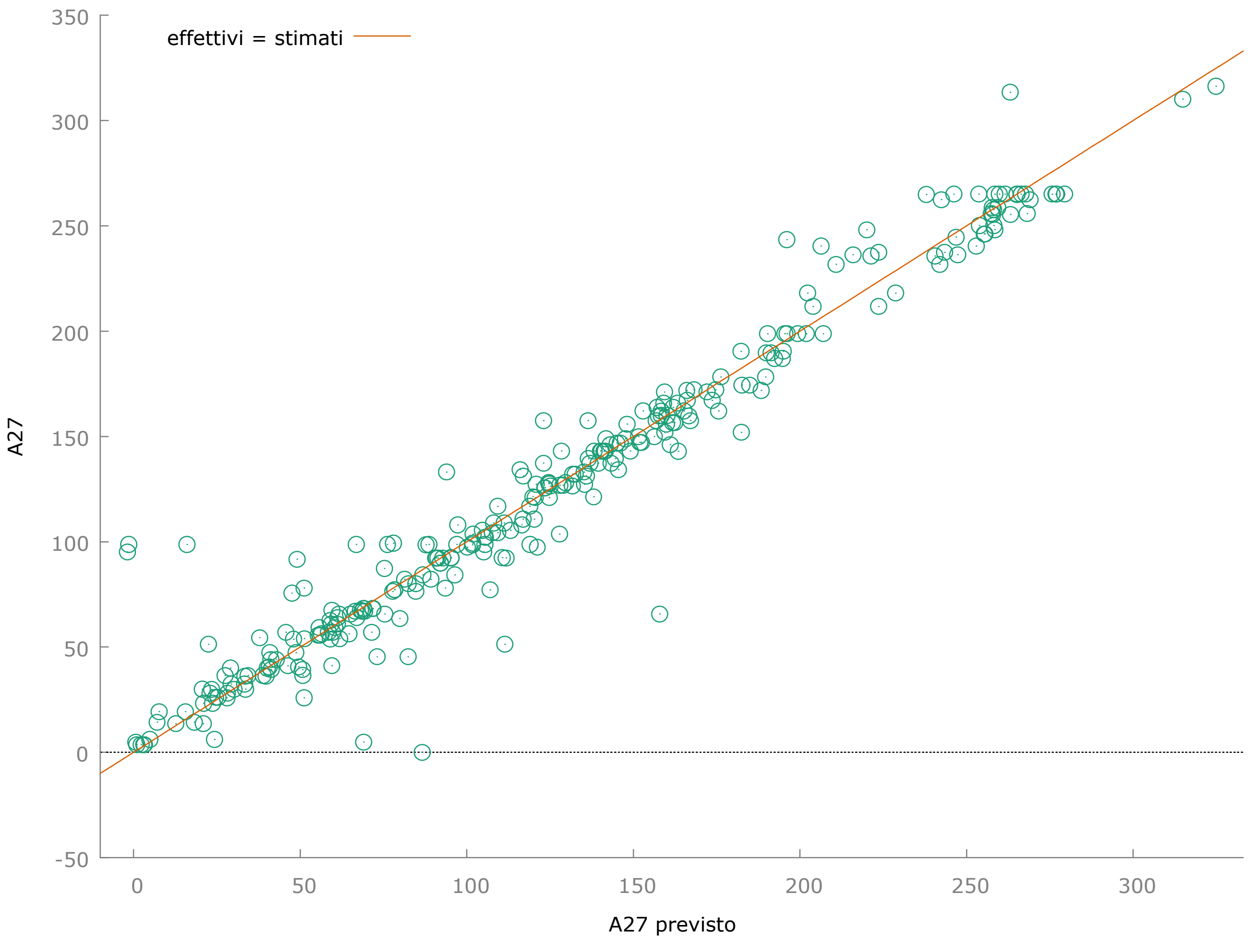


Modello 822: Panel dinamico a un passo, usando 288 osservazioni
Incluse 36 unità cross section
Matrice H conforme ad Ox/DPD
Variabile dipendente: A27

	Coefficiente	*Errore Std.*	*z*	*p-value*	
A27(-1)	0,0492691	0,0419695	1,174	0,2404	
const	−1,63253	1,25254	−1,303	0,1924	
A5	0,0875385	0,0442494	1,978	0,0479	**
A17	0,361759	0,174196	2,077	0,0378	**
A20	−0,746436	0,221106	−3,376	0,0007	***
A33	2,64433	0,272346	9,709	<0,0001	***
A37	−0,0644522	0,0289899	−2,223	0,0262	**
A38	0,112775	0,0433353	2,602	0,0093	***
A42	−0,382073	0,112185	−3,406	0,0007	***
A43	−1,40862	0,116399	−12,10	<0,0001	***
A45	−0,522863	0,0676150	−7,733	<0,0001	***
A50	0,687578	0,221234	3,108	0,0019	***
A59	−0,137131	0,0584913	−2,344	0,0191	**

Somma quadr. residui	87482,84	E.S. della regressione	17,83590

Numero di strumenti = 36
Test per errori AR(1): z = -1,39136 [0,1641]
Test per errori AR(2): z = -1,56642 [0,1173]
Test di sovra-identificazione di Sargan: Chi-quadro(23) = 66,971 [0,0000]
Test (congiunto) di Wald: Chi-quadro(12) = 4378,87 [0,0000]



Modello 823: Effetti fissi, usando 360 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche = 10
Variabile dipendente: A27

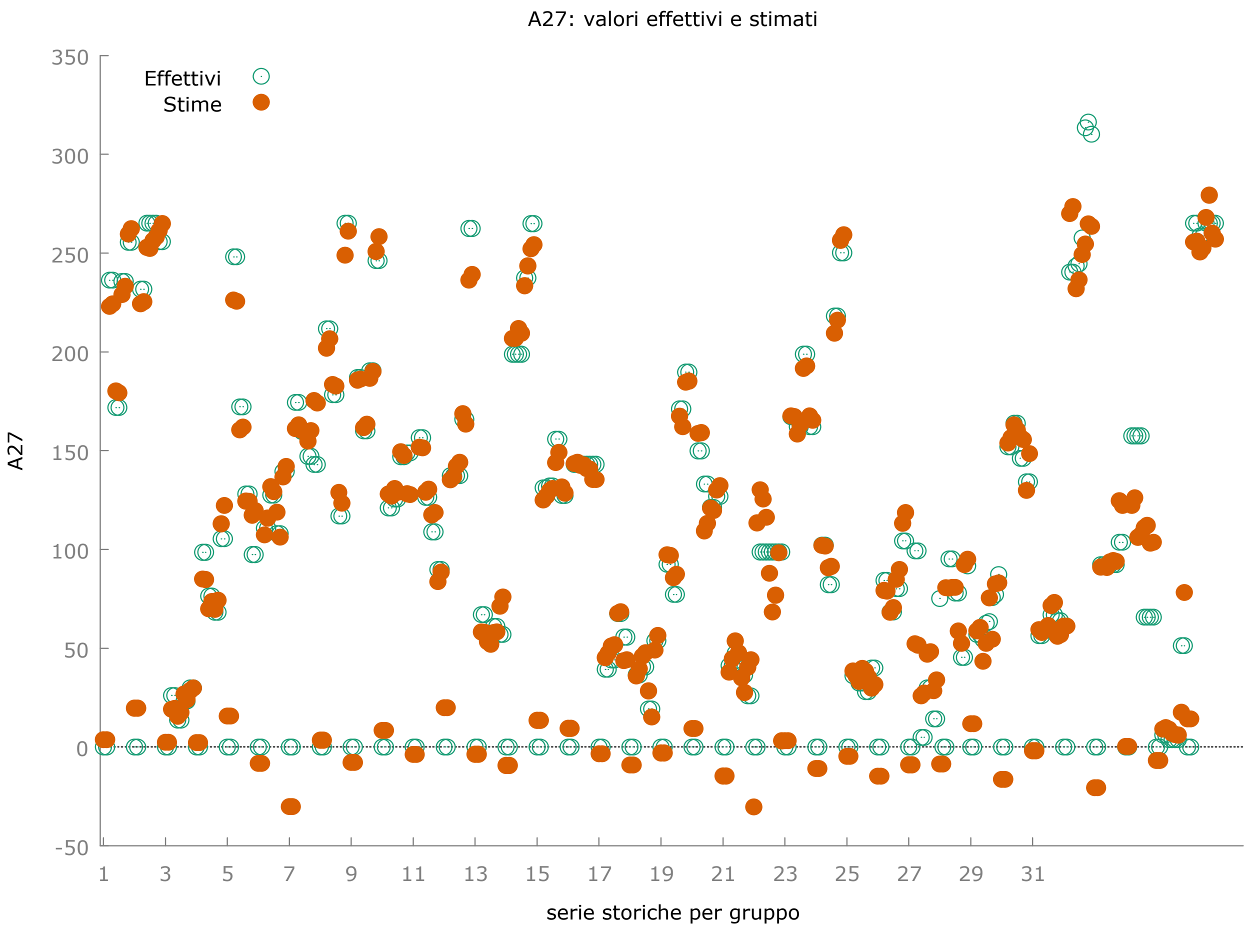
	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	−0,511076	2,12585	−0,2404	0,8102	
A5	0,0760765	0,0217155	3,503	0,0005	***
A17	0,511076	0,112219	4,554	<0,0001	***
A20	−0,512540	0,143997	−3,559	0,0004	***

A33	2,56871	0,0987637	26,01	<0,0001	***
A37	−0,100688	0,0498983	−2,018	0,0445	**
A38	0,0743446	0,0281662	2,639	0,0087	***
A42	−0,604590	0,0956246	−6,323	<0,0001	***
A43	−1,42668	0,0862796	−16,54	<0,0001	***
A45	−0,467406	0,0411579	−11,36	<0,0001	***
A50	0,607749	0,0989633	6,141	<0,0001	***
A59	−0,210267	0,0465560	−4,516	<0,0001	***

Media var. dipendente	100,1376	SQM var. dipendente	83,72179
Somma quadr. residui	112673,7	E.S. della regressione	18,97314
R-quadro LSDV	0,955223	R-quadro intra-gruppi	0,919504
LSDV F(46, 313)	145,1577	P-value(F)	2,3e-184
Log-verosimiglianza	−1545,124	Criterio di Akaike	3184,249
Criterio di Schwarz	3366,896	Hannan-Quinn	3256,873
rho	0,396112	Durbin-Watson	1,028478

Test congiunto sui regressori -
Statistica test: F(11, 313) = 325,037
con p-value = P(F(11, 313) > 325,037) = 6,0381e-164

Test per la differenza delle intercette di gruppo -
Ipotesi nulla: i gruppi hanno un'intercetta comune
Statistica test: F(35, 313) = 6,62785
con p-value = P(F(35, 313) > 6,62785) = 8,2889e-022



Modello 824: Effetti casuali (GLS), usando 360 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche = 10
Variabile dipendente: A27

