Life Expectancy at Birth in the Italian Regions

LEOGRANDE, ANGELO

Lum University Giuseppe Degennaro

2022

Life Expectancy at Birth in the Italian Regions

Between 2019 and 2020 it decreased by 0.87 years

Istat calculates the value of life expectancy at birth in the Italian regions. Life expectancy at birth is defined as "[…] *the average number of years a child born in a given calendar year can expect to live*".

Ranking of Italian regions by life expectancy at birth in the Italian regions in 2020. Umbria ranks first in terms of life expectancy at birth in 2020 with a value of 83.8 years, followed by Veneto and Lazio with a value of 83.1. Friuli-Venezia Giulia is in third place with an amount equal to 83 years. In the middle of the table there are Emilia-Romagna, Abruzzo and Basilicata with an amount equal to 82.4 years. Campania with a value equal to 81.4 years, Lombardy with a value equal to 81.2 years and Valle d'Aosta with an amount equal to 80.9 years close the ranking. On average in 2020, life expectancy at birth in the Italian regions was 82,345 years.

Ranking of the Italian regions by percentage change in life expectancy at birth. Lazio ranks first in terms of the percentage change in life expectancy at birth between 2004 and 2020 with a value equal to 3.74% equal to 3 years, followed by Friuli-Venezia Giulia with an amount equal to 3, 1% equal to 2.5 years and from Umbria with 2.82% equal to 2.3 years. In the middle of the table there are Calabria with an amount equal to 2.230% equal to 1.8 years, followed by Trentino-Alto Adige with an amount equal to 2.096% equal to 1.7 years, and by Tuscany with a value equal to 1 , 84% equal to 1.5 years. Puglia closes the ranking with a value of 1.23% equal to an amount of 1 year, followed by the Marche with a value equal to 0.978% equal to an amount of 0.8 years and by Lombardy with a value of 0.37 % equal to an amount of 0.3 years.

Italian Macro-Regions. The value of life expectancy at birth in the North has had a fluctuating trend between 2004 and 2020. In fact, if we consider the period between 2004 and 2020 in the North, life expectancy at birth has increased by 1.1 years equal to 1.3%. However, between 2015 and 2020 the value of life expectancy at birth decreased in the North by 0.7 years equal to a value of -0.84%. In Central Italy, life expectancy at birth increased between 2004 and 2020 by a value of 2.2 years or equal to 2.71%, a growth rate higher than that of the 2015-2020 subperiod equal to a value 0.5 years equal to a value of 0.6%. In the South, the growth in life expectancy between 2004 and 2020 grew by a value equal to 2 years or equal to a value of 2.49% higher than in the subperiod between 2015 and 2020 with a value equal to 0,6 years equal to a value of 0.7%. On average, also for Italy as a whole, the value of the absolute change in life expectancy at birth decreased from 1.6 years in the period between 2004 and 2020 to a value of 0 years in the period between 2015 and 2020. However, life expectancy decreased significantly between 2019 and 2020 in all Italian macro regions: North with a variation of -1.6 years equal to -1.9%; Center with a variation of -0.5 years equal to -0.59%; Mezzogiorno with an absolute change equal to -0.3 years equal to -0.36%; and in Italy the value decreased by a value of -0.9 years or equal to a value of -1.08%. It therefore follows that the value of life expectancy which had reached a peak in the Italian macro-regions in 2019 has undergone a significant reduction in 2020 in connection with the pandemic events of Covid-19.

Clustering with k-Means algorithm. The k-Means algorithm is applied below for the classification of Italian regions based on life expectancy at birth. The Silhouette coefficient is used to select the optimal number of clusters. The Silhouette coefficient varies between -1 and +1. The optimal number of clusters is generated in connection with the highest Silhouette coefficient value. The value of the highest Silhouette coefficient is equal to 0.533 units in connection with several clusters equal to 3. The clusters are made up as follows:

- Cluster 1: Piedmont, Liguria, Molise, Sardinia, Lazio, Basilicata, Calabria, Friuli Venezia Giulia, Abruzzo, Puglia, Aosta Valley;
- Cluster 2: Veneto, Tuscany, Umbria, Emilia-Romagna, Trentino-Alto Adige, Marche, Lombardy;
- Cluster 3: Campania, Sicily.

Therefore, if we take into consideration the value of the median of the clusters then we obtain the following ordering C2 = 82.8> C1 = 82.4> C3 = 81.7. The clusters map shows that cluster 2 substantially coincides with the center-north, cluster 1 coincides with the south and cluster 3 concerns only two regions, namely Campania and Sicily. The analysis therefore shows that life expectancy tends to grow with income. Furthermore, the data show the presence of the phenomenon of Central Italy which has very high life expectancy values compared to other regions.

Network Analysis. Network analysis is used to trace the relationships between regions or to verify the presence of specific connections between indicated areas. The analysis is carried out using the Euclidean distance. The result shows the presence of strong and significant links at the network level between three different regions, namely:

- Veneto and Emilia-Romagna with a value of 0.51 units;
- Veneto and Tuscany with a value of 0.51 units

Veneto appears to be an intermediate node between Emilia-Romagna and Tuscany. Finally, the analysis of the metric characteristics of the network analysis shows the following elements, namely:

- Number of nodes with a value of 20;
- Number of edges with a value of 24;
- Average degree with a value of 2.4.
- Density with a value of 0.1263.

Machine learning and predictions. Eight different machine learning algorithms are applied below for the prediction of the value of life expectancy at birth in the Italian regions. The algorithms were trained using 70% of the data and the prediction was made with the remaining 30% of the data. The algorithms were compared through their ability to maximize the R-square and minimize the following statistical errors, namely: "Mean Absolute Error", "Mean Squared Error", "Root Mean Squared Error". The analysis shows the following ordering of the algorithms, namely:

- Gradient Boosted Tree Regression with a payoff value of 4;
- Simple Regression Tree with a payoff value of 4;
- ANN-Artificial Neural Network and Tree Ensemble Regression with a payoff value of 14;
- Random Forest Regression with a payoff value of 22;
- PNN-Probabilistic Neural Network with a payoff value of 23;
- Polynomial Regression with a payoff value of 27;
- Linear Regression with a payoff value of 32.

Based on the application of the best performer algorithm which is the Gradient Boosted Tree Regression, the following values are predicted, namely:

- Valle d'Aosta with an increase from an amount of 80.90 up to a value of 82.03 units or equal to an amount of 1.13 units equal to a value of 1.40%;
- Emilia-Romagna with an increase from an amount of 82.40 units up to a value of 83.10 units or equal to a variation of 0.70 units equal to a variation of 0.85%;

- Tuscany with an increase from an amount of 83.00 units up to a variation of 83.10 units or equal to a variation of 0.10 units equal to a variation of 0.12%;
- Campania with an increase from an amount of 81.40 units up to a value of 82.77 units or equal to a variation of 1.37 units equal to an amount of 1.68%;
- Puglia with an increase from an amount of 82.10 units up to a value of 83.79 or equal to a variation of 1.69 units equal to a variation of 2.06%;
- Sicily with a variation from 82.00 units up to a value of 82.77 units or equal to a variation of 0.77 units equal to an amount of 0.94%.

