

Analysis of well-being in Italy

Geospatial Analysis and Representation for Data Science project

• INTRODUCTION

In the last decades, an increasing number of indexes and rankings have been deployed in order to monitor the development of countries in various economic, social, and cultural aspects. Each of these indicators provides its personal scores, based on the measurement and weighting of different categories of data. For what concerns well-being, there is no unique and official instrument to track the level reached by each country and in fact many of these nations implemented an individual method to measure and track their own score and compare it to the ones of the other countries.

For the purpose of this project, the analysis focus on Italy and the data were obtained by its Well-Being and Sustainability (BES) Index¹. The objective is to offer to the reader of this paper a representative overview of the Italian welfare status in 2021 in the areas of economy, health, and education. For each of these topics will be provided insights in specific aspects by combining the data with further datasets and attempting to uncover hidden patterns. In fact, the expected outcome of this paper is not to offer a descriptive analysis of the whole data provided by the Well-Being and Sustainability Index, but to investigate some of them, explore new potential relationships between phenomena, and stimulate further analysis.

• METHODOLOGICAL ASPECTS

A. BES data cleaning

For the purpose of this project, the data were mainly retrieved by the “Dataset Bes at local level 2022 edition”² provided by the Italian Institute of Statistics (ISTAT). The dataset consists on

¹ <https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being>

² <https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being/bes-at-local-level>

more than 14,000 rows and 26 columns. Each row contains data referring to a specific indicator of one of the domain under analysis, and the data are further classified according to the gender of the citizens and their region or province. The data included in the dataset were collected starting from 2004, but some of the categories were introduced later and therefore there are some missing entries. Furthermore, the data referring to the last years are sometimes just provisional or even not yet available. In the project, only the data of 2021 were used for the analysis. Moreover, other dataset have occasionally been used to provide further details and deepen the knowledge on specific areas.

B. Geospatial data joining

The BES dataset provided by ISTAT did not include any spatial reference and, for this reason, it was merged with a shapefile containing the geographical information of the Italian regions and provinces³. These information were stored as multipolygons and according to the WGS 84 / UTM zone 32N coordinate reference system.

C. Data forecasting

As a first step of the analysis, the data were divided accordingly to their domain and then the indicators were considered individually. As mentioned before, the data provided were not always homogeneous and in fact many of them did not provide any information for the year of interest of this project. In order to solve the issue, a time-series forecasting method was applied. The one that appeared the most suitable was the Exponential Smoothing method, that is useful for the forecasting of data without seasonality, and works by assigning a higher weight to the data closest to the year to forecast. The quality of the forecasts was adequate to provide a reasonable prediction of the values of 2021, despite the limited amount of data of previous years that were available for the computation.

D. Spatial regression analysis

One of the first steps of the analysis consisted in plotting the choropleth maps of the provided data. These maps are an effective way to express statistical information in a visual form by using a geographical representation and a palette of colours. Then, in order to proceed with the analysis, a spatial regression was applied to the data of each target. Such method was used in order to verify the existence of eventual relationships among the variables under analysis. In fact, according to Tobler's first law of geography, the geographical proximity of things increases the possibility of their mutual dependence. The first step of spatial regression consists in the creation of a spatial weight

³ <https://www.istat.it/it/archivio/222527>, generalised version 2021

matrix that contains information on the connectivity of the elements of the matrix itself. Among the different definition of neighbour that could be used for the analysis, it was determined that the most suitable for the analysis was the contiguity-based one, where the existence of a shared border between regions is the necessary condition. In the case of Italy, there are two regions (Sicily and Sardinia) that are islands and therefore do not share any border. For the purpose of the project, it was decided to create a connection with the regions closest to each of them (respectively Calabria and Lazio). Once the neighbourhood has been defined, the computation of the spatial weight matrix could be performed. The use of the global Moran's I model allows to verify the existence of spatial correlation among variables. This choice is based on the idea that the influence of these relationships is not limited to the local context, but that further propagates. The Monte Carlo approach is particularly effective for the purpose.

• ANALYSIS OF THE DOMAINS AND INDICATORS

1. HEALTH

For what concerns health, the data related to road mortality, avoidable mortality, and life expectancy in 2021 were used for the analysis. The first and last indicators were already provided with the data of the year, while the avoidable mortality of 2020 and 2021 had to be forecasted.

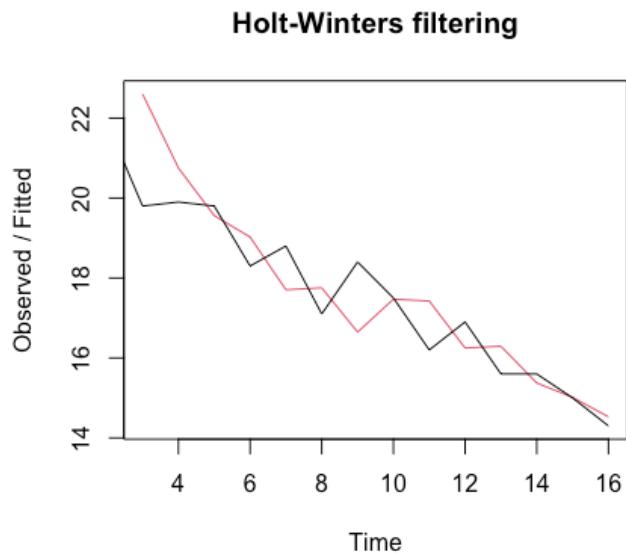


Figure 1: Observed and fitted data on avoidable mortality

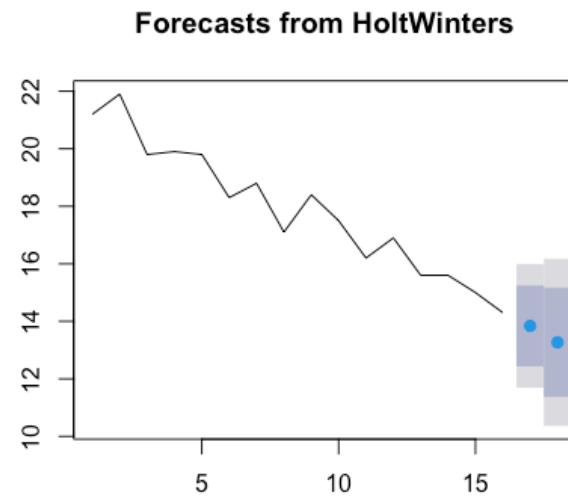


Figure 2: Forecast of 2020 and 2021 data

• ROAD MORTALITY

The road mortality data do not seem to suggest any relevant spatial pattern, as can be seen by plotting the data on the map of the Italian regions (see *Figure 3*), and this is also confirmed by the Moran's I tests and the Monte Carlo approach. A more detailed analysis performed on the provinces suggested that the area of Rovigo, in the Veneto region, has the highest mortality rate compared to all the other available data (see *Figure 4*).