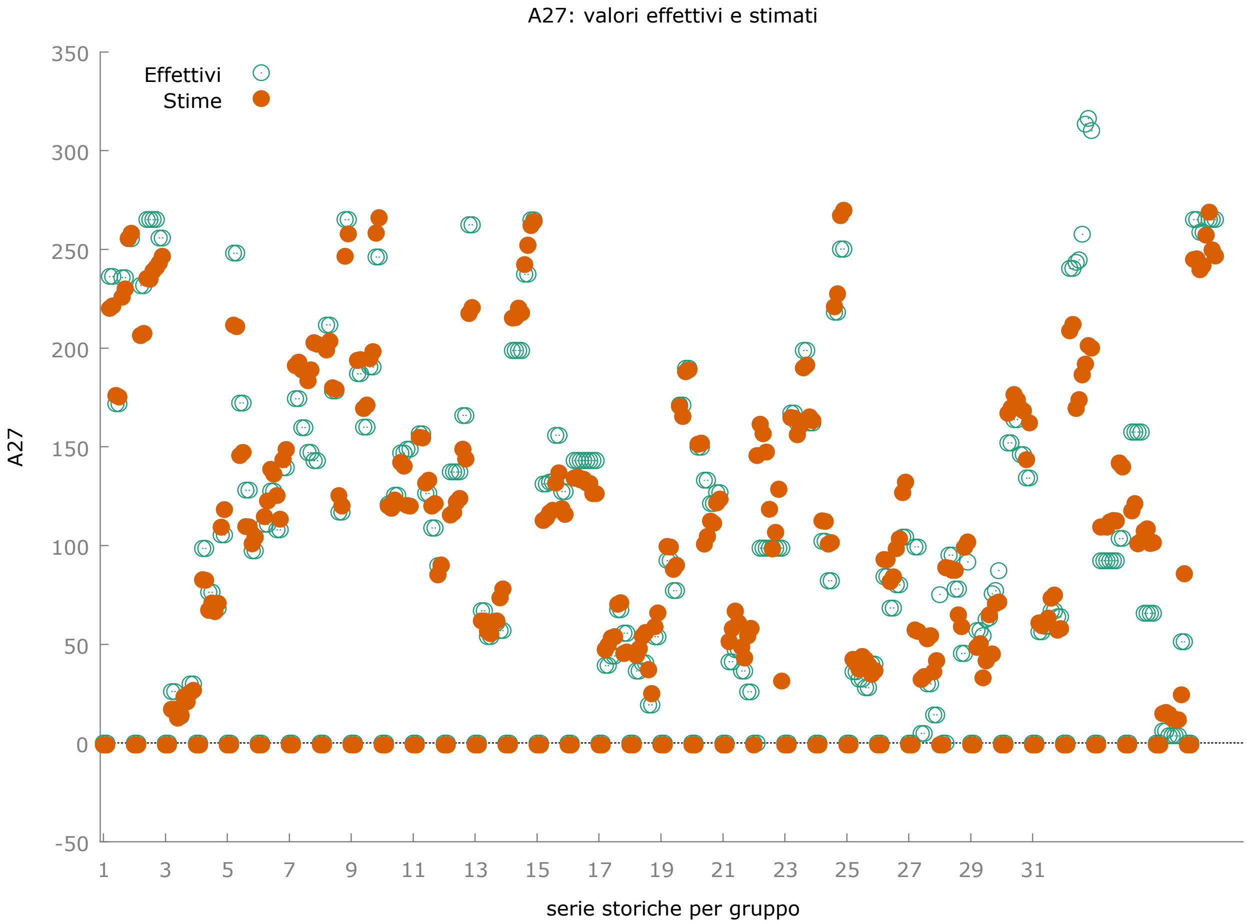
	Coefficiente	*Errore Std.*	*z*	*p-value*	
const	−0,908071	3,44961	−0,2632	0,7924	
A5	0,0759405	0,0196177	3,871	0,0001	***
A17	0,489314	0,0999653	4,895	<0,0001	***
A20	−0,469702	0,124467	−3,774	0,0002	***
A33	2,61679	0,0883666	29,61	<0,0001	***
A37	−0,101697	0,0440564	−2,308	0,0210	**
A38	0,0677468	0,0252248	2,686	0,0072	***
A42	−0,592895	0,0929575	−6,378	<0,0001	***
A43	−1,44138	0,0764026	−18,87	<0,0001	***
A45	−0,479701	0,0355506	−13,49	<0,0001	***
A50	0,562415	0,0897524	6,266	<0,0001	***
A59	−0,196025	0,0414012	−4,735	<0,0001	***

Media var. dipendente	100,1376	SQM var. dipendente	83,72179
Somma quadr. residui	200815,7	E.S. della regressione	23,98756
Log-verosimiglianza	−1649,145	Criterio di Akaike	3322,290
Criterio di Schwarz	3368,923	Hannan-Quinn	3340,832
rho	0,396112	Durbin-Watson	1,028478

Varianza 'between' = 278,119
Varianza 'within' = 359,98
Theta usato per la trasformazione = 0,661473
Test congiunto sui regressori -
Statistica test asintotica: Chi-quadro(11) = 3975,31
con p-value = 0

Test Breusch-Pagan -
Ipotesi nulla: varianza dell'errore specifico all'unità = 0
Statistica test asintotica: Chi-quadro(1) = 191,393
con p-value = 1,57862e-043

Test di Hausman -
Ipotesi nulla: le stime GLS sono consistenti
Statistica test asintotica: Chi-quadro(11) = 4,92062
con p-value = 0,93495



Modello 825: WLS, usando 360 osservazioni
Incluse 36 unità cross section
Variabile dipendente: A27
Pesi basati sulle varianze degli errori per unità

	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	0,802650	0,750670	1,069	0,2857	
A5	0,0456031	0,00643061	7,092	<0,0001	***
A17	0,225037	0,0300339	7,493	<0,0001	***
A20	−0,258555	0,0340651	−7,590	<0,0001	***
A33	2,96012	0,0262575	112,7	<0,0001	***
A37	−0,114256	0,0112799	−10,13	<0,0001	***
A38	0,0253664	0,00671383	3,778	0,0002	***
A42	−0,258257	0,0347699	−7,428	<0,0001	***
A43	−1,49518	0,0194185	−77,00	<0,0001	***
A45	−0,558165	0,0102422	−54,50	<0,0001	***
A50	0,257254	0,0392901	6,548	<0,0001	***
A59	−0,0941221	0,0129400	−7,274	<0,0001	***

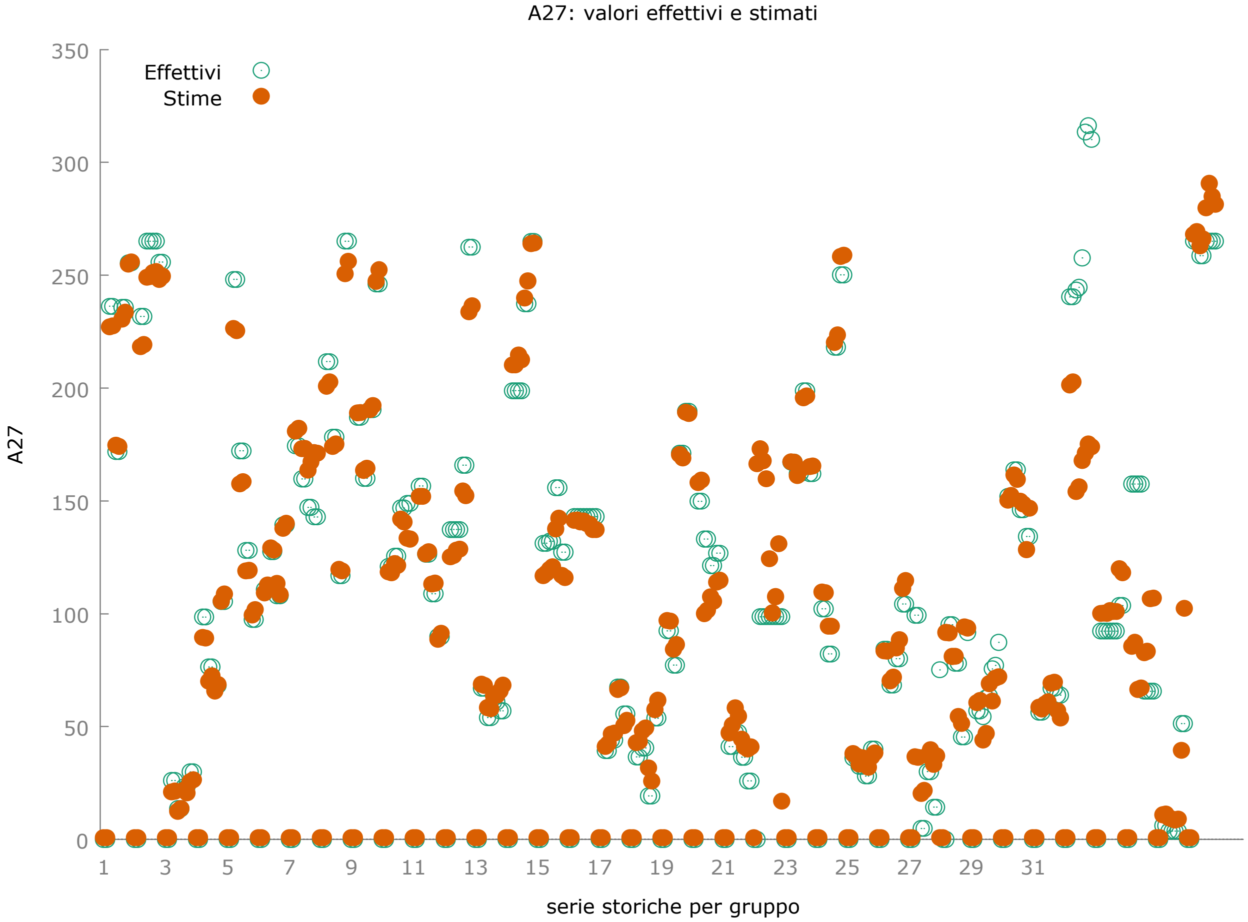
Statistiche basate sui dati ponderati:

Somma quadr. residui	230,3795	E.S. della regressione	0,813640
R-quadro	0,992546	R-quadro corretto	0,992311

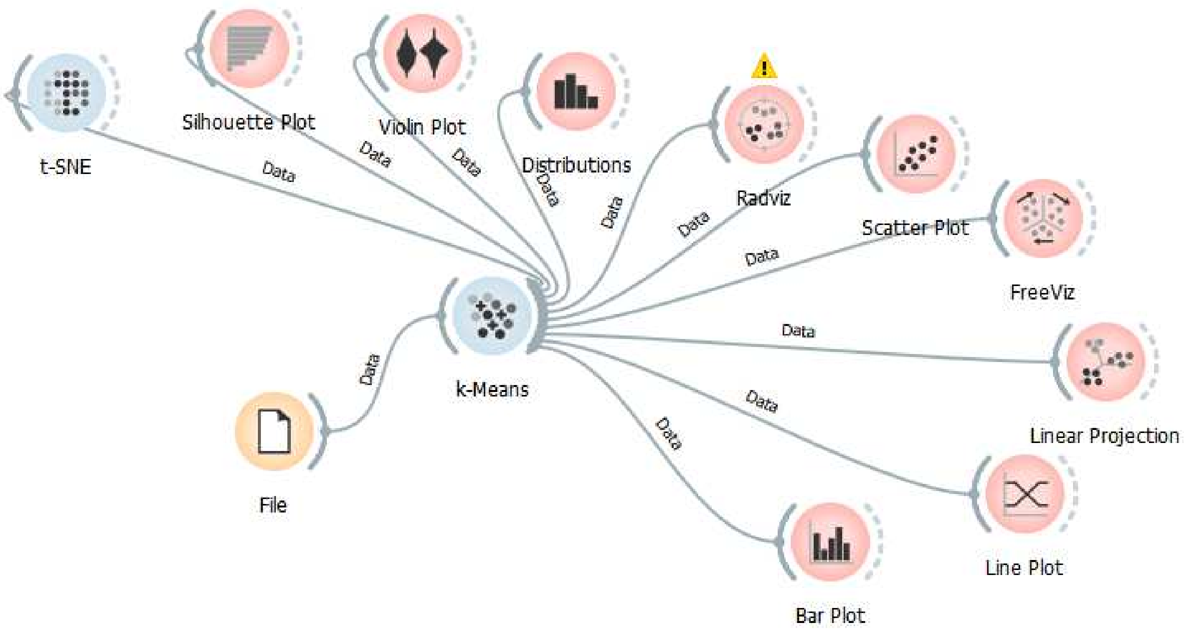
F(11, 348)	4212,740	P-value(F)	0,000000
Log-verosimiglianza	−430,4702	Criterio di Akaike	884,9403
Criterio di Schwarz	931,5736	Hannan-Quinn	903,4826

