REGIONAL COMPETITIVENESS ANALYSIS – A CASE STUDY ON ITALY

Analysis based on RCI 2.0 and its indicators for the Year 2022

A **Data Science Project** for the **University of Milano-Bicocca**
Department of Computer Science, Systems and Communication
Master's Degree in Data Science – Data Science Lab Course
Academic Year 2023-2024

Authors:
Nava Sara, mat. 870885, s.nava38@campus.unimib.it
Saresini Giulia, mat. 864967, g.saresini@campus.unimib.it

Index

1. Introduction

Competitiveness is defined by the Organization for Economic Co-operation and Development[1] (OECD) as the ability of a nation to produce goods and services that can compete internationally, while simultaneously maintaining and increasing the real income of its population in the long term, within a free and fair market context (OECD, 1994).

The European Commission supports EU Member States in creating sustainable and competitive economies using the **Regional Competitiveness Index** (RCI), a specifically developed indicator designed to assess and quantify the level of competitiveness of each region within the Union's member countries.

Italy has historically exhibited a **fragmented internal landscape**, characterized by heterogeneous territories with different opportunities and specific strengths and weaknesses.

But how do these various aspects contribute to defining the competitiveness level of each region? What are the most relevant factors in Italy, and how are they synthesized within the RCI?

To answer these questions, it is necessary to analyze the indicators selected by the European Commission that contribute to the construction of the RCI, identifying those that are most significant and relevant for Italy, with reference to **2022 data**. Once the predominant factors are identified, a global clustering analysis should be conducted based on all indicators to identify regions with similar levels of competitiveness. This overall clustering should then be compared with that obtained by applying the same analysis to subgroups of indicators specific to different macro-areas, with the aim of understanding whether the distribution of competitiveness in the macro-areas reflects the overall scenario or if significant differences emerge between the various regional contexts.

2. Data Overview

2.1 RCI – Regional Competitiveness Index

The **Regional Competitiveness Index** (RCI)[2,3] is a comprehensive tool developed to measure and compare the **competitiveness of regions** within the **European Union**. It's obtained through a combination of **68 indicators**, 48 of which are at the regional level, divided into **three sub-indices**: *Basic, Efficiency*, and *Innovation*. Within these sub-indices, the indicators are divided into **pillars** that describe various macro aspects of regional competitiveness:

1. **Basic Sub-Index**: It covers foundational drivers necessary for all types of economies, comprising five pillars:
   - **Institutions**: Assesses the quality and efficiency of institutions, perceived corruption, and regulatory framework.

- **Macroeconomic Stability**: Evaluates the overall economic climate and its stability, crucial for long-term investments.
- **Infrastructure**: Examines infrastructural quality, including connectivity and accessibility, vital for efficient economic functioning.
- **Health**: Focuses on the health and well-being of the workforce, which impacts productivity and participation in the labor market.
- **Basic Education**: Measures outcomes of compulsory education, highlighting the quality and effectiveness of educational systems.

2. **Efficiency Sub-Index**: Relevant to more developed economies, it includes three pillars:
   - **Higher Education, Training, and Lifelong Learning**: Reflects the role of education in enhancing productivity and supporting economic growth.
   - **Labor Market Efficiency**: Assesses the efficiency and flexibility of labor markets, which are essential for optimal resource allocation.
   - **Market Size**: Considers the scale of markets available to firms, which influences their ability to leverage economies of scale and encourages innovation.

3. **Innovation Sub-Index**: Focused on advanced stages of economic development, it comprises:
   - **Technological Readiness**: Measures the adoption and use of existing technologies by households and enterprises, crucial for competitiveness.
   - **Business Sophistication**: Indicates the level of specialization and diversification within an economy, reflecting its capacity to adapt and compete.
   - **Innovation**: Assesses the region's capacity for producing cutting-edge products and processes, a key factor in maintaining a competitive edge.

The fifth edition, known as **RCI 2.0**, introduces an improved framework to facilitate comparisons over time, with recalculated indices for 2019 and 2016, termed RCI 2.0 2019 and RCI 2.0 2016. This new version maintains the original structure but adopts enhanced methodologies to better reflect current regional dynamics and ensure clarity in historical comparisons.