In summary, on average, for the regions analyzed, the growth in the value of life expectancy at birth is predicted to increase from a value of 81.97 to a value of 92.93 units or equal to an amount of 0.96 units equal to a value of 1.17%. However, it must be considered that this prediction, although methodologically correct, is still incorrect since the data does not consider either the persistence of the pandemic in 2021 or the Ukrainian crisis.

Conclusions. Life expectancy at birth in the Italian regions has grown from an amount of 80.72 years in 2004 to a value of 82,345 units in 2020 or a growth of 2%. However, it must be considered that between 2019 and 2020 the value of life expectancy at birth on average for the Italian regions decreased from 83,215 units to a value of 82,345 units in connection with Covid. The data show a generic positive relationship between income and life expectancy at birth and in particular highlight a dominance of the regions of Central Italy. Certainly, it is probable that also analyzing the data of 2021 and 2022 there is, contrary to what is predicted by the machine learning algorithms, still a reduction in the value of life expectancy at birth due to the covid 19 crisis also in conjunction with the Ukrainian crisis. The growth in life expectancy is therefore certainly to relate to the growth of per capita income, but also with macroeconomic elements such as health and financial crises that can reduce life expectancy at birth. In fact, if we take for example the change in life expectancy at birth between 2019 and 2020 we see that Lombardy has undergone a negative change of 2.5 years, Valle d'Aosta equal to 1.8 years and Marche equal at a value of 1.4 years. It therefore follows that the pandemic has substantially changed the structure of life expectancy in the Italian regions, also regardless of per capita income.

References

Laureti, L., Costantiello, A., & Leogrande, A. (2022). Satisfaction with the Environmental Condition in the Italian Regions between 2004 and 2020. University Library of Munich, Germany.

Leogrande, Angelo, Alberto Costantiello, Lucio Laureti, and Domenico Leogrande. The Determinants of Landscape and Cultural Heritage Among Italian Regions in the Period 2004-2019. University Library of Munich, Germany, 2021.

Magazzino, C., & Leogrande, A. (2021). Subjective Well-Being In Italian Regions: A Panel Data Approach. Applied Econometrics and International Development, 21(1), 1-18.

Angelo Leogrande. (2022). The Risk of Poverty in the Italian Regions. In Il Sud Est. Zenodo. https://doi.org/10.5281/zenodo.6194023

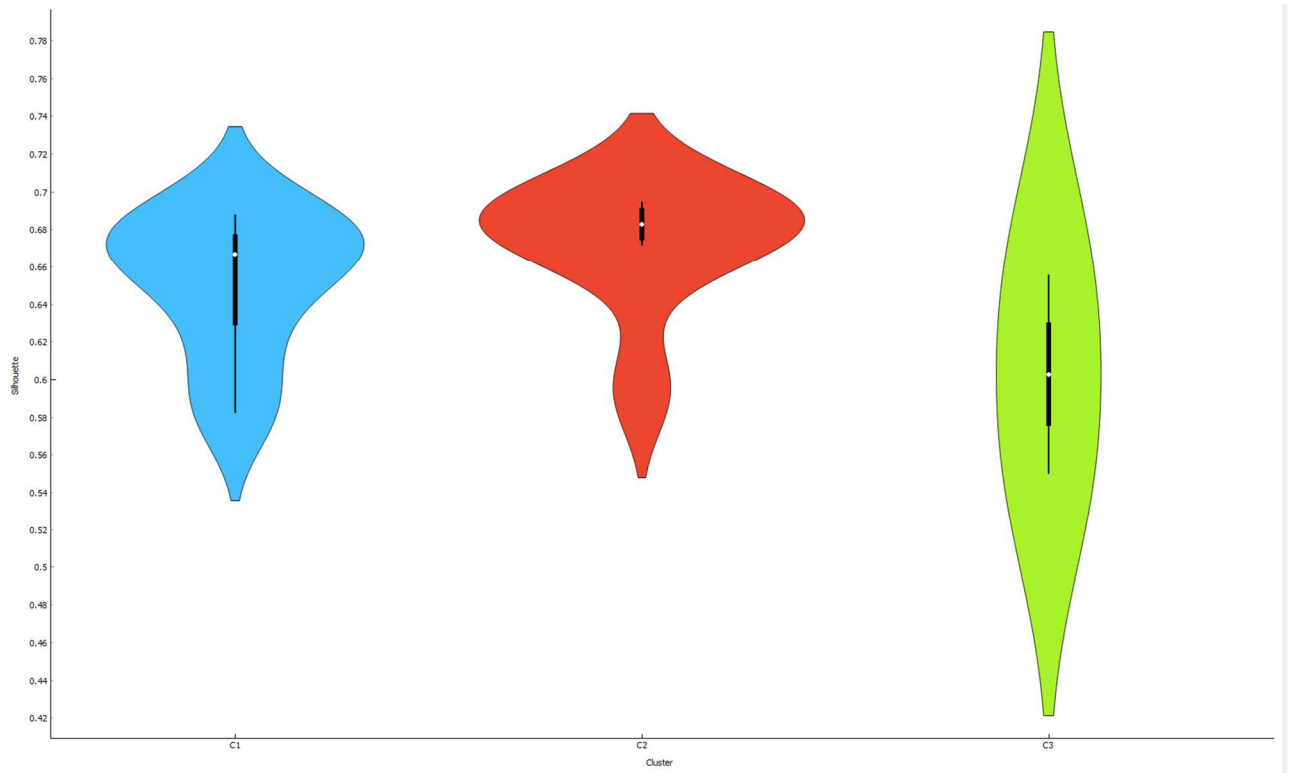
Angelo Leogrande. (2022). Social Participation in the Italian Regions. In Il Sud Est. Zenodo. https://doi.org/10.5281/zenodo.6212069