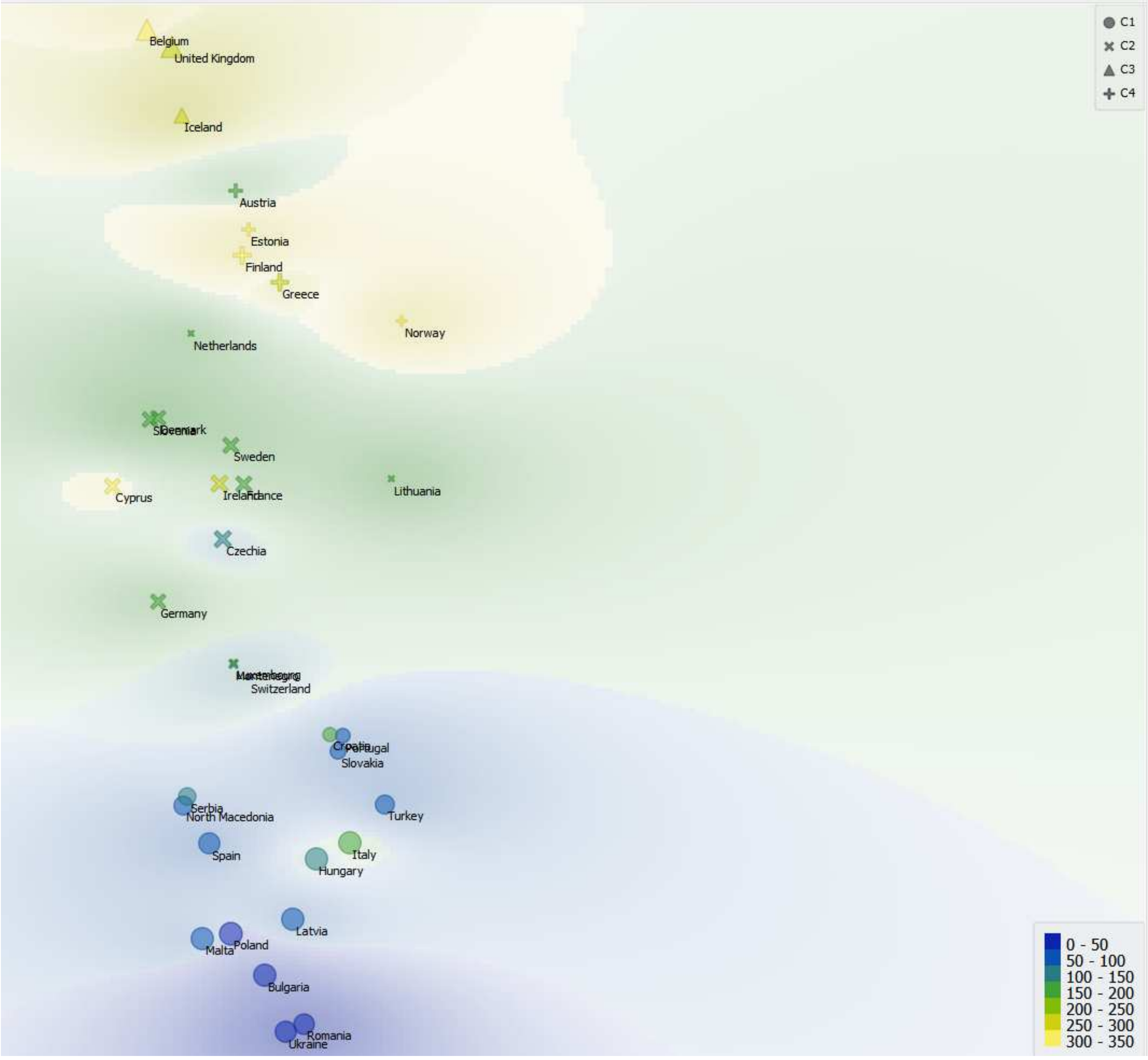
Statistiche basate sui dati originali:

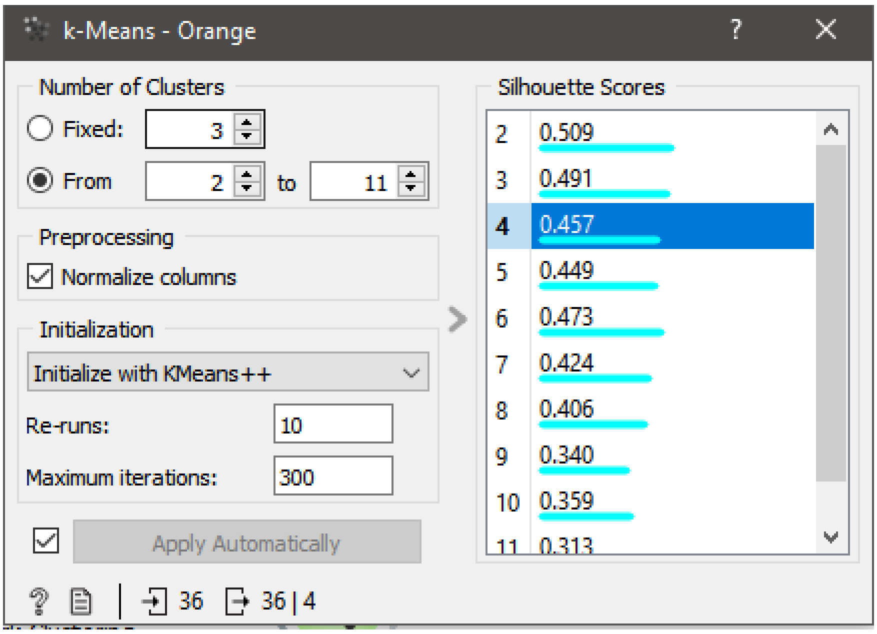
Media var. dipendente	100,1376	SQM var. dipendente	83,72179
Somma quadr. residui	217432,1	E.S. della regressione	24,99610

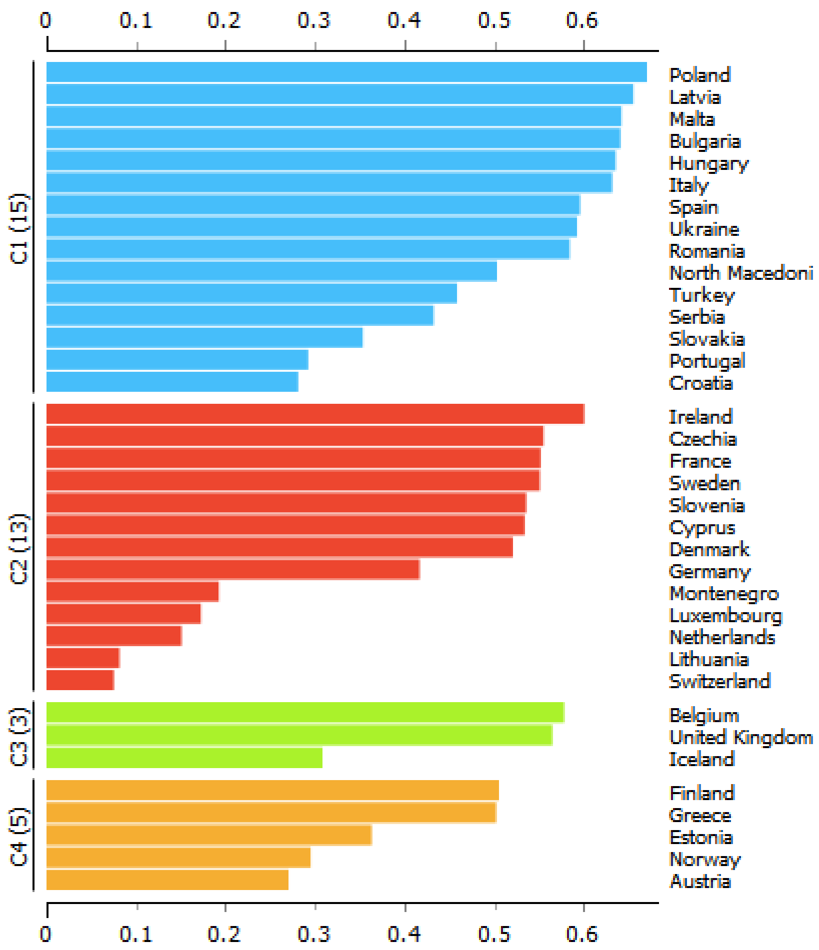


Clusterization

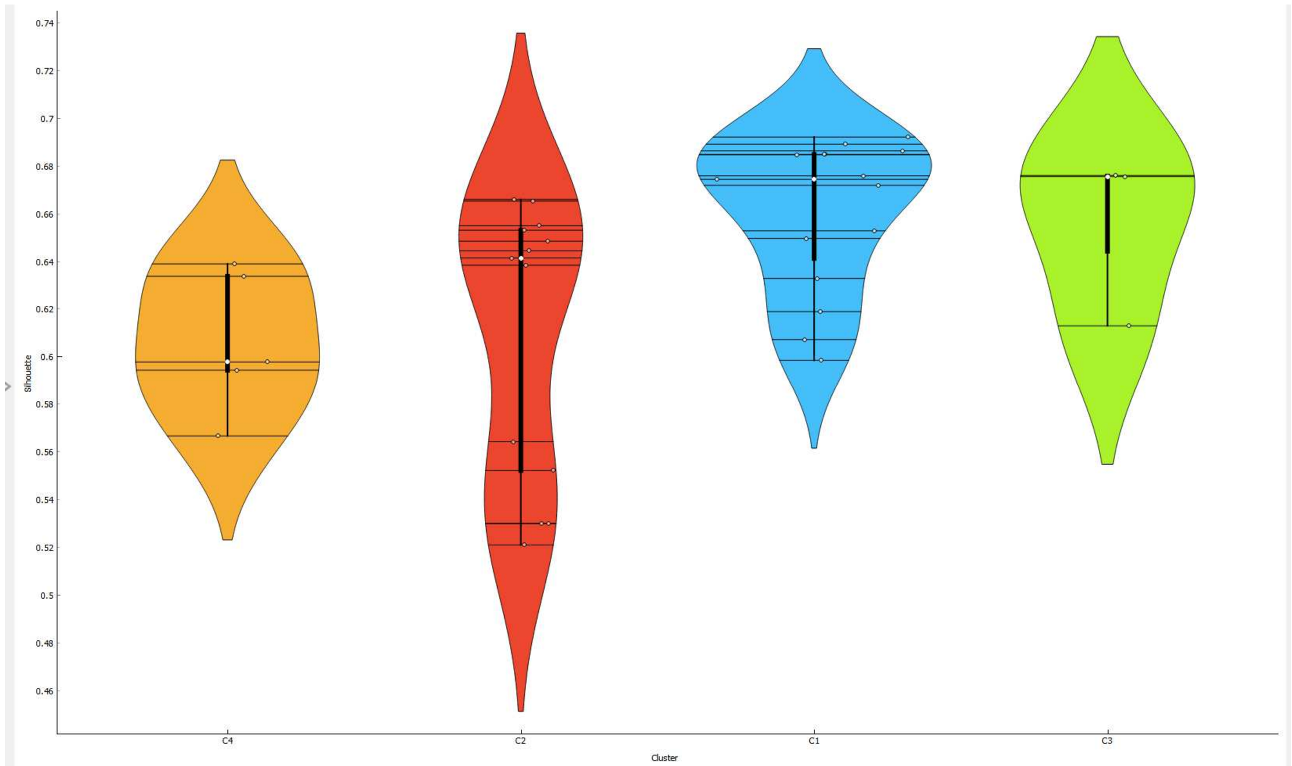

t-SNE
Silhouette Plot
Violin Plot
Distributions
Radviz
Scatter Plot
FreeViz
Linear Projection
Line Plot
Bar Plot
k-Means
File
Data