Figure 3: Road mortality in 2021

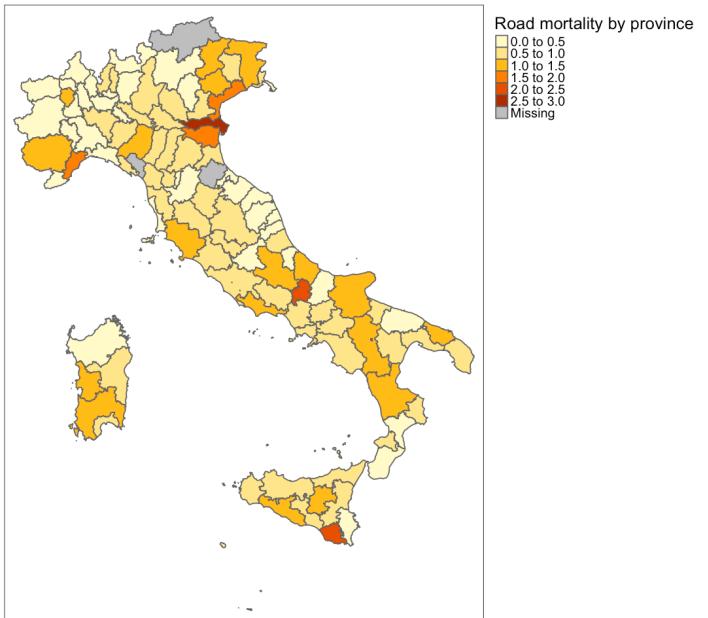


Figure 4: Road mortality by province

According to the statistics, Rovigo does not present a very high number of accidents compared to the other provinces of Veneto but the rate of fatalities is significantly higher. The Italian Institute of Statistics published a report concerning the road accidents in Veneto in 2021 and registered a 9.6 mortality rate in Rovigo, compared to 7.5 of Belluno, 7.2 of Treviso, 7.1 of Venice, 5.1 of Verona, 4.9 of Padova, and 3.8 of Vicenza⁴.

In order to assess the cities or towns where the majority of these occurrences happened, a further dataset⁵ customised on the basis of the data provided by ISTAT was analysed. The outcomes suggest that Rovigo itself is the city of the province where the majority of the road accidents related injuries and deaths happened.

⁴ https://www.istat.it/it/files//2022/11/FOCUS_VENETO-2021.pdf

⁵ http://dati.istat.it/Index.aspx?DataSetCode=DCIS_MORTIFERITISTR1&Lang=en

ITTER107	Territory	TIPO_DATO22	Data type	ESITO	Result	TIME	Select time	Value
1	029041	Rovigo	KILLINJ	killed and injured	F	injured	2021	228
3	029001	Adria	KILLINJ	killed and injured	F	injured	2021	61
63	029033	Occhiobello	KILLINJ	killed and injured	F	injured	2021	51
77	029052	Porto Viro	KILLINJ	killed and injured	F	injured	2021	31
55	029029	Lendinara	KILLINJ	killed and injured	F	injured	2021	28

Figure 5: top 5 locations with highest number of injured in the province of Rovigo

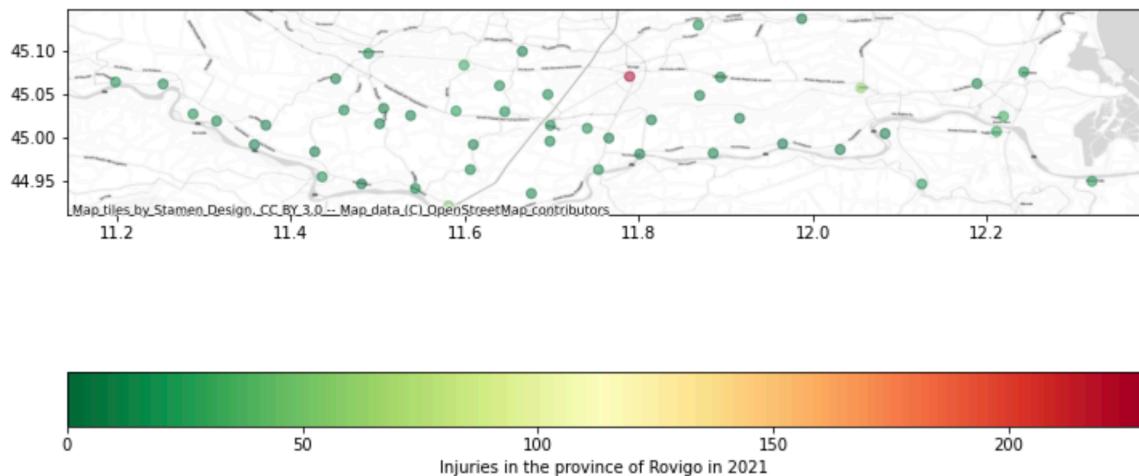


Figure 6: road injuries in the province of Rovigo in 2021

ITTER107	Territory	TIPO_DATO22	Data type	ESITO	Result	TIME	Select time	Value
0	029041	Rovigo	KILLINJ	killed and injured	M	killed	2021	6
8	029004	Badia Polesine	KILLINJ	killed and injured	M	killed	2021	2
78	029040	Rosolina	KILLINJ	killed and injured	M	killed	2021	2
30	029015	Ceregnano	KILLINJ	killed and injured	M	killed	2021	2
6	029003	Arquà Polesine	KILLINJ	killed and injured	M	killed	2021	1

Figure 7: top 5 locations with highest number of fatalities in the province of Rovigo