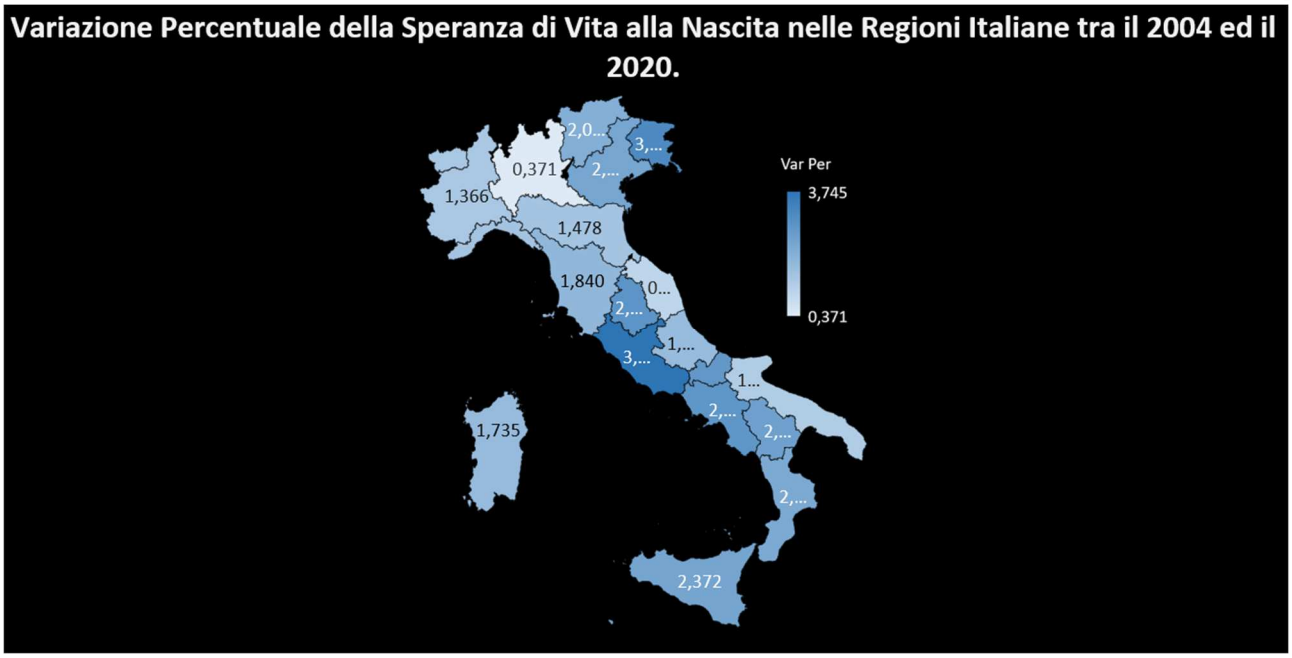
Angelo Leogrande. (2022). The Reduction of NEETs in the Italian Regions. In Il Sud Est. Zenodo. https://doi.org/10.5281/zenodo.6212291

Angelo Leogrande. (2022). The Employment Rate in the Italian Regions. In Il Sud Est. Zenodo. https://doi.org/10.5281/zenodo.6211793

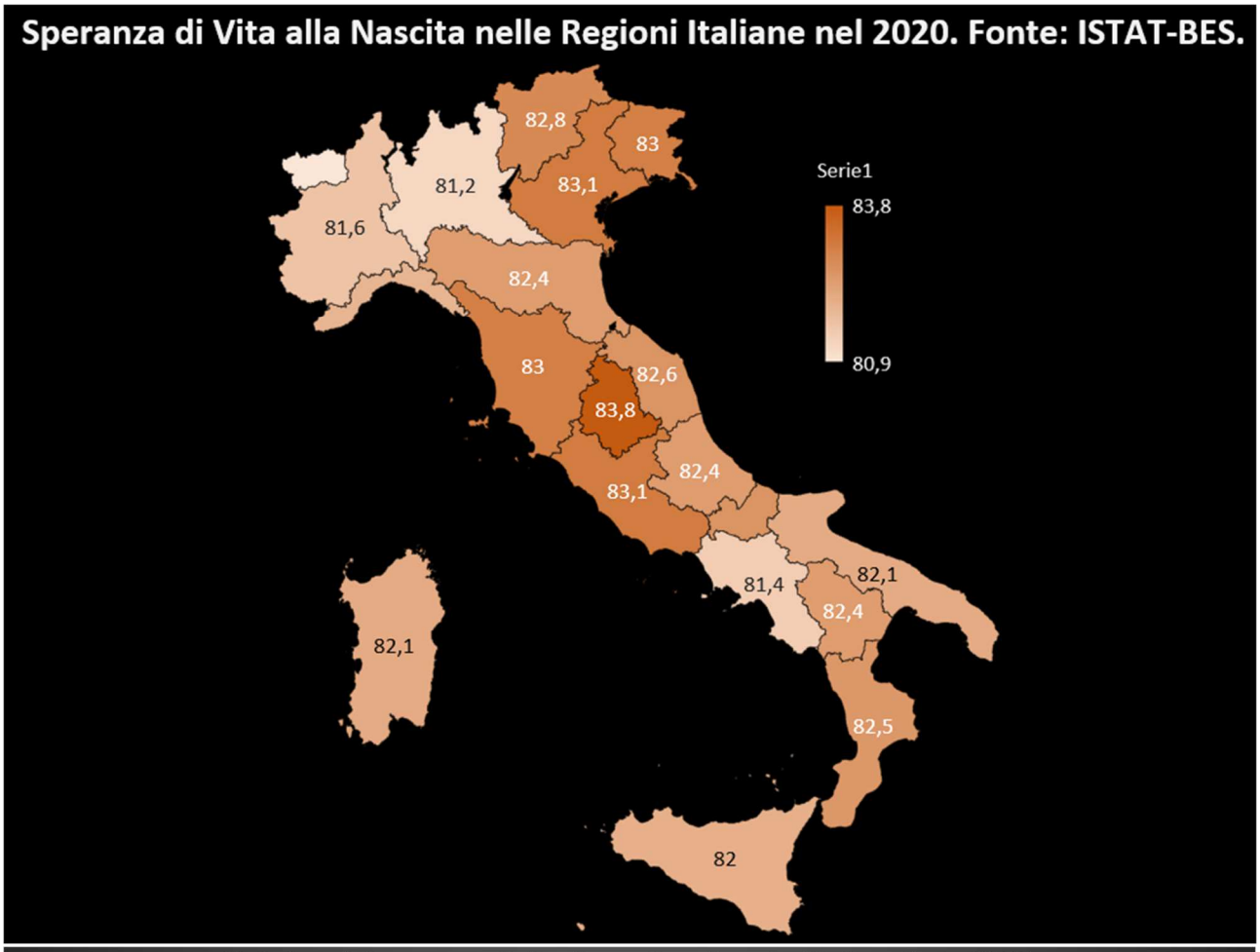
Angelo Leogrande. (2022). The Duration of Civil Proceedings in the Italian Regions. In Il Sud Est. Zenodo. https://doi.org/10.5281/zenodo.6188615

Angelo Leogrande. (2022). The Crowding of Prisons in the Italian Regions. Il Sud Est. https://doi.org/10.5281/zenodo.6395200