C1
C2
C3
C4
Belgium
United Kingdom
Iceland
Austria
Estonia
Finland
Greece
Norway
Netherlands
Sweden
Lithuania
Cyprus
Czechia
Germany
Switzerland
Slovakia
Turkey
Serbia
North Macedonia
Spain
Italy
Hungary
Latvia
Malta
Poland
Bulgaria
Romania
Ukraine
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350

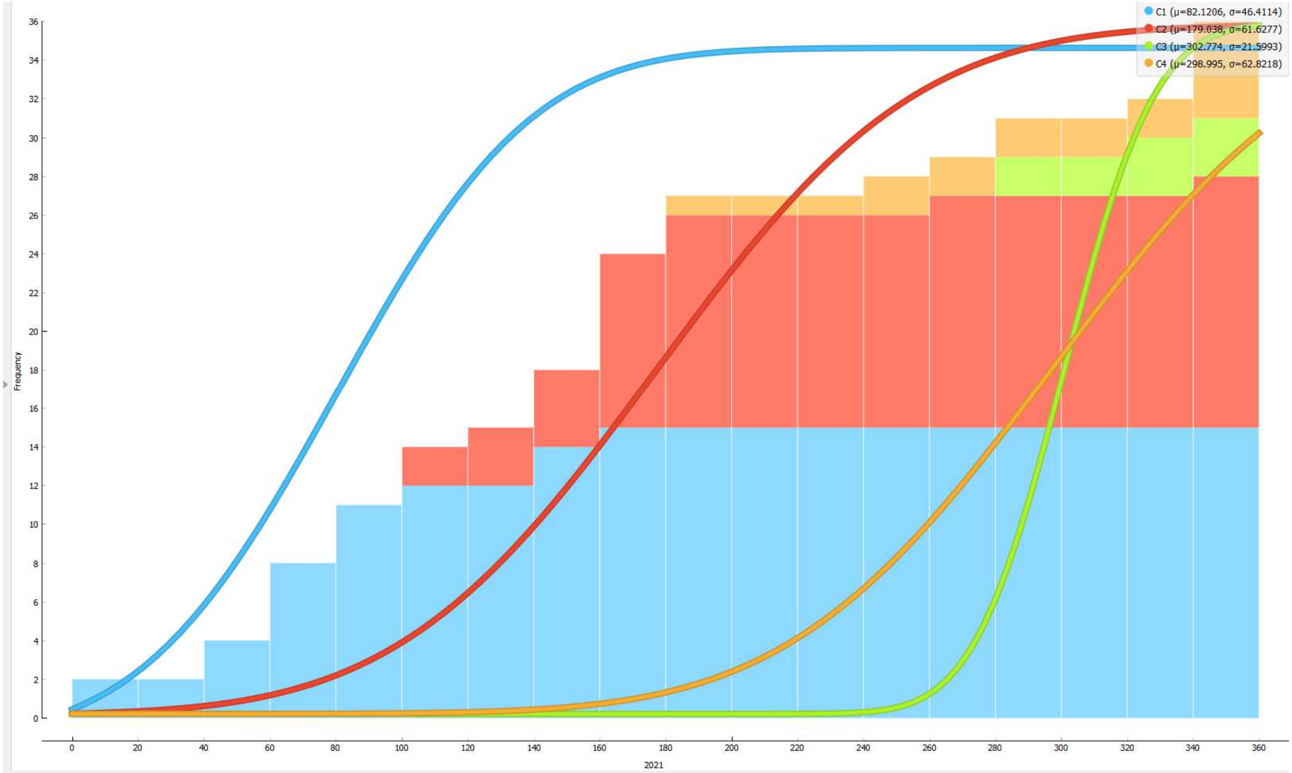

k-Means - Orange
Number of Clusters
Fixed: 3
From 2 to 11
Preprocessing
Normalize columns
Initialization
Initialize with KMeans++
Re-runs: 10
Maximum iterations: 300
Apply Automatically
36
36 | 4
Silhouette Scores
2 0.509
3 0.491
4 0.457
5 0.449
6 0.473
7 0.424
8 0.406
9 0.340
10 0.359
11 0.313

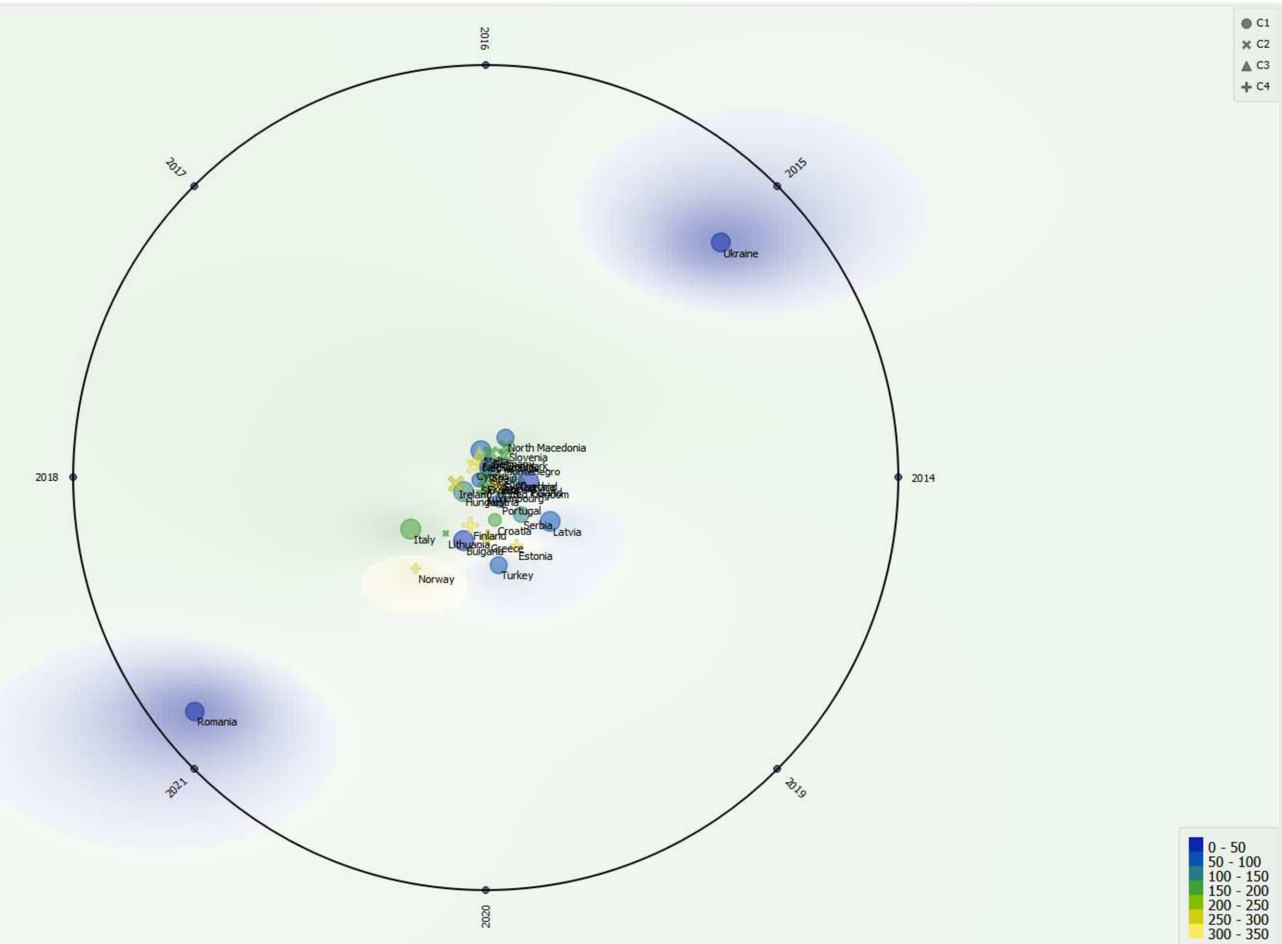


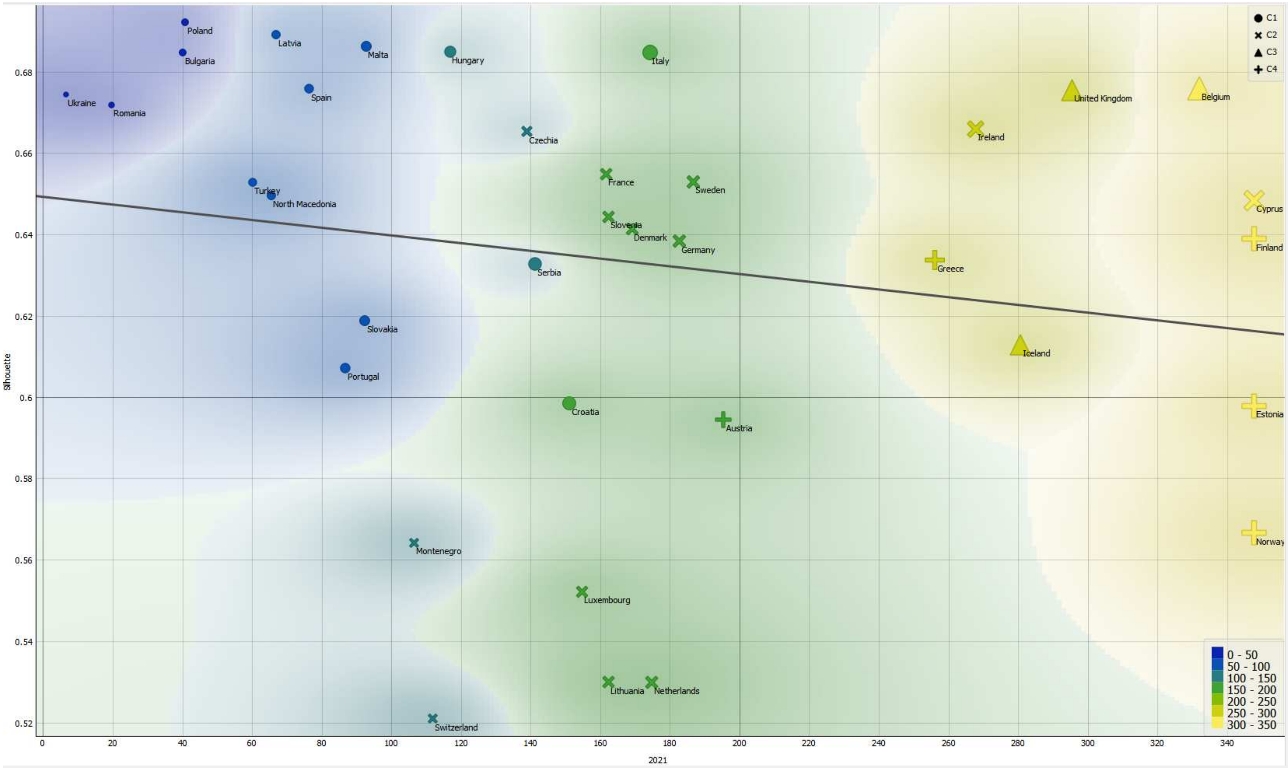
0
0.1
0.2
0.3
0.4
0.5
0.6
C1 (15)
Poland
Latvia
Malta
Bulgaria
Hungary
Italy
Spain
Ukraine
Romania
North Macedoni
Turkey
Serbia
Slovakia
Portugal
Croatia
C2 (13)
Ireland
Czechia
France
Sweden
Slovenia
Cyprus
Denmark
Germany
Montenegro
Luxembourg
Netherlands
Lithuania
Switzerland
C3 (3)
Belgium
United Kingdom
Iceland
C4 (5)
Finland
Greece
Estonia
Norway
Austria