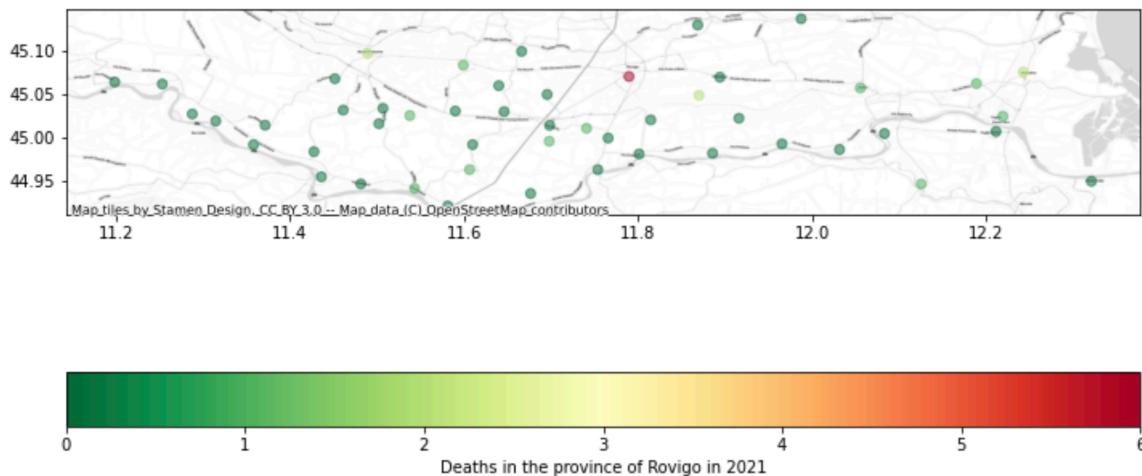


Figure 8: road fatalities in the province of Rovigo in 2021

• AVOIDABLE MORTALITY

Regarding the topic of avoidable mortality, the statistical tests suggest a form of global spatial correlation among regions. In particular, it appears that the regions of South Italy and the Islands have a higher incidence of this type of deaths compared to the northern regions, while the ones of the Center present intermediate values.

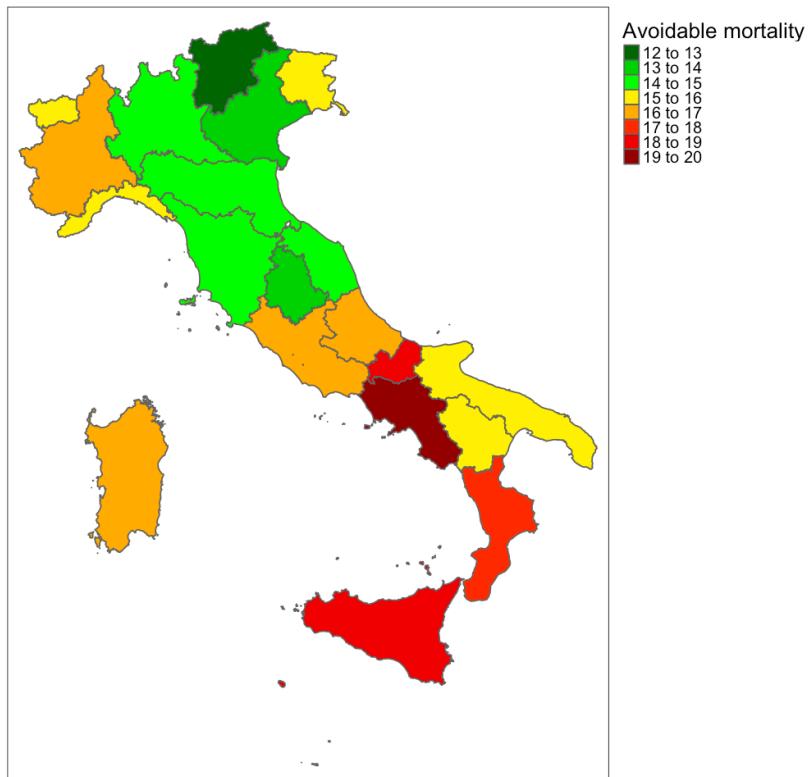


Figure 9: incidence of avoidable mortality

The incidence of avoidable mortality cases is mostly generated by the lack of sanitary structures and services for preventable diseases and injuries⁶. On the basis of the provided data, it could be inferred that the region Campania has the lowest ratio of healthcare structures per density of population and therefore does not have enough structures to provide its inhabitants with the needed services. In order to verify such claim on the basis of data provided by ISTAT⁷ (the data refer to the year 2019 but are considered appropriate for the analysis) the number of hospitals in Campania was compared to the one of Lazio, since the two regions have a similar population (respectively 5,839,084 and 5,898,124 residents). The analysis shows that the region Campania can count on 108 hospitals, whereas Lazio relies on 114 structures. *Figure 10* and *Figure 11* show a representative overview of the distribution of the structures in the regions.

⁶ <https://www.oecd-ilibrary.org/sites/6cf53429-en/index.html?itemId=/content/component/6cf53429-en>