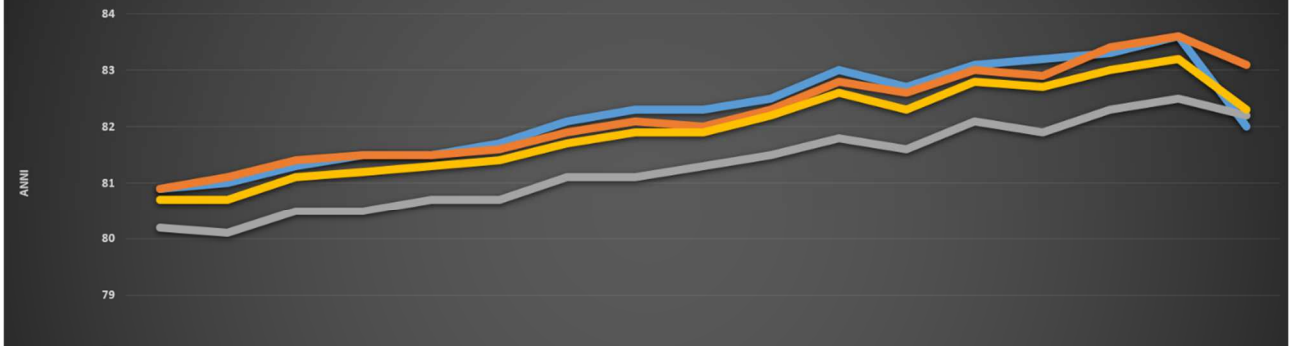




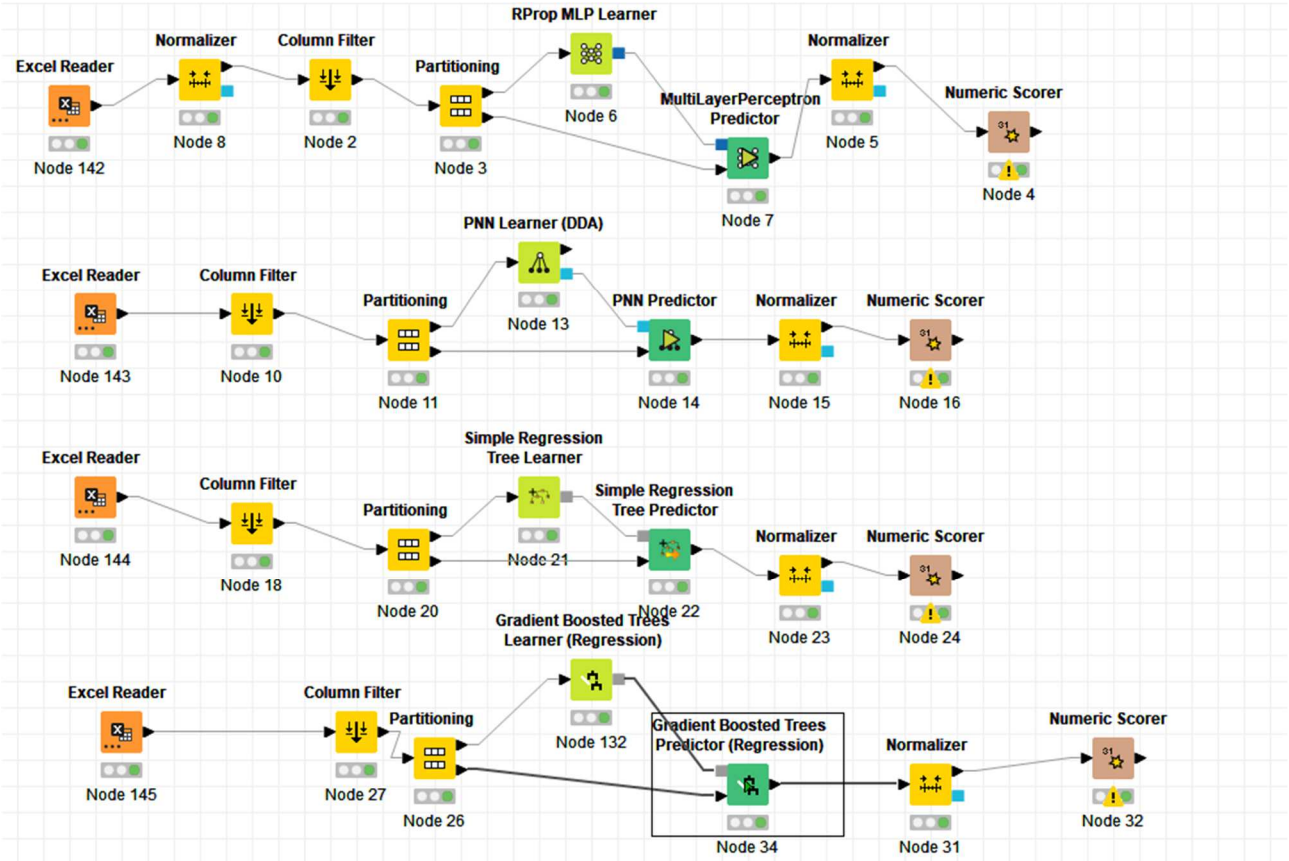
Speranza di Vita alla Nascita nelle Regioni Italiane nel 2020. Fonte: ISTAT-BES.



Variazione della speranza di vita alla nascita nelle regioni italiane. Fonte: ISTAT.