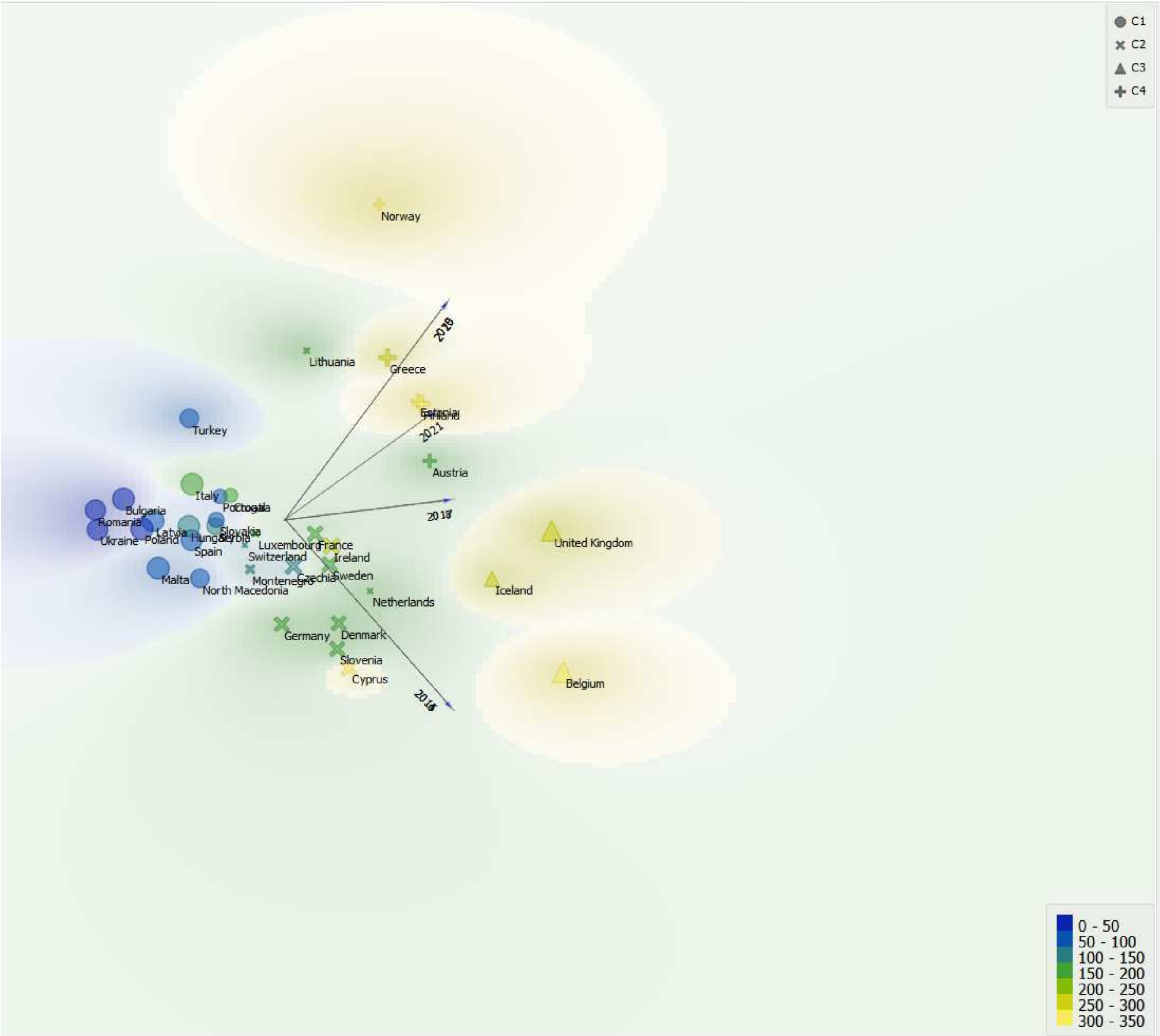
Silhouette
Cluster
C4
C2
C1
C3

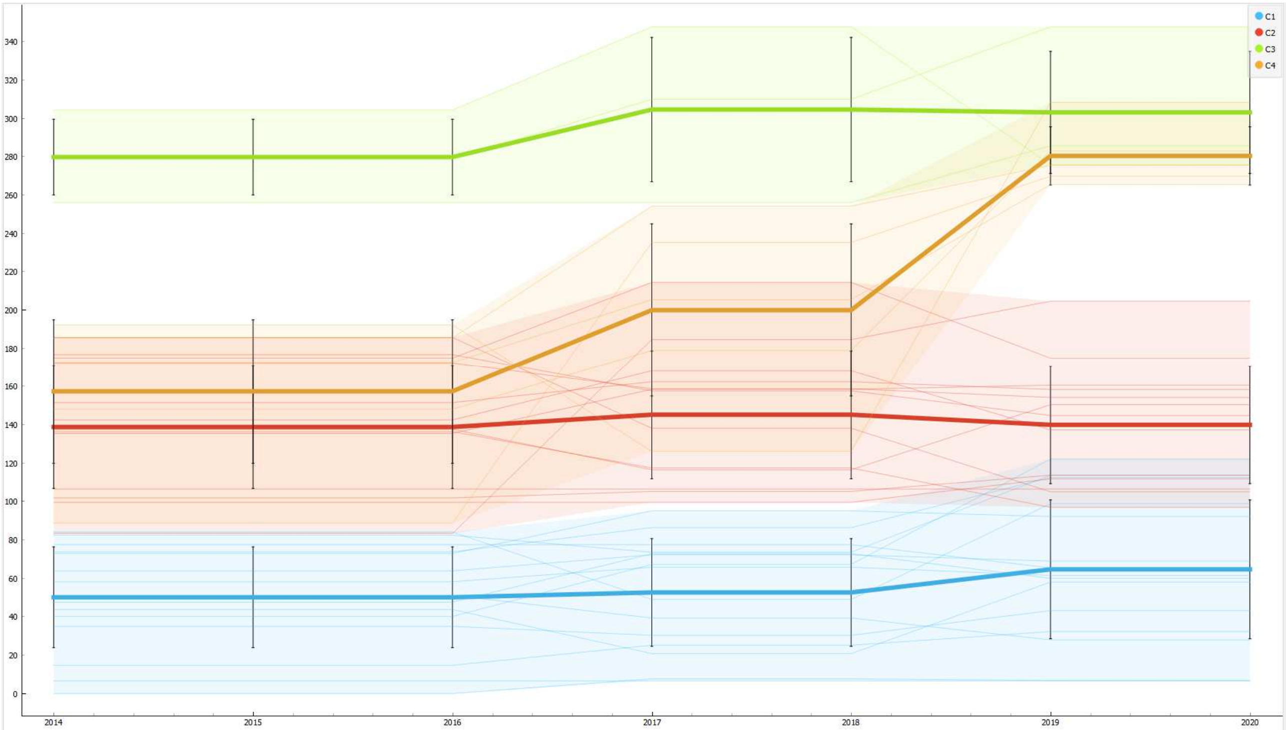

C1 (μ=82.1206, σ=46.4114)
C2 (μ=179.038, σ=61.6277)
C3 (μ=302.774, σ=21.5993)
C4 (μ=298.995, σ=62.8218)
Frequency
2021

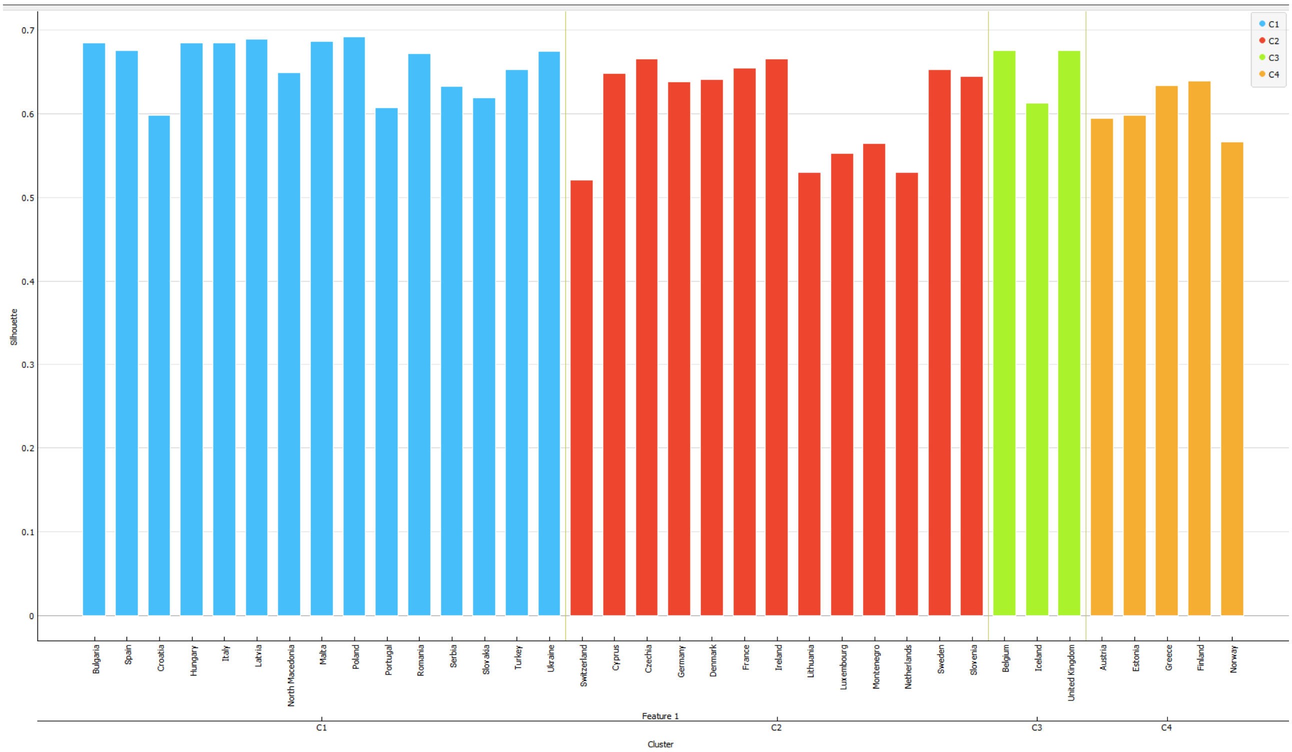

C1
C2
C3
C4
2016
2015
2014
2019
2020
2021
2018
2017
Ukraine
Romania
North Macedonia
Slovenia
Ireland
Hungary
Portugal
Italy
Croatia
Serbia
Latvia
Finland
Lithuania
Bulgaria
Greece
Estonia
Norway
Turkey
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350