⁷ <http://dati.istat.it/Index.aspx?QueryId=11099&lang=en#>

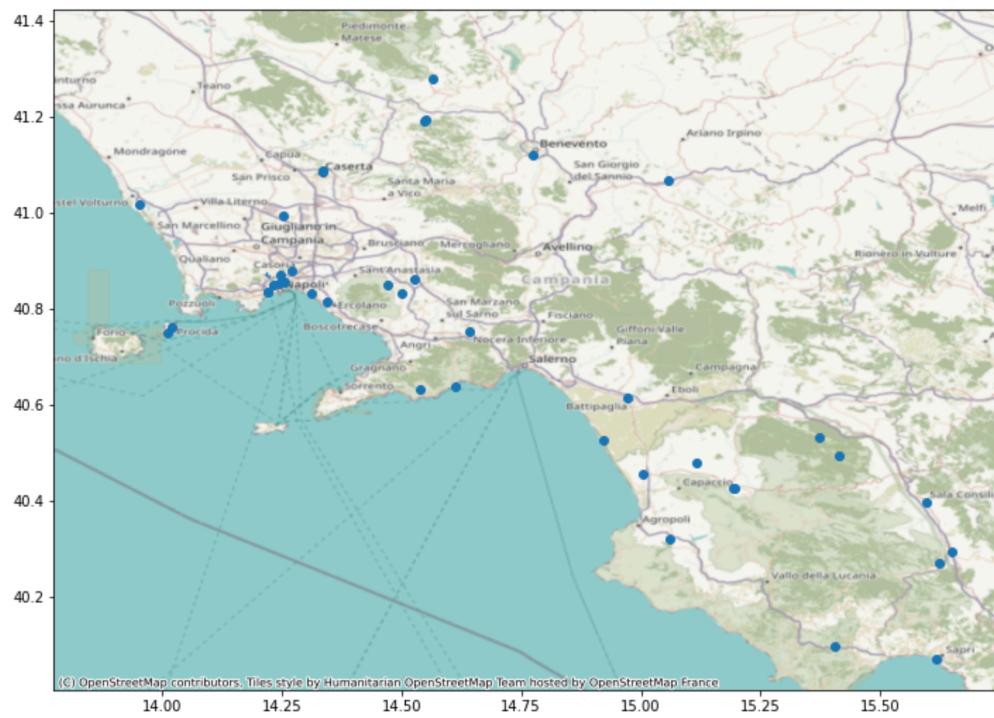


Figure 10: hospitals in Campania

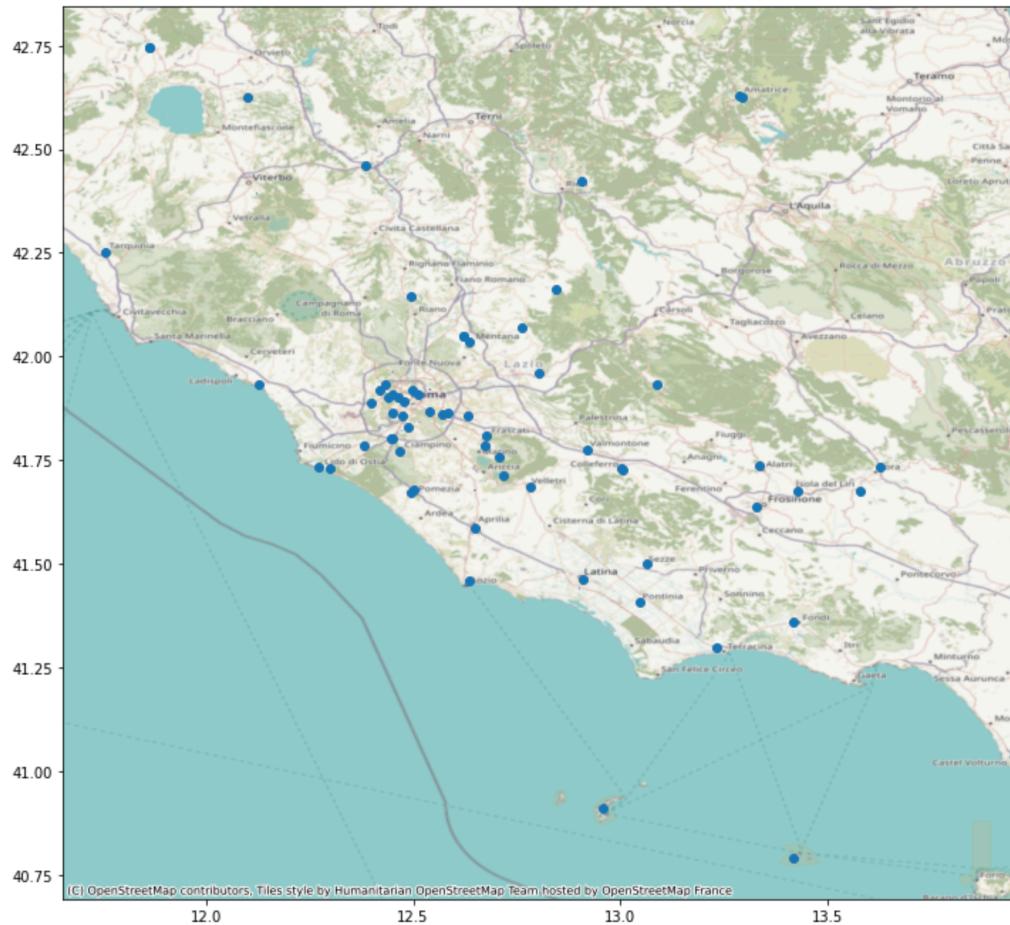


Figure 11: hospitals in Lazio

The numbers are particularly significant if compared to the one of Trentino-Alto Adige, where there are 23 hospitals for a total of 1,062,860 people. The comparison with these data suggests the need to increase the number of structures both in Lazio and Campania up to about 150 in order to reach the level of coverage offered by Trentino-Alto Adige.

• LIFE EXPECTANCY AT BIRTH

Similarly to the case of avoidable mortality, the life expectancy at birth is strongly characterised by a spatial component. The regions of Southern Italy are the ones with the lowest longevity, with Campania and Sicily at the bottom of the ranking.

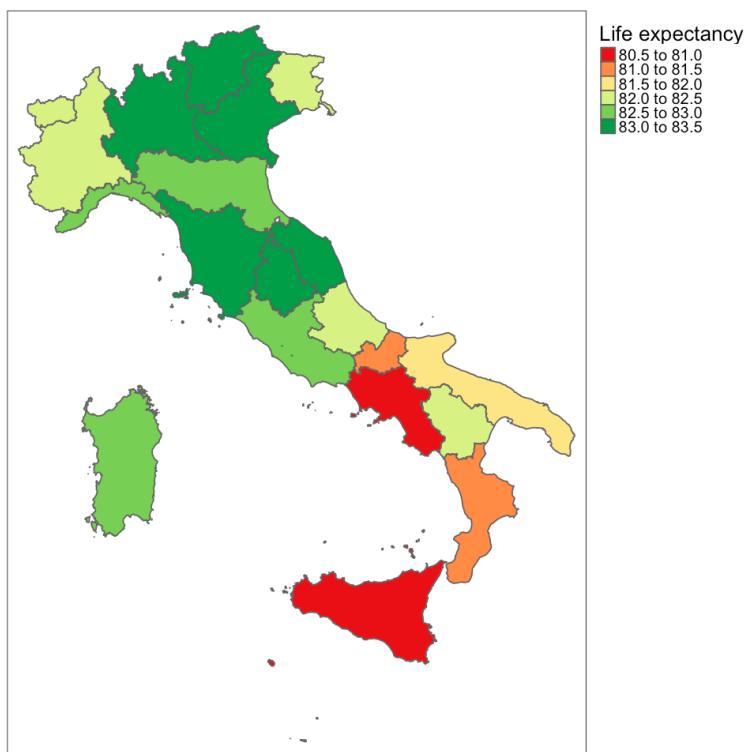


Figure 12: life expectancy in 2021

There are many factors that can affect the length of people's life, especially the ones related to their lifestyles. In order to understand whether the habits of people can affect their longevity, the correlation between life expectancy and overweight, smokers rate, alcohol intake, inactivity, and adequate nutrition was analysed with the use of a correlation matrix. The outcome suggests that longevity is highly negatively correlated with overweight and inactivity. Assessing the multicollinearity of the variables, it was possible to detect this phenomenon in both the highly correlated attributes. In order to solve the issue, a new analysis excluding overweight was performed and it was possible to assess that the life expectancy is, indeed, strictly negatively related to inactivity.

The choice of including inactivity instead of overweight was due to the higher significance of the former attribute, as emerged in the analysis.