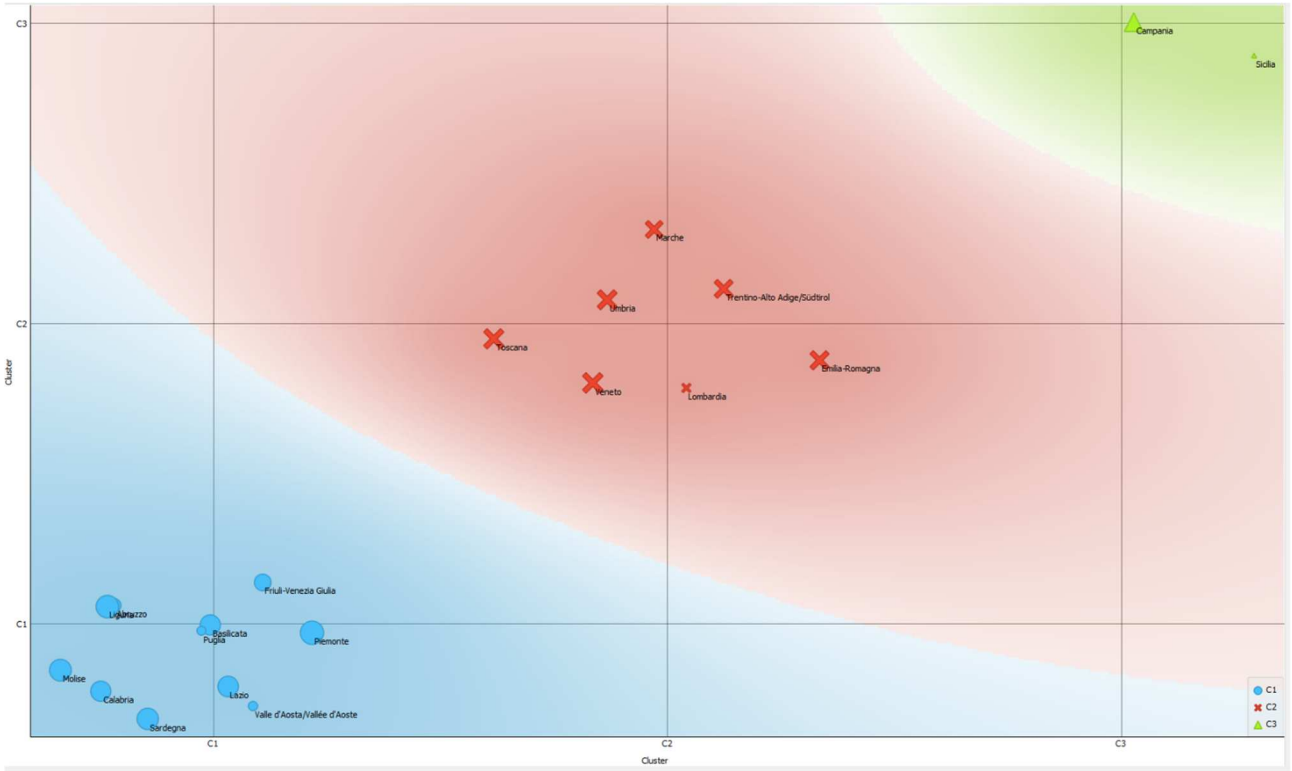


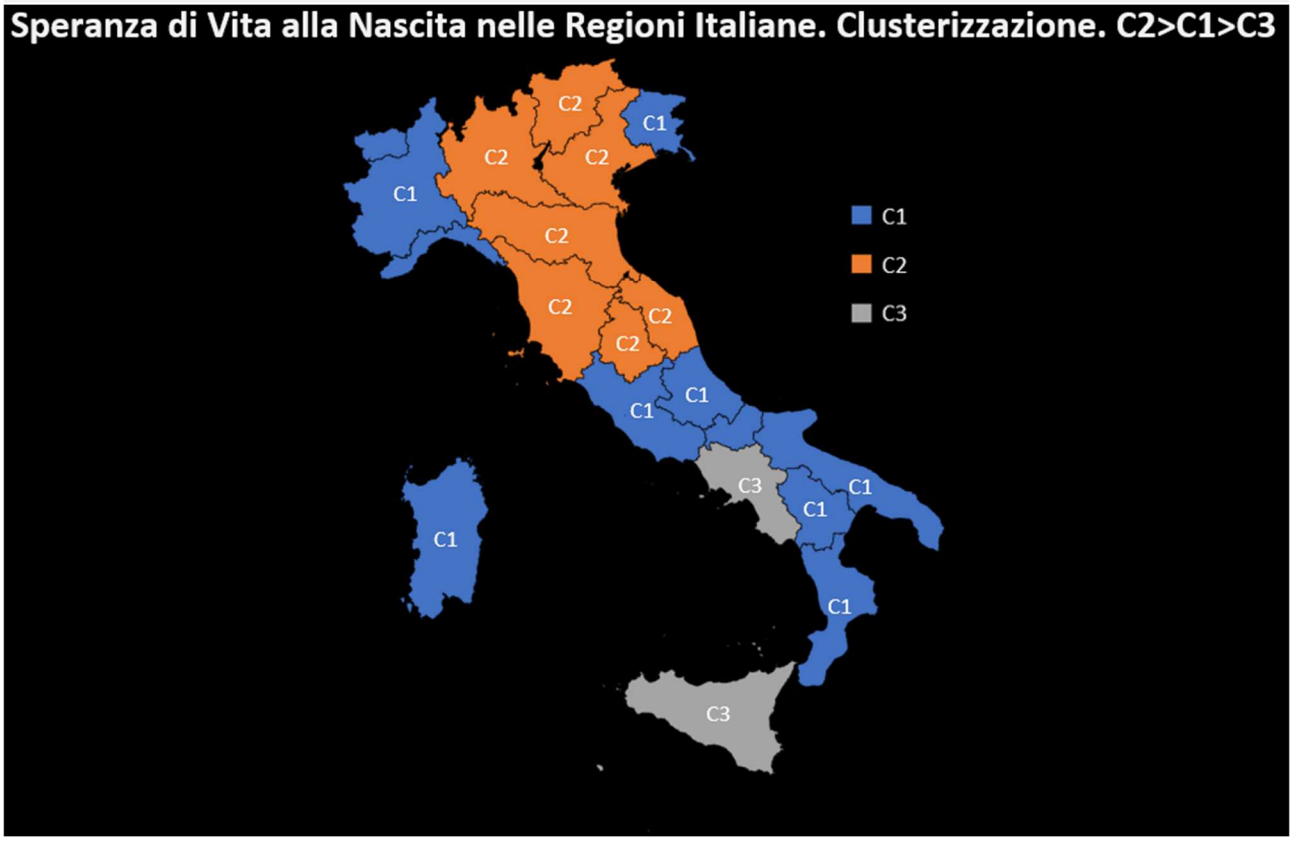
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Nord	80,9	81	81,3	81,5	81,5	81,7	82,1	82,3	82,3	82,5	83	82,7	83,1	83,2	83,3	83,6	82
Centro	80,9	81,1	81,4	81,5	81,5	81,6	81,9	82,1	82	82,3	82,8	82,6	83	82,9	83,4	83,6	83,1
Mezzogiorno	80,2	80,1	80,5	80,5	80,7	80,7	81,1	81,1	81,3	81,5	81,8	81,6	82,1	81,9	82,3	82,5	82,2
Italia	80,7	80,7	81,1	81,2	81,3	81,4	81,7	81,9	81,9	82,2	82,6	82,3	82,8	82,7	83	83,2	82,3