2.2 RCI 2.0 – 2022 distribution in the Italian regions

Before analyzing the indicators used to construct the Competitiveness Index, it is important to provide a **preliminary overview of the distribution of the RCI 2.0 – 2022** in the **Italian regions** and **its evolution since 2016**.

It should be noted that the data provided by the European Commission for Italy includes the division of the Trentino-Alto Adige region into the two autonomous provinces of Bolzano and Trento. Therefore, in the analyses conducted in this research, a customized geographical representation of Italy is used, in which the boundaries of the Trentino-Alto Adige region are delineated by the respective autonomous provinces.

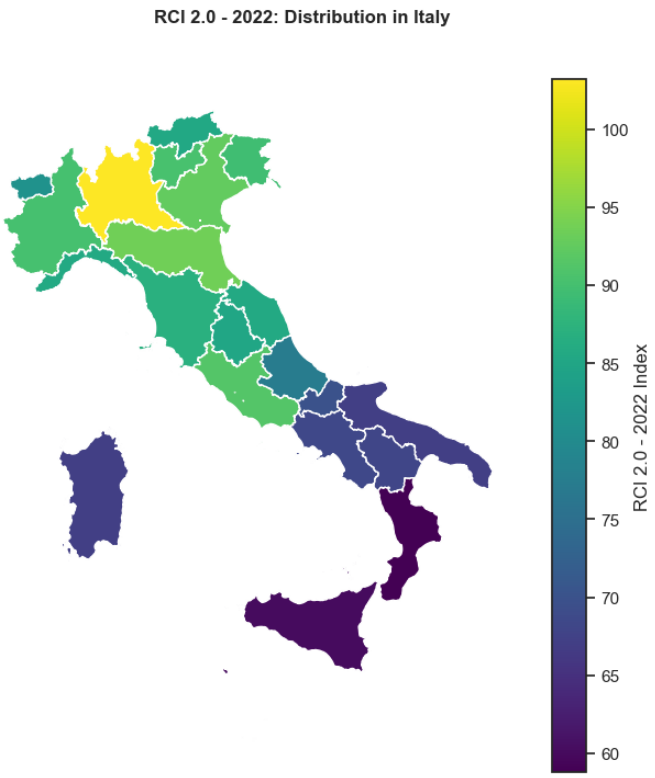

Figure 1: RCI 2.0 - 2022 Index Distribution in Italy

As illustrated in *Figure 1*, Italy presents a scenario of considerable variety: the **Regional Competitiveness Index ranges** from **58.8** to **103.2**. This wide variation within the same country highlights significant regional disparities that require thorough analysis. However, such data is not surprisingly new: the stark imbalance between Northern and Southern Italy has long been known, with the South rich in resources, capital, and labor, while the North, as Goethe remarked, acts as a "large magnet that irresistibly draws people back".

It is important to mention that 2022 was a particularly significant year, marked by the **severe effects of the COVID-19 pandemic** and the onset of the **conflict in Ukraine**. Italy was one of the countries hardest hit by the pandemic, with a particularly severe impact on the South, exacerbating poverty and slowing economic growth compared to the North. The Svimez 2021 report indicates that while the South's GDP grew by 5%, the North's GDP increased by 6.8%, with a more pronounced rise in poverty in the South. Additionally, the war in Ukraine further widened the gap, leading to lower economic growth in the South and an increase in absolute poverty. Persistent issues include high unemployment, low wages, and a significant percentage of NEET (Not in Education, Employment, or Training) youth. Although past policies have had limited success in improving the situation, the National Recovery and Resilience Plan (PNRR) represents a potential opportunity to reduce the gap, if managed effectively[4].

A closer examination of the chart reveals a sharp contrast between **Lombardia and its neighboring regions**, which, **along with Lazio**, rank among the **most competitive areas in the country**, just behind Lombardia itself. **Moving towards the center**, including regions such as Valle d'Aosta, Liguria and Provincia Autonoma di Bolzano, there is a **more gradual decline in the competitiveness index**, which

then **drops significantly in the southern regions**, with a particularly noticeable decrease in Calabria and Sicilia.