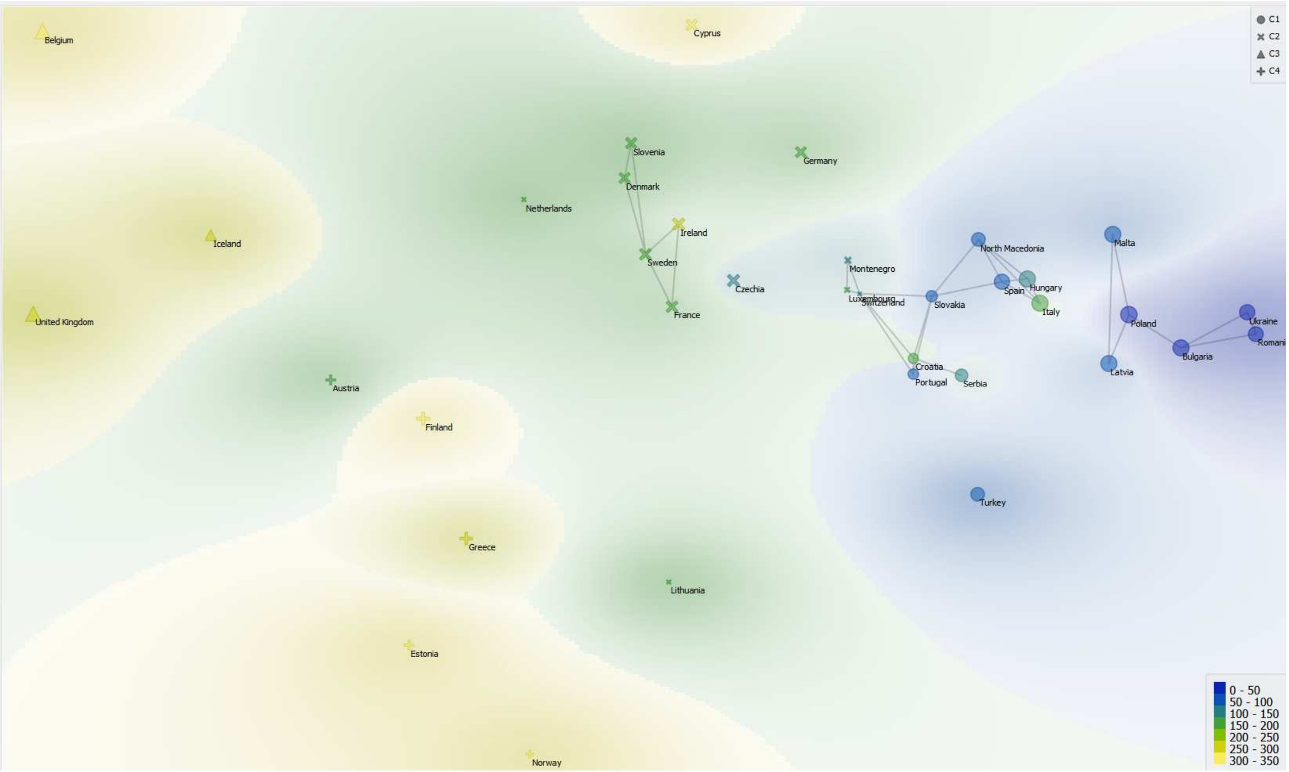


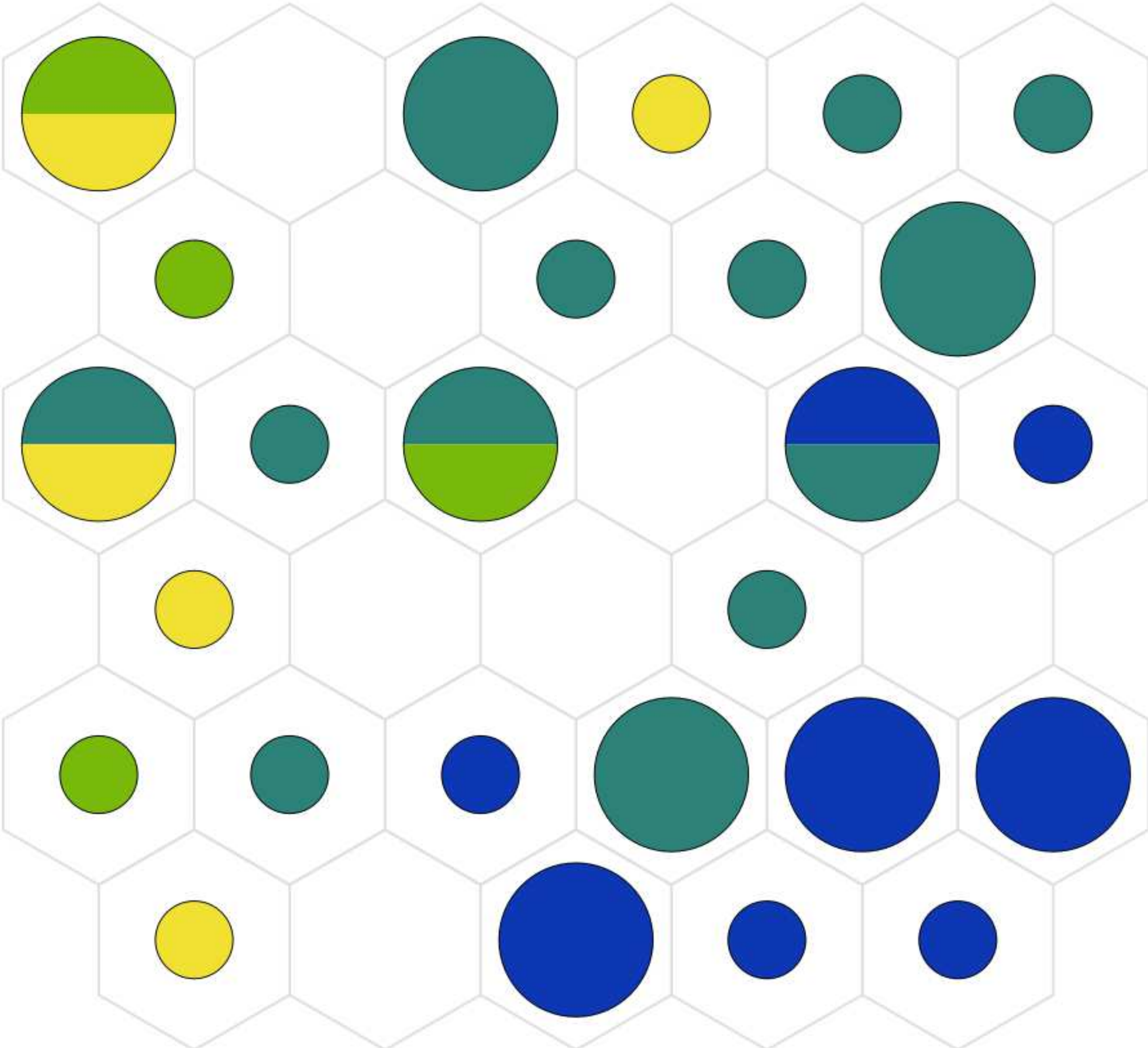
C1
C2
C3
C4
Poland
Latvia
Malta
Hungary
Italy
Bulgaria
Ukraine
Romania
Spain
United Kingdom
Belgium
Ireland
Czechia
France
Sweden
Turkey
North Macedonia
Cyprus
Slovenia
Denmark
Germany
Finland
Serbia
Greece
Slovakia
Iceland
Portugal
Croatia
Austria
Estonia
Montenegro
Norway
Luxembourg
Lithuania
Netherlands
Switzerland
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350
Silhouette
2021


C1
C2
C3
C4
Norway
Lithuania
Greece
Turkey
Austria
2021
Italy
Bulgaria
Romania
Ukraine
Poland
Latvia
Spain
Slovakia
Luxembourg
France
Switzerland
Ireland
Czechia
Sweden
Montenegro
Malta
North Macedonia
Netherlands
United Kingdom
Iceland
Germany
Denmark
Slovenia
Cyprus
Belgium
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350


C1
C2
C3
C4
2014
2015
2016
2017
2018
2019
2020

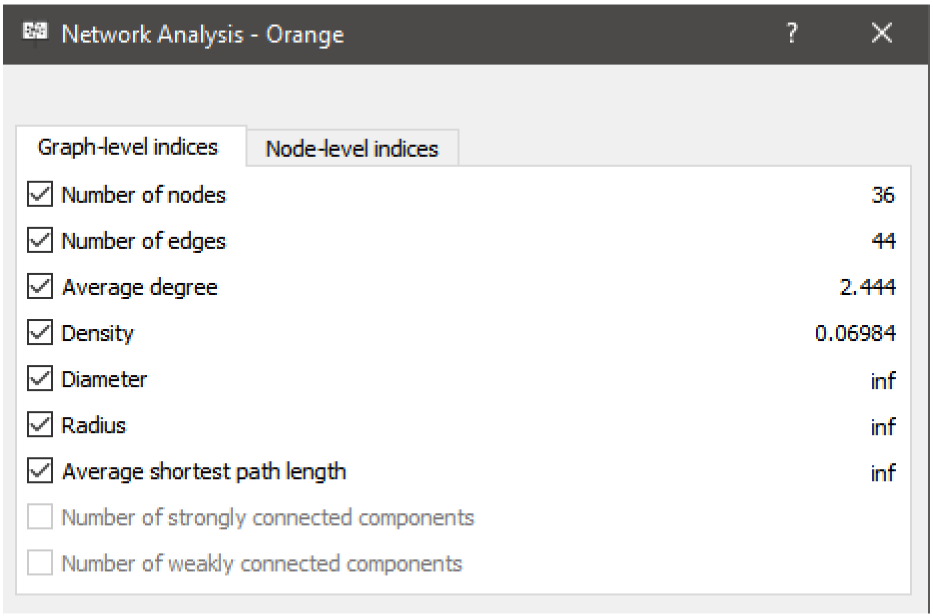

C1
C2
C3
C4
Silhouette
Bulgaria
Spain
Croatia
Hungary
Italy
Latvia
North Macedonia
Malta
Poland
Portugal
Romania
Serbia
Slovakia
Turkey
Ukraine
Switzerland
Cyprus
Czechia
Germany
Denmark
France
Ireland
Lithuania
Luxembourg
Montenegro
Netherlands
Sweden
Slovenia
Belgium
Iceland
United Kingdom
Austria
Estonia
Greece
Finland
Norway
Feature 1
Cluster