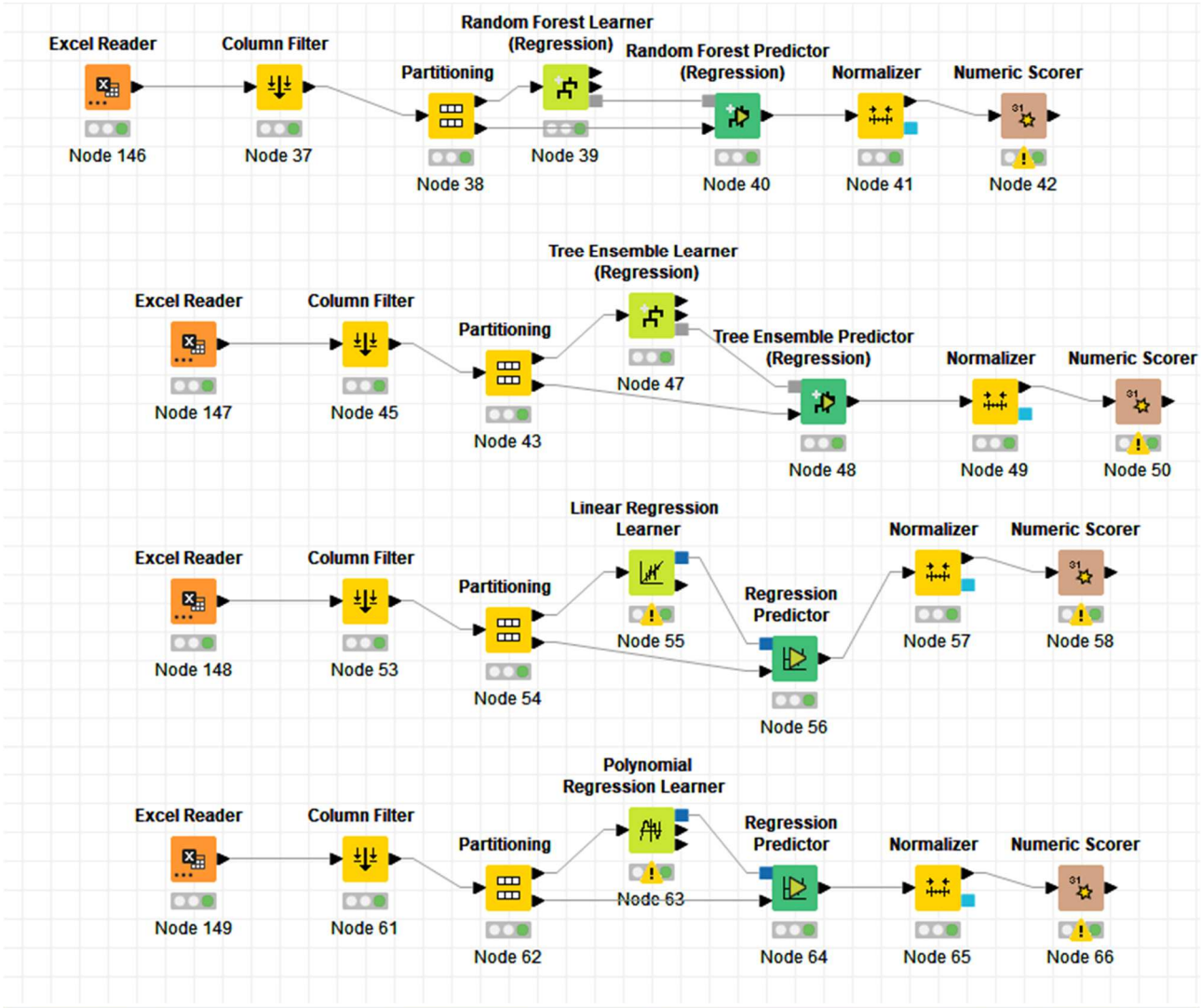


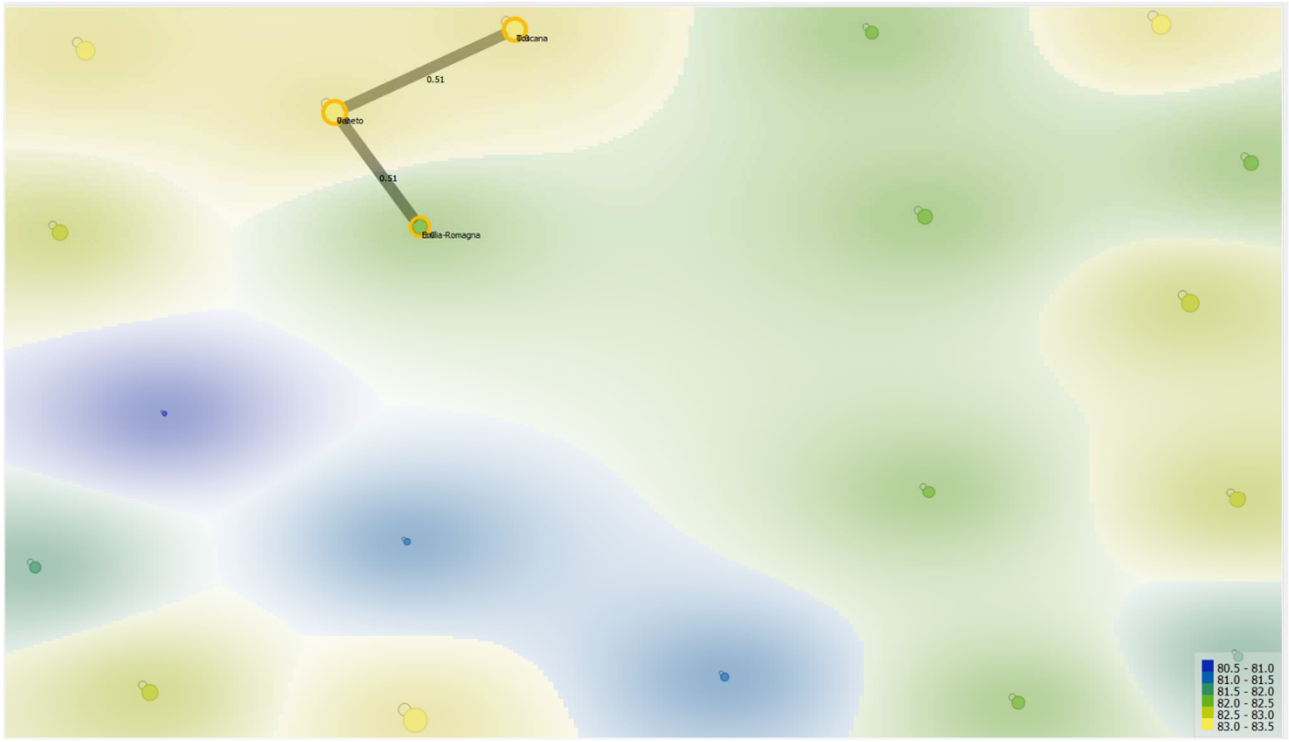
	2020	Prediction	Var Ass	Var Per
Valle D'Aosta	80,90	82,03	1,13	1,40
Emilia-Romagna	82,40	83,10	0,70	0,85
Toscana	83,00	83,10	0,10	0,12
Campania	81,40	82,77	1,37	1,68
Puglia	82,10	83,79	1,69	2,06
Sicilia	82,00	82,77	0,77	0,94
Media	81,97	82,93	0,96	1,17

Algorithm	R^2	mean absolute error	mean squared error	root mean squared error	Totale
Gradient Boosted Tree Regression	1	1	1	1	4
Simple Regression Tree	2	2	2	2	8
ANN	3	3	4	4	14
Tree Ensemble Regression	4	4	3	3	14
Random Forest Regression	5	5	6	6	22
PNN	6	7	5	5	23
Polynomial Regression	7	6	7	7	27
Linear Regression	8	8	8	8	32

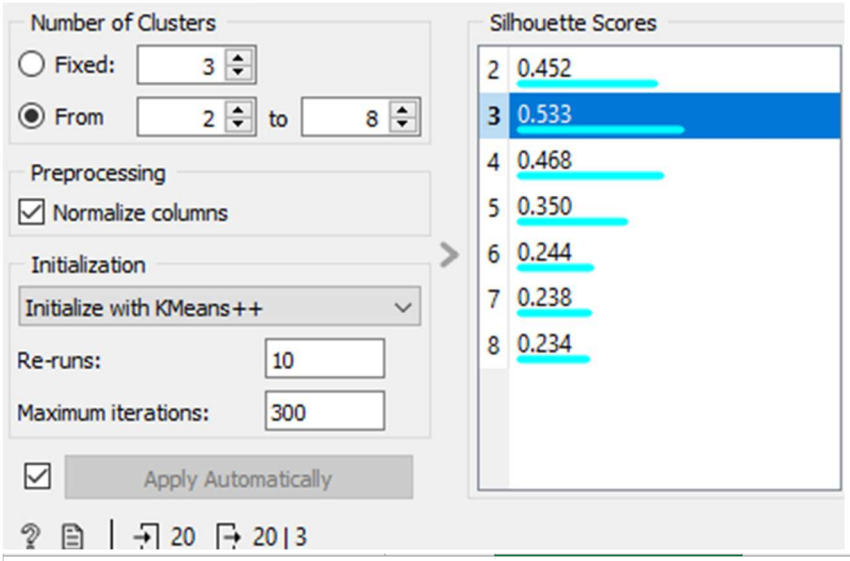

C3
C2
C1
Cluster
Campania
Sicilia
Marche
Umbria
Trentino-Alto Adige/Südtirol
Toscana
Emilia-Romagna
Veneto
Lombardia
Friuli-Venezia Giulia
Liguria
Basilicata
Puglia
Piemonte
Molise
Calabria
Lazio
Valle d'Aosta/Vallée d'Aoste
Sardegna
C1
C2
C3
C1
C2
C3
Cluster