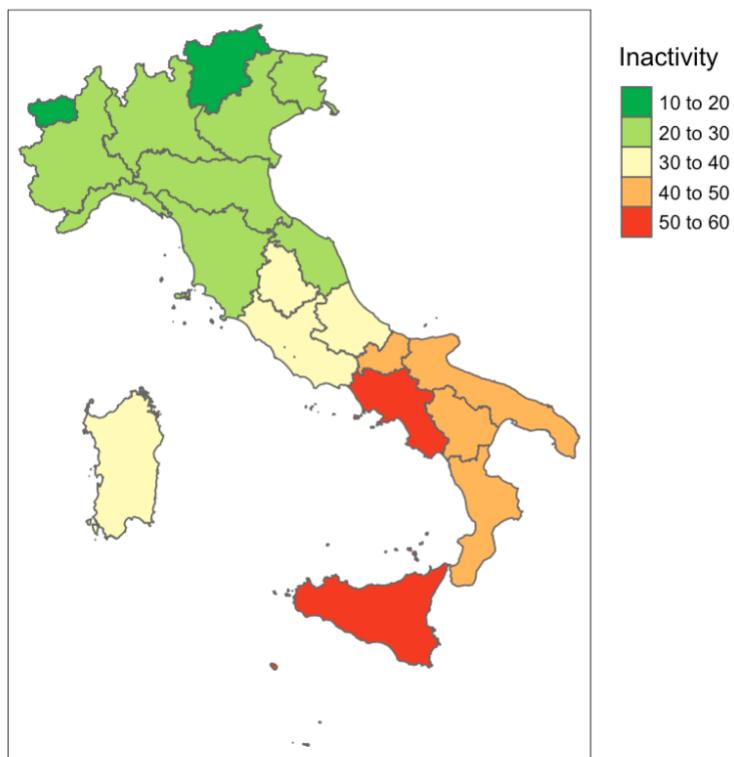


Figure 13: inactivity rate in 2021

2. ECONOMIC WELL-BEING

The topic of economic well-being is particularly relevant in addressing the overall welfare status of a country. The dataset provided by the Institute of Statistics offers data related to the average pensions, salaries, and incomes. Furthermore, also the percentage of low pensions, bad loan entry rates, and the assets per capita have been collected.

As done for the missing 2021 health data, the exponential smoothing approach was applied to the groups of indicators that did not include any data for the year of interest. Subsequently, the Moran's I test and Monte Carlo approach were computed for each of the indicators of economic well-being. The results highlighted a strong spatial correlation for all the indicators, with the exclusion of the bad loan entry rates that seem quite homogeneous across the peninsula.

In order to provide a general overview of the overall economic situation of each region, comprehensive values of well-being based on the available indicators have been computed. The computation of such data was performed by standardising the data of the available indicators and

determining the mean of the results for each region with the use of the RowMean function. For this computation, the 6 indicators were assumed to be equivalent in weight (each of them accounts for one sixth of the overall value).

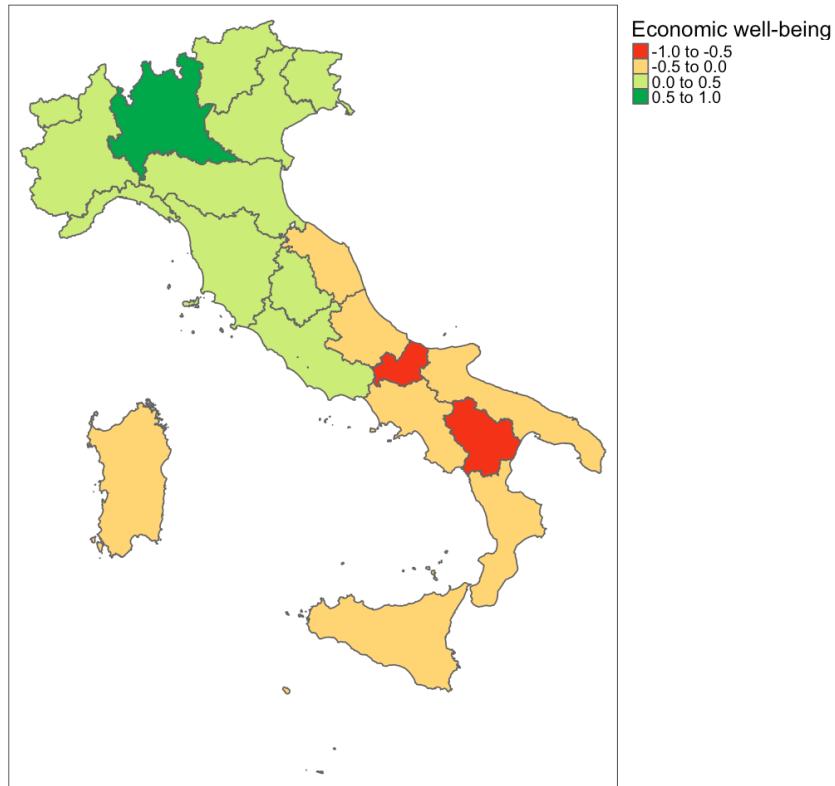


Figure 14: overall economic well-being in 2021

According to the map, a spatial component in the economic welfare of Italy appears quite clearly. The regions of Northern Italy seem to be significantly wealthier than the ones of the South and Islands, while the central area of the peninsula is equally divided between regions that are economically wealthy and others that still struggle financially. A significantly high value has been reached by Lombardy, while the worst results have been registered in Molise and Basilicata. The different values reached in the regions could be due to each of the previously aggregated variables or to a combination of them, even if the salary is commonly considered as the main factor of those under analysis in determining the economic well-being of a territory. In order to understand whether the salary disparity actually matches the level of welfare reached by the regions, it was performed an analysis considering the average salary of each single province of the three regions, since the data suggest a significant geographical disparity (from 14,000€ up to about 30,000€). The outcome shows that the average salary in Molise and Basilicata is quite homogenous within the provinces of each region, while the value of Lombardy is significantly influenced by the result of the

province of Milan. In fact, the salaries in the other provinces of Lombardy are much more similar to the ones of Molise and Basilicata than expected.