Lombardia clearly stands out as a **strong outlier in this distribution**, potentially masking less significant differences among the surrounding regions in the area highlighted in green. The choropleth map in *Figure 2* shows the distribution of the RCI 2.0 – 2022 with Lombardia excluded: the differences appear more pronounced, highlighting a group of more competitive regions in light green, consisting of Emilia-Romagna, Veneto, Lazio, Provincia Autonoma di Trento, Friuli-Venezia Giulia, and Piemonte, which seem to have Regional Competitiveness Index values above approximately 87. This group is followed by a set with an index roughly in the range of [80, 85], comprising Liguria, Toscana, Umbria, Marche, Provincia Autonoma di Bolzano, and Valle d'Aosta. Abruzzo appears as a transitional region, slightly less competitive than the previous ones but significantly superior to the southern regions. Even in this version of the chart, the southern regions do not show significant differences among themselves, with a gap of about 10 RCI points between Abruzzo and Molise, Campania, Sardegna, and Puglia, and an even more pronounced gap for Calabria and Sicilia.

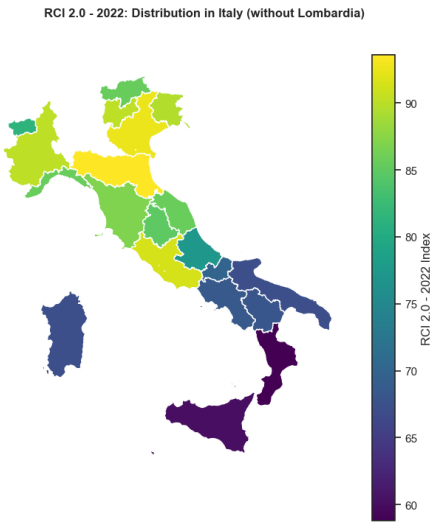

Figure 2: RCI 2.0 - 2022 Distribution in Italy, excluding Lombardia

By **separating the first two groups** (characterized by shades of green, roughly corresponding to Northern and Central Italy) from the southern group, including the islands, the significant gap becomes clear. The first group (*Figure 3*) shows an RCI variation range of 20 points, with values between approximately 80 and 100. In contrast, the second group (*Figure 4*) is not practically adjacent to the first, as the RCIs recorded fall within a narrower and distinctly separate range, oscillating between about 60 and 70. The historical considerations previously mentioned are fully confirmed by this preliminary analysis.

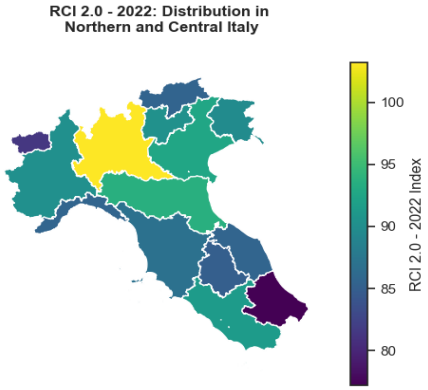

Figure 3: RCI 2.0 - 2022 Distribution in Northern and Central Italy

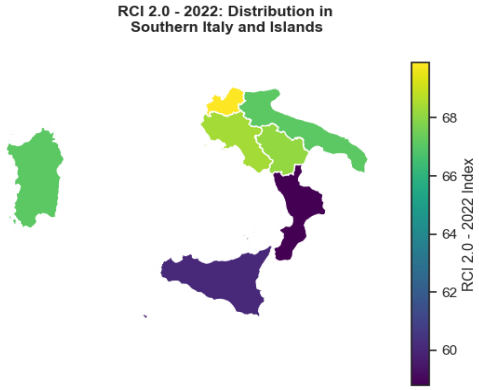

Figure 4: RCI 2.0 - 2022 Distribution in Southern Italy and Islands

2.3 RCI 2.0 over time (2016 – 2019 – 2022)

As previously indicated, versions of RCI 2.0 **for the years 2016, 2019, and 2022** have been made available. Considering the Italian context and taking into account historical analyses, it may be interesting to provide a brief description of the **regional trends of the RCI 2.0** over these years.