Speranza di Vita alla Nascita nelle Regioni Italiane. Clusterizzazione. C2>C1>C3
C2
C1
C2
C2
C1
C2
C2
C2
C2
C1
C1
C1
C3
C1
C1
C1
C1
C3
C1
C2
C3

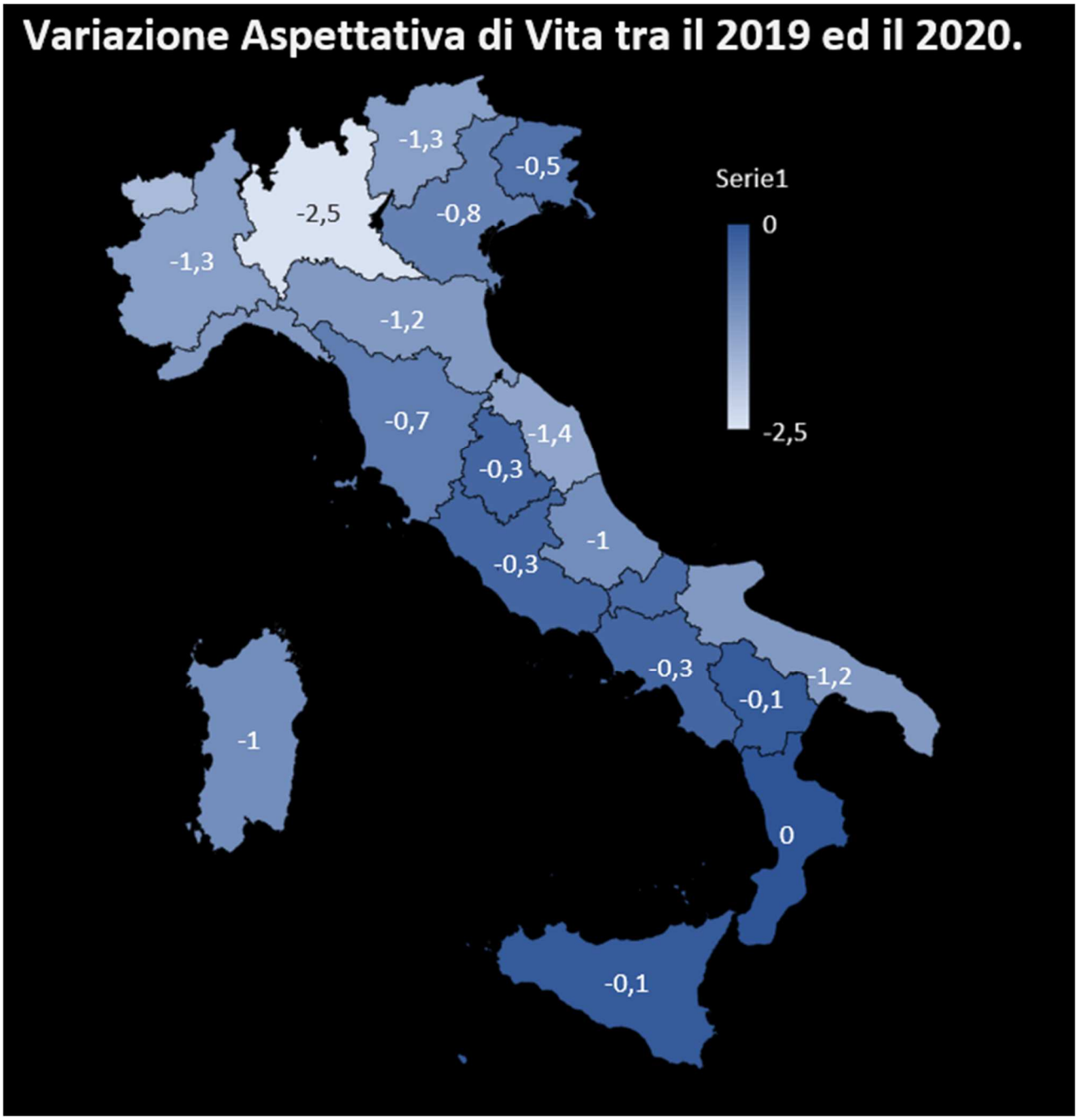

Excel Reader
Column Filter
Random Forest Learner (Regression)
Partitioning
Random Forest Predictor (Regression)
Normalizer
Numeric Scorer
Node 146
Node 37
Node 38
Node 39
Node 40
Node 41
Node 42
Tree Ensemble Learner (Regression)
Excel Reader
Column Filter
Partitioning
Tree Ensemble Predictor (Regression)
Normalizer
Numeric Scorer
Node 147
Node 45
Node 43
Node 47
Node 48
Node 49
Node 50
Linear Regression Learner
Excel Reader
Column Filter
Partitioning
Regression Predictor
Normalizer
Numeric Scorer
Node 148
Node 53
Node 54
Node 55
Node 56
Node 57
Node 58
Polynomial Regression Learner
Excel Reader
Column Filter
Partitioning
Regression Predictor
Normalizer
Numeric Scorer
Node 149
Node 61
Node 62
Node 63
Node 64
Node 65
Node 66