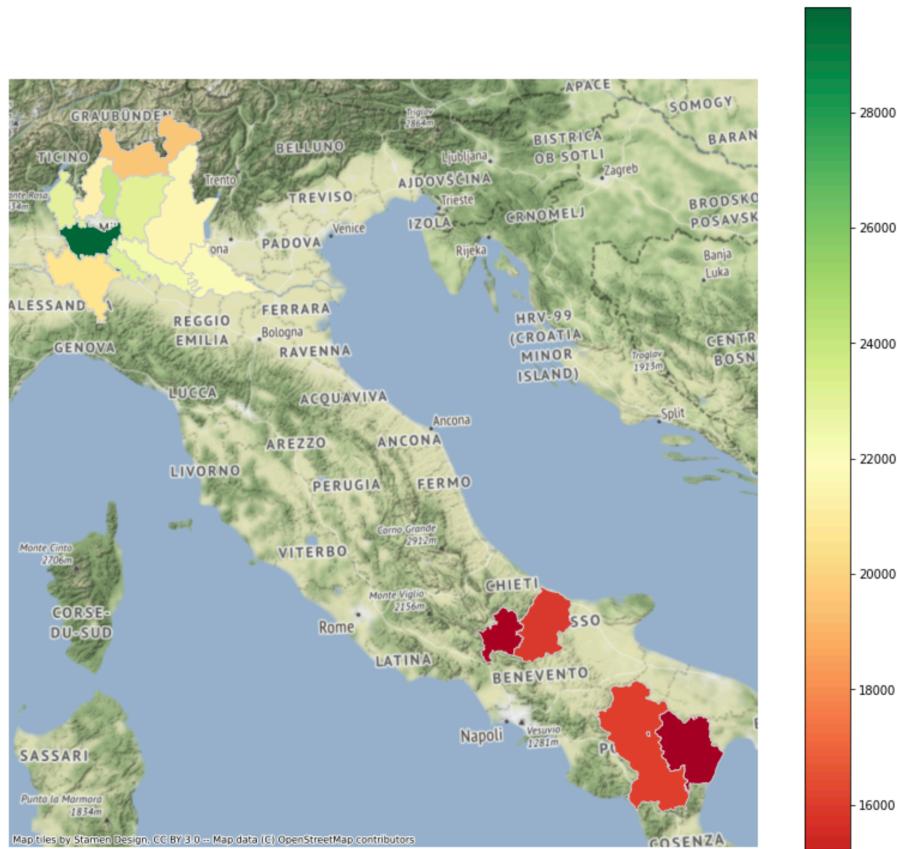


Figure 15: average salary in Lombardy, Molise, and Basilicata

3. EDUCATION AND TRAINING

Education is another vital indicator of the welfare status of a country. Educated people are, indeed, more easily employed, earn higher salaries, and have lower probabilities of living below the poverty level. According to the ISTAT dataset, there are several measures that contribute to the definition of the quality of the Italian educative system. The attendance to school from an early age up to university is taken into account, as well as the literacy and numeracy level of young students, the percentage of NEETs⁸, and the participation rate to life-long learning.

⁸ NEET: Not engaged in Education, Employment or Training

• LITERACY AND NUMERACY

The data on literacy and numeracy levels are particularly insightful for the analysis. In fact, despite the mandatory school attendance of children since a young age, a significant percentage of them still presents inadequate proficiency in those areas in the 8th grade. The rate of students with inadequate understanding of mathematical concepts appears to be even higher than the one of inadequate literary knowledge. Even in these cases, the statistical tests performed on the data suggest the relevance of the spatial component.

A more detailed analysis provides further insights into the distribution of these rates, showing that males tend to have a lower literacy proficiency than females. The situation is particularly severe in Calabria and Sicily, with peaks of about 65%. The worst results have been reached in the province of Crotone (Calabria), with percentages of respectively 51.3 and 64.2 for females and males. The lowest rate of inadequacy has been registered in the province of Belluno (Veneto), with rates of 20.9 for females and 28.4 for males.

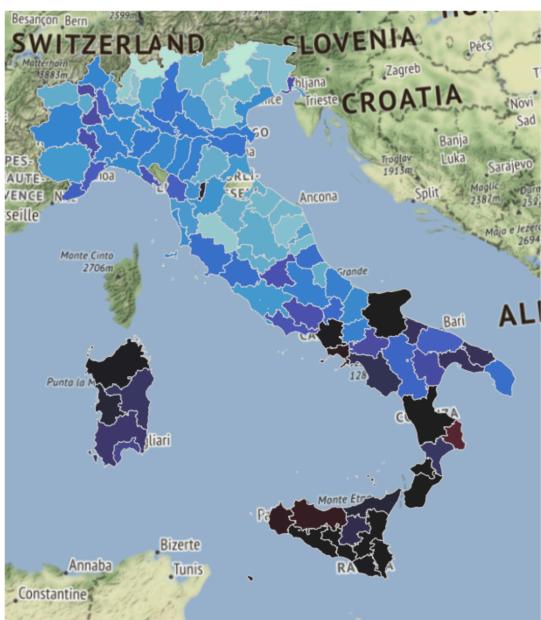


Figure 16: share of inadequate literacy among females

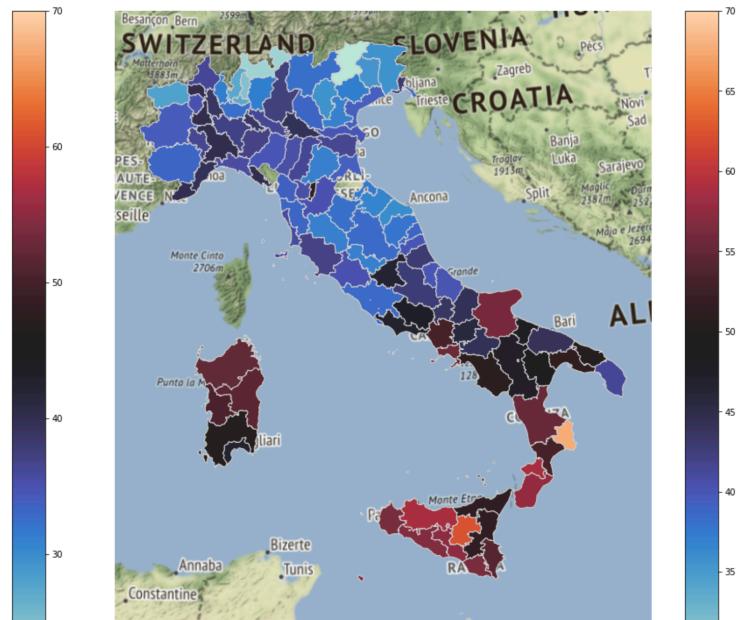


Figure 17: share of inadequate literacy among males

Also in the case of mathematical proficiency, the worst results have been registered in the province of Crotone, with values for females and male of respectively 74.8 and 73.5. However, the differences between genders are not as significant as in the previous case. On the other hand, the best performances have been registered in the province of Sondrio (Lombardy) and amounted to 23.4 for females and 22.1 for males.