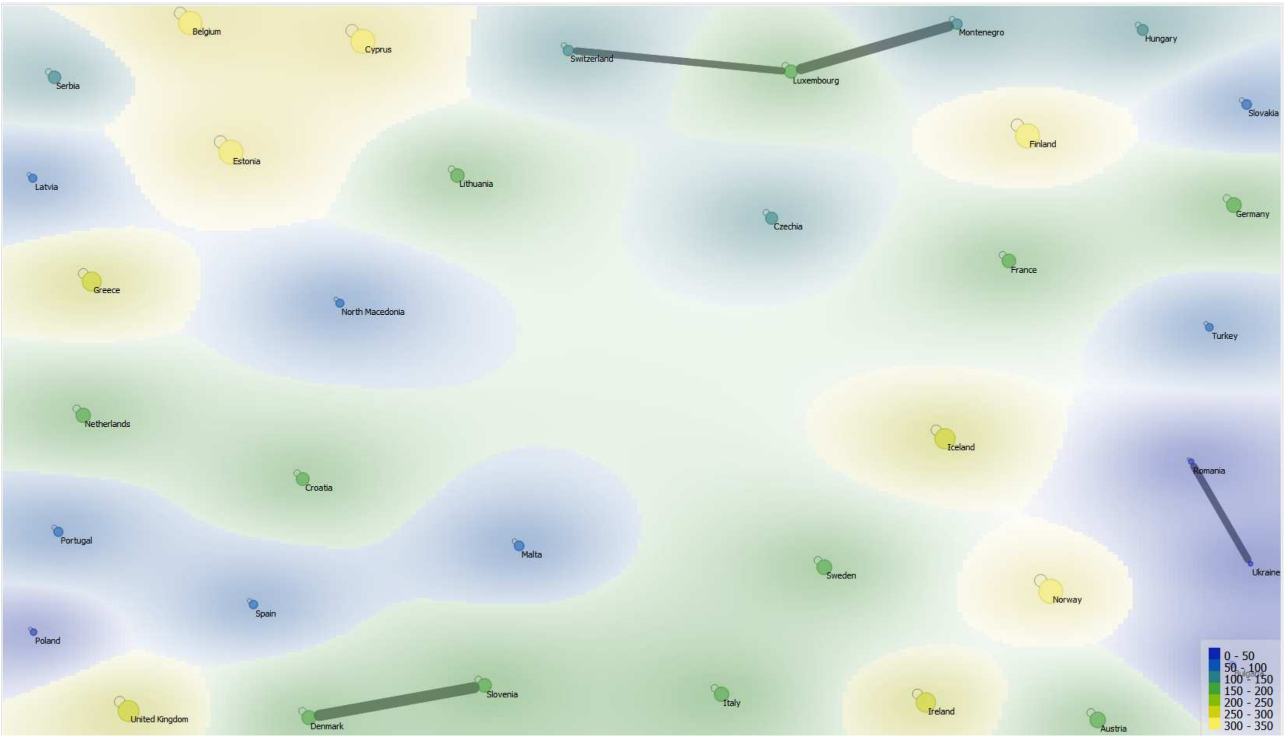

C1
C2
C3
C4
Belgium
Cyprus
Slovenia
Germany
Denmark
Netherlands
Ireland
Iceland
Sweden
Czechia
Montenegro
North Macedonia
Malta
Hungary
Spain
Luxembourg
Switzerland
Slovakia
Italy
Poland
Ukraine
Romani
United Kingdom
France
Bulgaria
Croatia
Portugal
Serbia
Latvia
Austria
Finland
Turkey
Greece
Lithuania
Estonia
Norway
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350

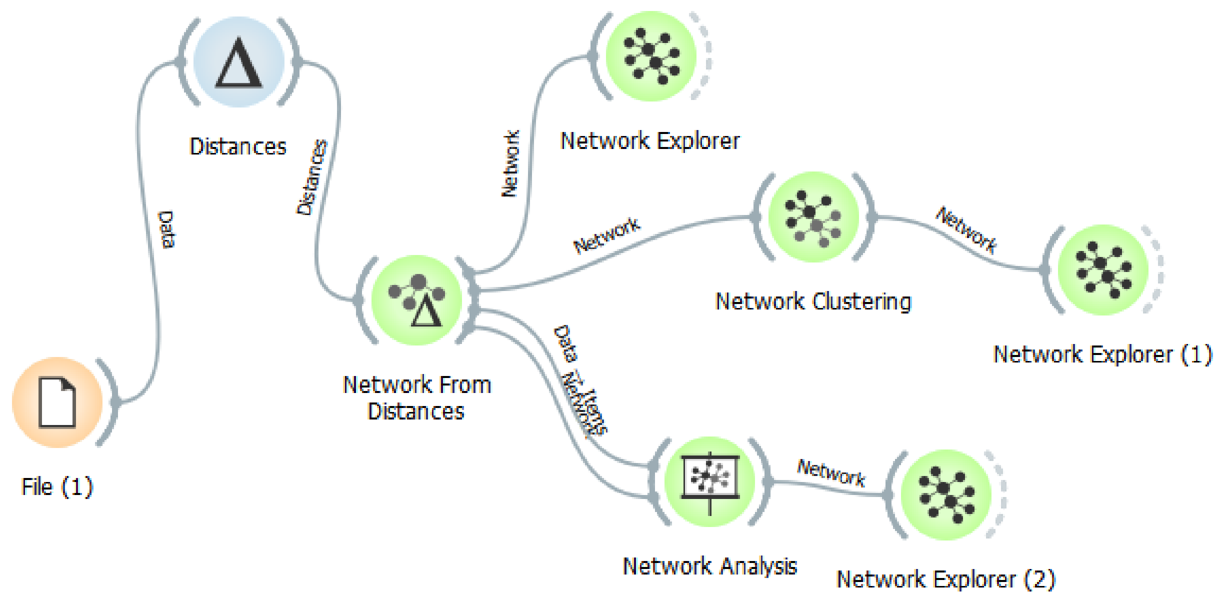


	2021	Feature 1	Cluster	Silhouette
3	40.1326	Bulgaria	C1	0.684847
11	76.3975	Spain	C1	0.675903
14	151.156	Croatia	C1	0.59856
15	116.921	Hungary	C1	0.684899
18	174.263	Italy	C1	0.68473
21	66.93	Latvia	C1	0.689188
23	65.4546	North Macedonia	C1	0.64968
24	92.7024	Malta	C1	0.68639
27	40.8346	Poland	C1	0.69217
28	86.8025	Portugal	C1	0.607272
29	19.6842	Romania	C1	0.671941
30	141.351	Serbia	C1	0.632904
33	92.45	Slovakia	C1	0.618988
34	60.1333	Turkey	C1	0.652876
35	6.59656	Ukraine	C1	0.674431
4	111.799	Switzerland	C2	0.521024
5	347.832	Cyprus	C2	0.648495
6	139.009	Czechia	C2	0.665333
7	182.658	Germany	C2	0.638446
8	169.124	Denmark	C2	0.641465
13	161.656	France	C2	0.654983
16	267.71	Ireland	C2	0.666037
19	162.404	Lithuania	C2	0.53001
20	154.751	Luxembourg	C2	0.552227
22	106.482	Montenegro	C2	0.564311
25	174.832	Netherlands	C2	0.530006
31	186.761	Sweden	C2	0.653108
32	162.477	Slovenia	C2	0.644461
2	332.092	Belgium	C3	0.676062
17	280.691	Iceland	C3	0.613039
36	295.538	United Kingdom	C3	0.675583
1	195.368	Austria	C4	0.594442
9	347.832	Estonia	C4	0.597882
10	256.109	Greece	C4	0.633804
12	347.832	Finland	C4	0.638998
26	347.832	Norway	C4	0.566688