To obtain the regional trend, the percentage variation (**growth rate**) of the index between two years is simply computed as follows:

$$\boldsymbol{\Delta_{RCI}\%} = \frac{(RCI_{T_1} - RCI_{T_0})}{RCI_{T_0}} * 100$$

where T_1 is the latest year while T_0 is the earlier one.

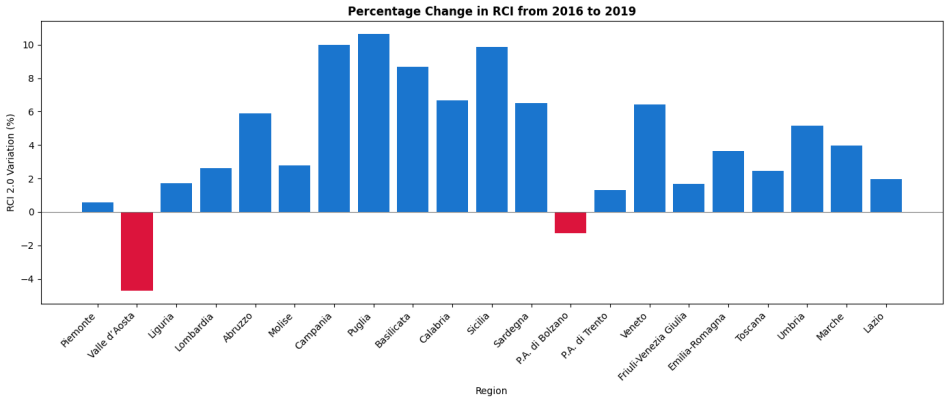

Figure 5: Percentage Change in RCI 2.0 from 2016 to 2019 for each Italian region.

As shown in *Figure 5*, a **positive trend** is observed for most Italian regions **between 2016 and 2019**, with particularly significant growth in the southern regions, which recorded higher RCI 2.0 growth rates compared to other regions.

However, in the subsequent period **between 2019 and 2022**, a **generalized decline** is observed across the Italian territory, albeit with lower absolute rates with the respect to the previous case (*Figure 5*), as shown in *Figure 6*. In this context, the **southern regions are the most vulnerable**: although they had recorded the highest growth rates for the index between 2016 and 2019, they appear to be more weakened in the second period analyzed.

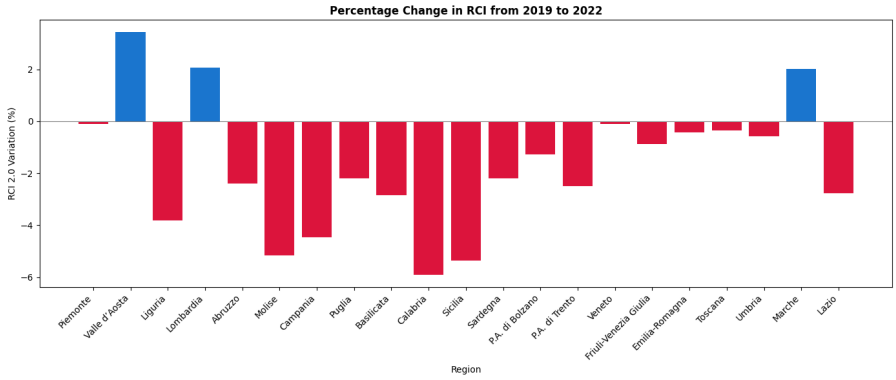

Figure 6: Percentage Change in RCI 2.0 from 2019 to 2022 for each Italian region.

This finding is consistent with previous historical observations related to the **North-South divide in Italy**, highlighting how the southern regions have been more markedly affected by the negative impacts of the COVID-19 pandemic and the conflict in Ukraine, both of which are temporally located within the second analysis period.

Examining the **extended period from 2016 to 2022** and considering the impact of the overall negative effects between 2019 and 2022, an **overall growth of the RCI 2.0** index still emerges, as highlighted in *Figure 7*. This suggests that, despite adverse factors, **Italy is making progress in competitiveness at the European level**.