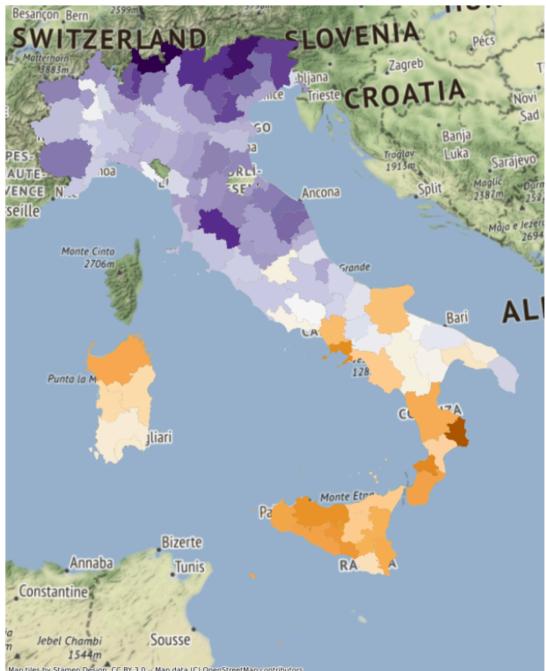


Figure 18: share of inadequate numeracy among females

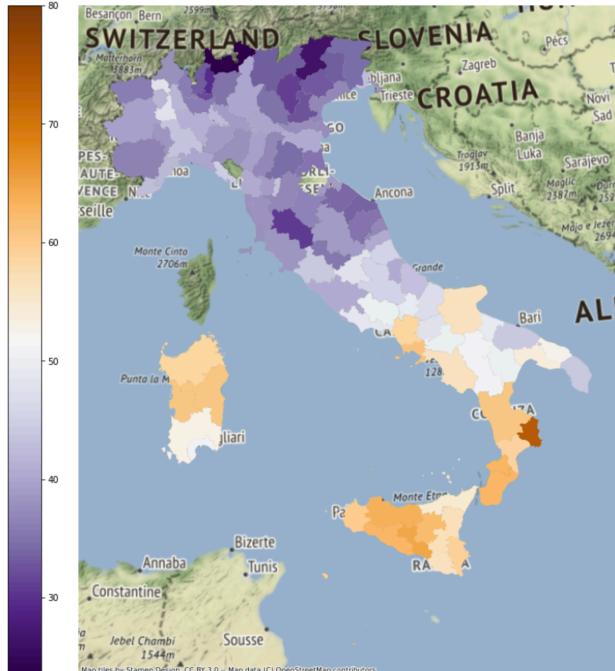


Figure 19: share of inadequate numeracy among males

• LIFE-LONG LEARNING AND NEETS

Other interesting measures of the goodness of the Italian education and training systems are surely represented by the share of NEETs and life-long learners. The pattern follows the same distribution of the data on literacy and numeracy, revealing an higher number of life-long learners and lower amount of NEETs in the northern regions compared to the rest of the country.

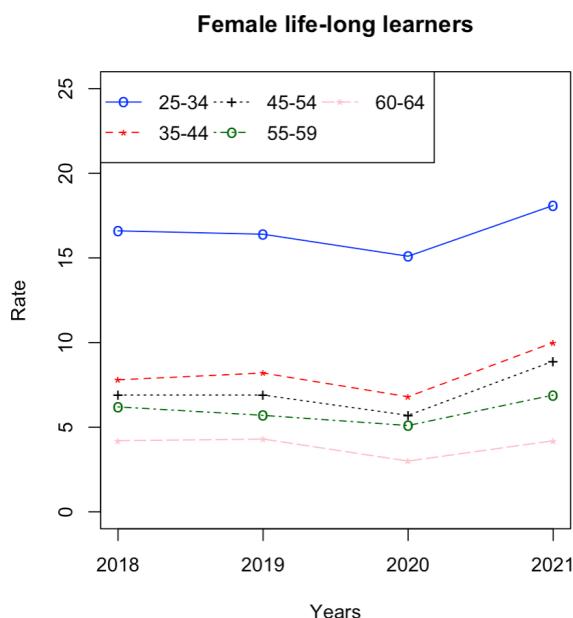


Figure 22: female life-long learners trend (2018-2021)

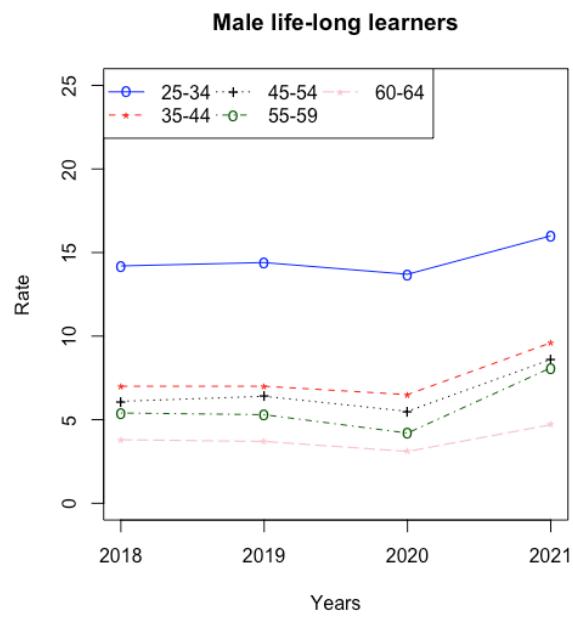


Figure 23: male life-long learners trend (2018-2021)

For the topic of life-long learning, an interesting aspect to notice is that the female group tends to have a slightly higher percentage of learners in almost all the considered age classes (*Figure 22* and *Figure 23*). The overall trend of the last 4 years is positive for both genders despite the slight decline of 2020 that was probably due to the consequences of the COVID19 pandemics.

Concerning regional disparity of such data across the country, an interesting aspect to consider is the presence of high-ranked universities in the territory. In fact, as suggested by the UNESCO Institute for Lifelong Learning, higher education institutes are the main sources of promotion of life-long learning.⁹ For this analysis, the data of Censis¹⁰ were aggregated and used to locate the best universities in Italy. The result shows that the 15 best universities of 2021 are mainly concentrated in the Northern and Central regions of Italy, with the only exceptions of the University of Salerno (Campania, South Italy) and the University of Sassari (Sardinia, Islands). As a consequence, the rates of life-long learners could be affected by such uneven distribution of high quality institutions across the country.