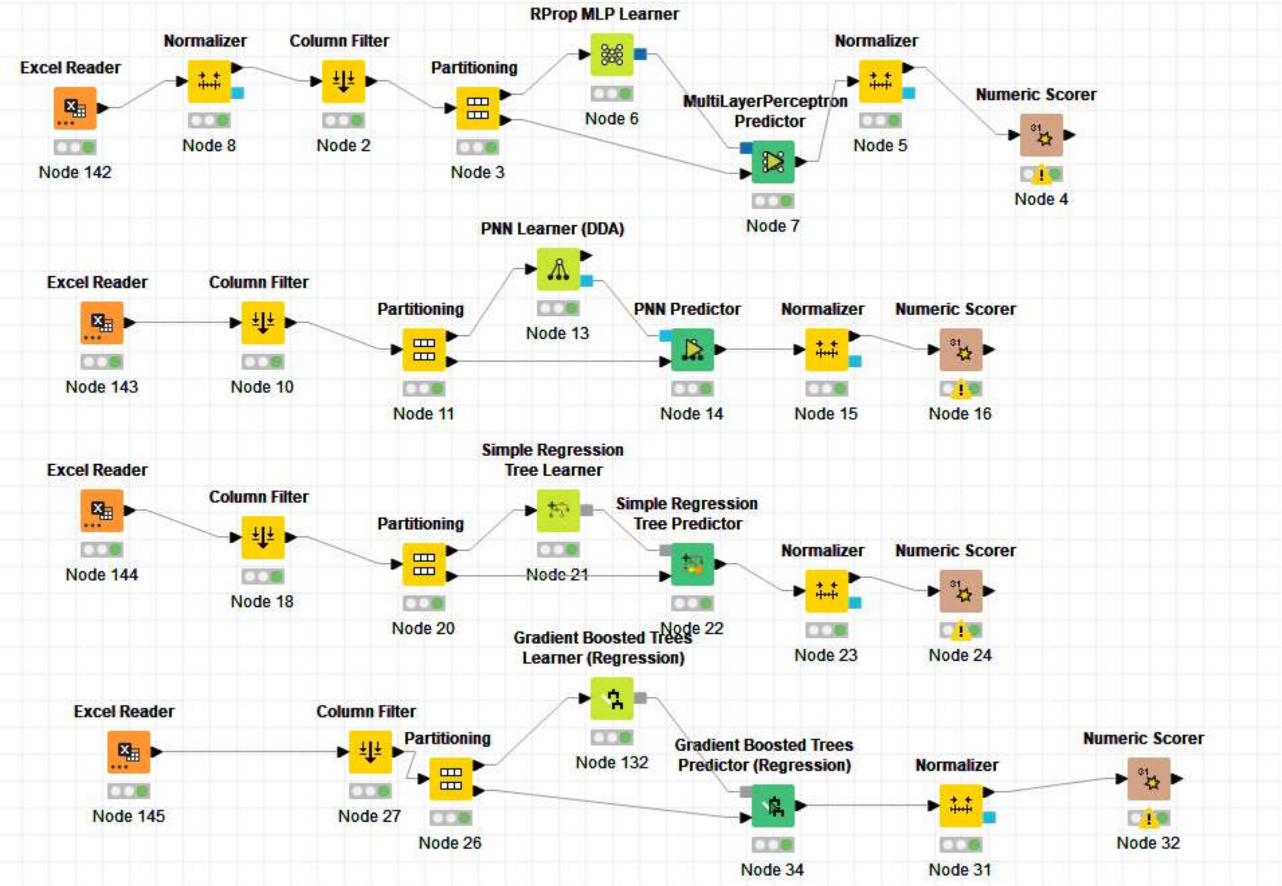
Network Analysis

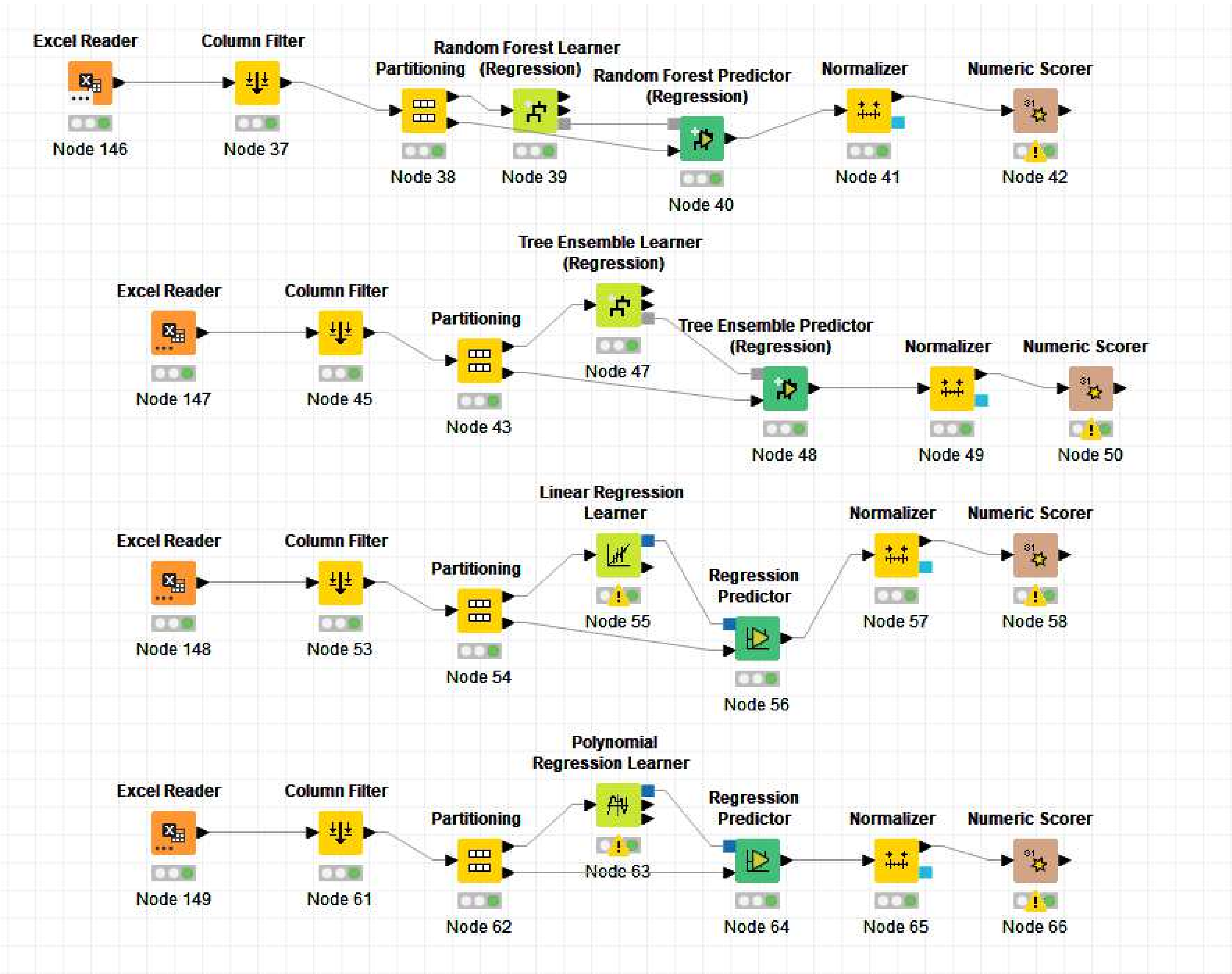

Network Analysis - Orange
Graph-level indices
Node-level indices
Number of nodes 36
Number of edges 44
Average degree 2.444
Density 0.06984
Diameter inf
Radius inf
Average shortest path length inf
Number of strongly connected components
Number of weakly connected components


Belgium
Cyprus
Switzerland
Luxembourg
Montenegro
Hungary
Serbia
Slovakia
Finland
Estonia
Latvia
Lithuania
Czechia
Germany
France
Greece
North Macedonia
Turkey
Netherlands
Iceland
Croatia
Romania
Portugal
Malta
Ukraine
Sweden
Norway
Spain
Poland
Slovenia
Italy
Ireland
United Kingdom
Denmark
Austria
0 - 50
50 - 100
100 - 150
150 - 200
200 - 250
250 - 300
300 - 350



Machine Learning and Predictions


Excel Reader
Node 142
Normalizer
Node 8
Column Filter
Node 2
Partitioning
Node 3
RProp MLP Learner
Node 6
MultiLayerPerceptron Predictor
Node 7
Normalizer
Node 5
Numeric Scorer
Node 4
Excel Reader
Node 143
Column Filter
Node 10
Partitioning
Node 11
PNN Learner (DDA)
Node 13
PNN Predictor
Node 14
Normalizer
Node 15
Numeric Scorer
Node 16
Excel Reader
Node 144
Column Filter
Node 18
Partitioning
Node 20
Simple Regression Tree Learner
Node 21
Simple Regression Tree Predictor
Node 22
Normalizer
Node 23
Numeric Scorer
Node 24
Excel Reader
Node 145
Column Filter
Node 27
Partitioning
Node 26
Gradient Boosted Trees Learner (Regression)
Node 132
Gradient Boosted Trees Predictor (Regression)
Node 34
Normalizer
Node 31
Numeric Scorer
Node 32



	Statistical results of the machine learning algorithms used for the prediction with original data			
	ANN	**PNN**	**Simple Regression Tree**	**Gradient Boosted Trees Regression**
R^2	0,128886177	-0,919301063	-0,320310002	-0,656173488
mean absolute error	0,242103293	0,29537217	0,217223243	0,297576378
mean squared error	0,104386709	0,143026775	0,096160247	0,124674364
root mean squared error	0,323089321	0,378188808	0,310097157	0,353092572
	Random Forest Regression	**Tree Ensemble Regression**	**Linear Regression**	**Polynomial Regression**
R^2	0,033274166	-0,454346323	-0,905431621	-0,583239856
mean absolute error	0,186156067	0,258031995	0,288359552	0,270084841
mean squared error	0,074290063	0,111901106	0,122678859	0,145957411
root mean squared error	0,272562035	0,334516226	0,35025542	0,38204373

Algorithms	R^2	Mean absolute error	Mean squared error	Root mean squared error	Sum
Random Forest Regression	2	1	1	1	5
Simple Regression Tree	3	2	2	2	9
ANN	1	3	3	3	10
Tree Ensemble Regression	4	4	4	4	16
Linear Regression	7	6	5	5	23
Polynomial Regression	5	5	8	8	26
Gradient Boosted Trees Regression	6	8	6	6	26
PNN	8	7	7	7	29

Prediction with the original data using the Random Forest Regression algorithm					
Country	**2021**	**Prediction**	**Country**	**2021**	**Prediction**
Austria	195,368	327,021	*Lithuania*	162,404	173,456
Belgium	332,092	279,495	*Luxembourg*	154,751	132,809
Bulgaria	40,133	36,722	*Latvia*	66,930	103,551
Switzerland	111,799	135,087	*Montenegro*	106,482	125,950
Cyprus	347,832	192,342	*North Macedonia*	65,455	94,065
Czechia	139,009	220,356	*Malta*	92,702	110,631
Germany	182,658	154,690	*Netherlands*	174,832	215,741
Denmark	169,124	207,744	*Norway*	347,832	195,897
Estonia	347,832	292,699	*Poland*	40,835	67,281
Greece	256,109	251,547	*Portugal*	86,802	94,565
Spain	76,398	106,621	*Romania*	19,684	21,288
Finland	347,832	188,446	*Serbia*	141,351	111,392
France	161,656	171,553	*Sweden*	186,761	177,784
Croatia	151,156	121,987	*Slovenia*	162,477	229,854
Hungary	116,921	96,132	*Slovakia*	92,450	95,245
Ireland	267,710	172,953	*Turkey*	60,133	88,984
Iceland	280,691	294,423	*Ukraine*	6,597	31,591
Italy	174,263	83,761	*United Kingdom*	295,538	294,525

Augmented Data

	Statistical results of the algorithms used for prediction through the use of augmented data			
	ANN	**PNN**	**Simple Regression Tree**	**Gradient Boosted Tree Regression**
R^2	0,811910721	0,683748646	0,524600593	0,774358703
mean absolute error	0,112803213	0,114650962	0,154343354	0,107745554
mean squared error	0,019859537	0,026396318	0,036879259	0,027459858
root mean squared error	0,14092387	0,162469437	0,192039732	0,165710162
	Random Forest Regression	**Tree Ensemble Regression**	**Linear Regression**	**Polynomial Regression**
R^2	0,918671301	0,895135143	0,914747902	0,840260767
mean absolute error	0,054200754	0,066683868	0,065663828	0,114509614
mean squared error	0,00642922	0,009137328	0,007166789	0,017133227
root mean squared error	0,080182414	0,095589372	0,084656893	0,130893953

Ranking of machine learning algorithms by predictive ability					
Algorithms	**R^2**	**mean absolute error**	**mean squared error**	**root mean squared error**	**Sum**
Random Forest Regression	1	1	1	1	4
Linear Regression	2	2	2	2	8
Tree Ensemble Regression	3	3	3	3	12
Polynomial Regression	4	6	4	4	18
ANN	5	5	5	5	20
Gradient Boosted Tree Regression	6	4	7	7	24
PNN	7	7	6	6	26
Simple Regression Tree	8	8	8	8	32

Prediction using augmented data obtained using the Random Forest Regression algorithm				
Country	**Prediction 2021**	**Prediction Augmented Data**	**Absolute Variation**	**Percentage Variation**
Ciprus	192,342	227,778	35,436	18,423
Finland	188,446	254,688	66,242	35,152
France	171,553	181,084	9,531	5,556
Italy	83,761	105,555	21,794	26,019
Lithuania	173,456	157,855	-15,601	-8,994
Norway	195,897	184,822	-11,075	-5,653
Romania	21,288	54,363	33,075	155,369
Slovenia	229,854	216,674	-13,18	-5,734
Slovakia	95,245	106,423	11,178	11,736
Ukraine	31,591	54,363	22,772	72,084
United Kingdom	294,525	289,262	-5,263	-1,787
Mean	152,542	166,624	14,083	9,232

Comparison between statistical errors and R-squared between values obtained by prediction with original data and prediction with augmented data				
Statistical Metrics	**Original Data**	**Augmented Data**	**Absolute Variation**	**Percentage Variation**
	Random Forest Regression	**Random Forest Regression**		
R^2	0,033274166	0,918671301	0,885	2660,915
Mean absolute error	0,186156067	0,054200754	-0,132	-70,884
Mean squared error	0,074290063	0,00642922	-0,068	-91,346
Root mean squared error	0,272562035	0,080182414	-0,192	-70,582
Mean of Errors	0,177669388	0,046937463	-0,131	-73,582