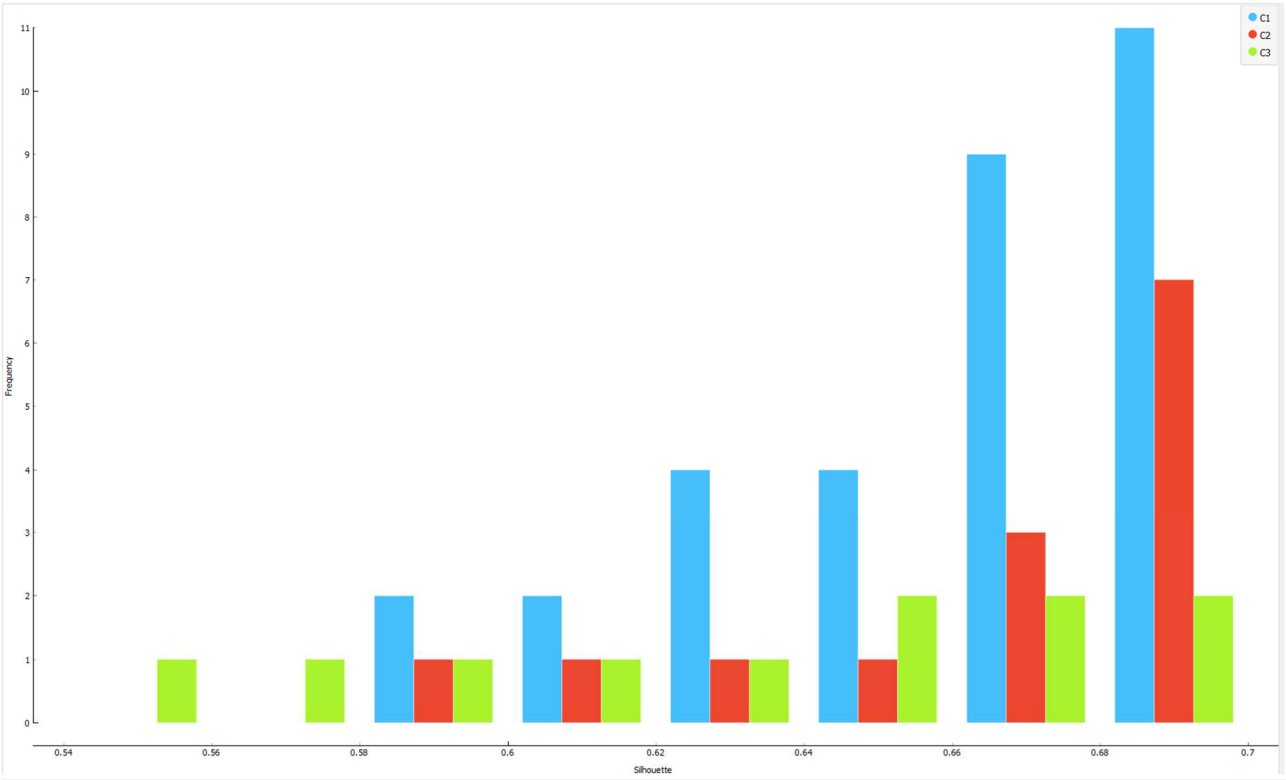

Toscana
0.51
Veneto
0.51
Emilia-Romagna
80.5 - 81.0
81.0 - 81.5
81.5 - 82.0
82.0 - 82.5
82.5 - 83.0
83.0 - 83.5

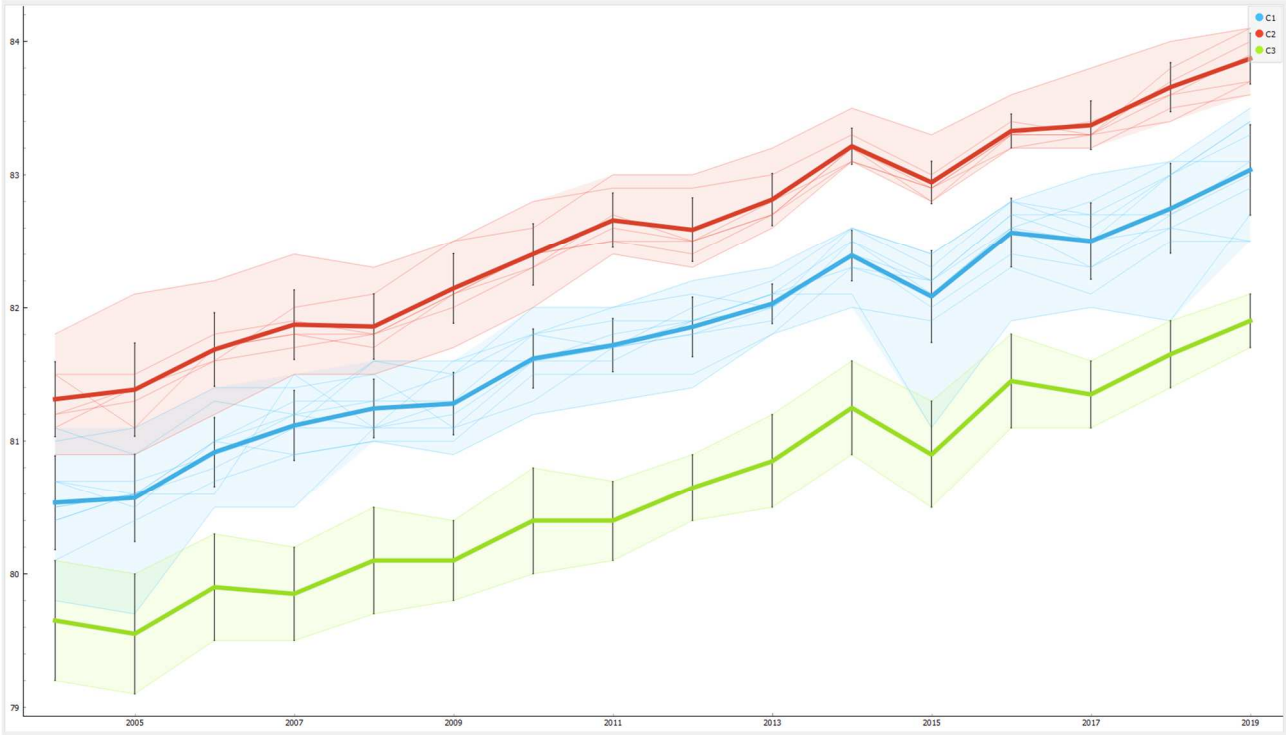
	2020	Regione	Cluster
13	82.4	Abruzzo	C1
17	82.4	Basilicata	C1
18	82.5	Calabria	C1
15	81.4	Campania	C3
8	82.4	Emilia-Romagna	C2
7	83.0	Friuli-Venezia G...	C1
12	83.1	Lazio	C1
3	81.9	Liguria	C1
4	81.2	Lombardia	C2
11	82.6	Marche	C2
14	82.6	Molise	C1
1	81.6	Piemonte	C1
16	82.1	Puglia	C1
20	82.1	Sardegna	C1
19	82.0	Sicilia	C3
9	83.0	Toscana	C2
5	82.8	Trentino-Alto A...	C2
10	83.8	Umbria	C2
2	80.9	Valle d'Aosta/V...	C1
6	83.1	Veneto	C2



Statistical Results of Machine Learning Algorithms				
Algorithm	**R^2**	**mean absolute error**	**mean squared error**	**root mean squared error**
ANN	-0,382	0,269680695	0,180784299	0,425187369
PNN	-0,8246	0,385964912	0,203854842	0,45150287
Simple Regression Tree	-0,3084	0,261363636	0,138310185	0,371900773
Gradient Boosted Tree Regressio	0,36544	0,202421622	0,065550561	0,256028439
Random Forest Regression	-0,5986	0,332676433	0,213421121	0,461975238
Tree Ensemble Regression	-0,4167	0,291913565	0,151228485	0,388881068
Linear Regression	-2	0,561403509	0,454293629	0,674013078
Polynomial Regression	-0,9014	0,380952381	0,306122449	0,553283335


Variazione Aspettativa di Vita tra il 2019 ed il 2020.
-1,3
-0,5
-2,5
-0,8
-1,3
-1,2
-0,7
-1,4
-0,3
-1
-0,3
-0,3
-1,2
-0,1
-1
0
-0,1
Serie1
0
-2,5





Variazione della Speranza di Vita alla Nascita nelle Macro-Regioni Italiane

	2004-2020		2015-2020		2019-2020	
	Var Ass	Var Per	Var Ass	Var Per	Var Ass	Var Per
Nord	1,10	1,36	-0,70	-0,85	-1,60	-1,91
Centro	2,20	2,72	0,50	0,61	-0,50	-0,60
Mezzogiorno	2,00	2,49	0,60	0,74	-0,30	-0,36
Italia	1,60	1,98	0,00	0,00	-0,90	-1,08