	University	Average points
0	Liberà Università di Bolzano	101.0
1	Università di Camerino UNICAM	98.2
2	Università degli Studi di Trento	97.3
3	Università di Milano Bocconi	96.2
4	Università di Roma LUISS	94.2
5	Università degli Studi di Siena	94.0
6	Università degli Studi di Perugia	93.3
7	Università degli Studi di Milano Statale	93.3
8	Università degli Studi di Sassari	92.8
9	Università degli Studi di Udine	92.8
10	Università degli Studi di Trieste	92.0
11	Università degli Studi di Bologna	91.8
12	Università degli Studi di Salerno	91.8
13	Università Politecnica delle Marche	91.3
14	Università degli Studi di Pavia	91.2

Figure 24: Censis ranking of the best Italian universities

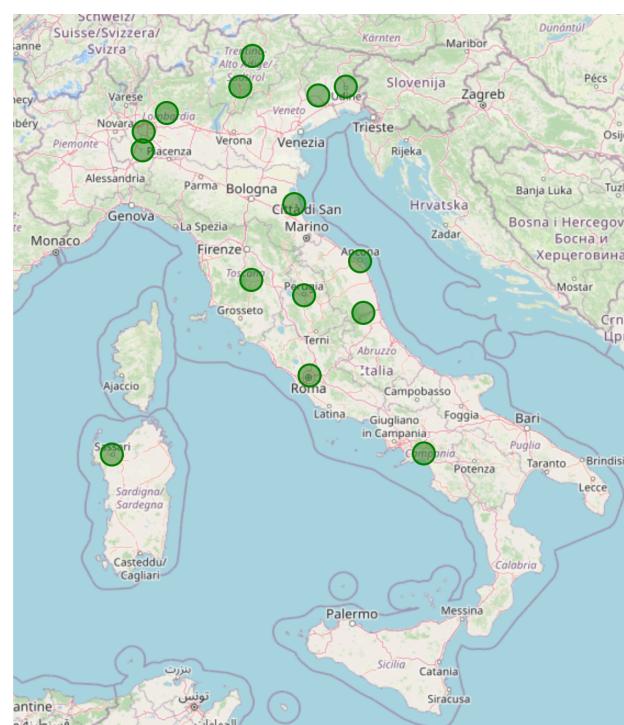


Figure 25: map of the best Italian universities in 2021

For what concerns NEETs, the rate is significantly higher in the Southern regions and in the Islands, especially in Sicily.

⁹ <https://www.uil.unesco.org/en/role-higher-education-institutions-lifelong-learning>

¹⁰ [https://www.censis.it/sites/default/files/downloads/Classifica Censis delle Università 2021-2022_0.pdf](https://www.censis.it/sites/default/files/downloads/Classifica%20Censis%20delle%20Universit%202021-2022_0.pdf)

The last years trends highlight the significant gender difference between the rates. In fact, despite the percentages being quite similar up to 24 years old, they change dramatically in the 25-29 years range. Women are much more prone to become NEETs during adulthood, while men's percentage decreases. The trends have been quite stable in the last 4 years and are a possible reflection of the Italian cultural heritage, where women are more likely than men to engage in activities of full-time care of the household and children, hence falling in the NEETs category.

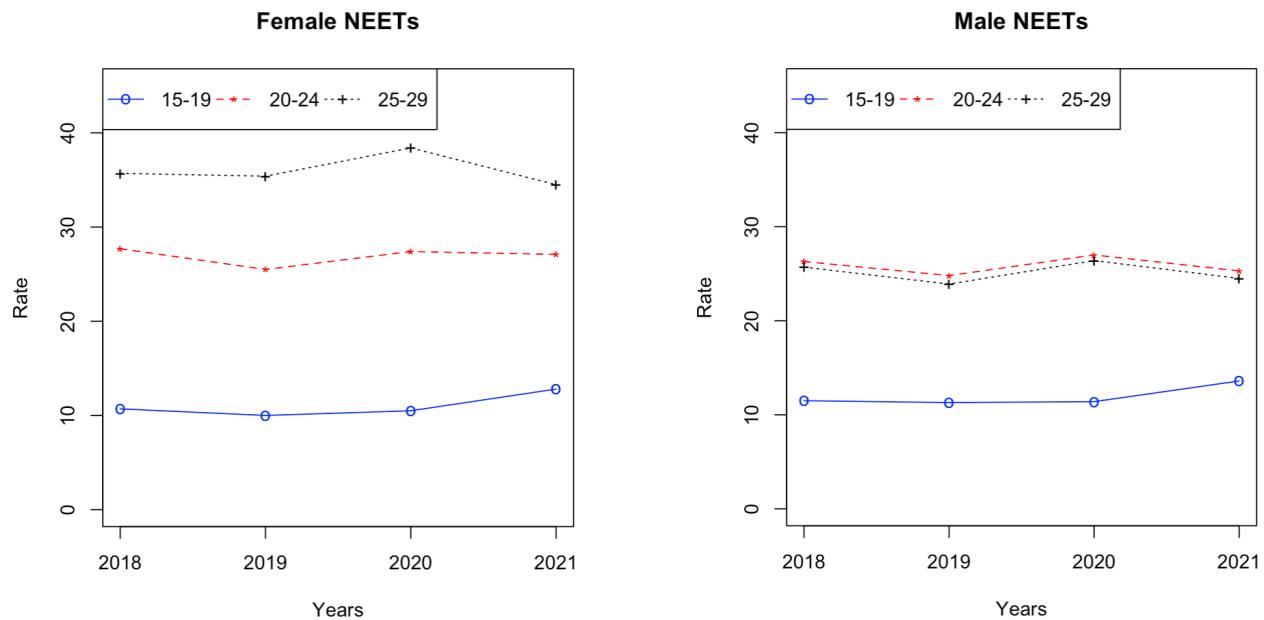


Figure 26: female NEETs trend (2018-2021)

Figure 27: male NEETs trend (2018-2021)

• RESULTS AND CONCLUSION

The results obtained with this analysis suggest that a comprehensive Italian measures of well-being is difficult to define because of the many territorial differences. Overall, the analysis conducted points towards the conclusion that the regions of North and Central Italy in 2021 enjoyed a higher well-being in terms of health, economy, and education compared to the regions of South Italy. For what concerns the two islands, Sardinia had values similar to the ones of Central Italy, while Sicily was much more similar to the regions of the South. The only aspects analysed in the paper that were not strongly affected by a geographical component were bad loan entry rates and road mortality. The conclusion that can be derived from this paper is that, if the trends of 2021 will be maintained in the following years, new support measures targeting the southern area of the peninsula should be adopted in order to improve the welfare of the country in the three areas.