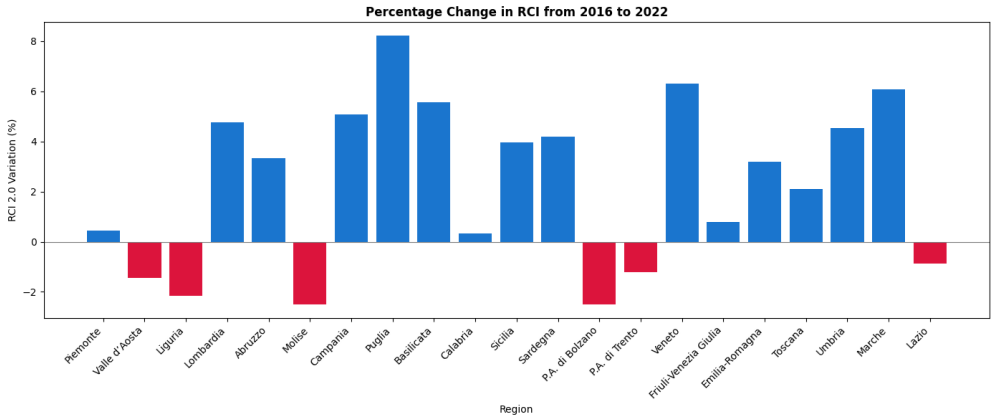

Figure 7: Percentage Change in RCI 2.0 from 2016 to 2022 for each Italian region.

However, as shown in the heatmap in *Figure 8,* it remains evident that, despite a generally uniform trend across the national territory, the **significant gap between Northern and Southern Italy** has **remained** largely **unchanged**.

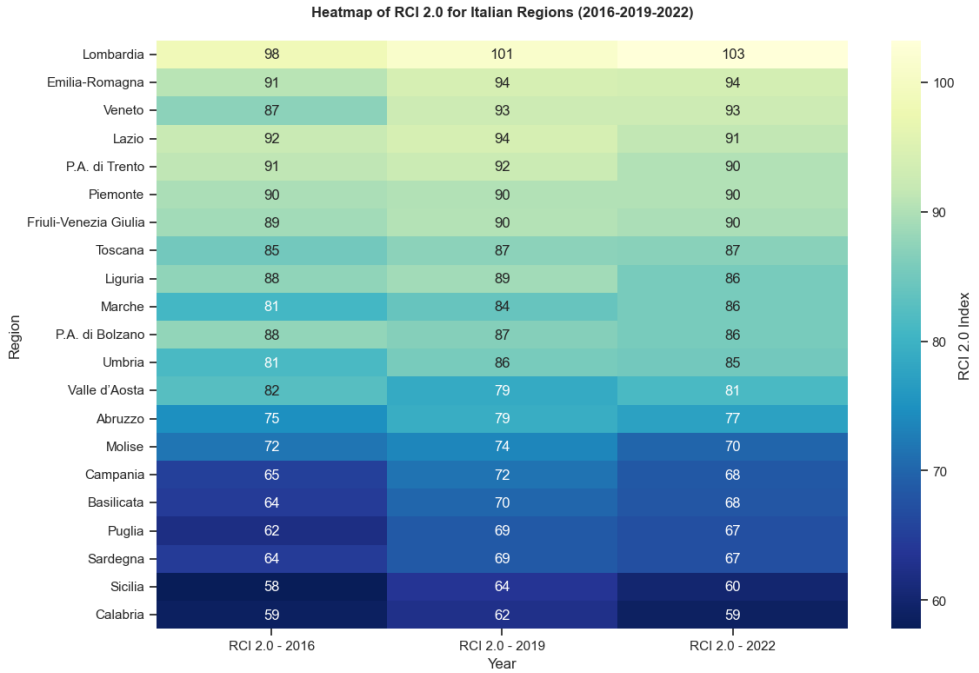

Figure 8: Heatmap of RCI 2.0 for Italian Regions (2016-2019-2022).

3. Data Pre-Processing

This section outlines the steps undertaken to **prepare the dataset for clustering**. Initially, certain variables were excluded based on key criteria, such as the coefficient of variation (CV) and correlation analysis, and in case they refer to a national – level information. After selecting the variables, data standardization was performed to ensure comparability across the indicators. A wide-format dataset was created, consolidating all information related to the retained indices. Additionally, sub-datasets, also in wide format, were generated to group the indices according to their respective Pillar categories. This process resulted in a reduction of indicators, decreasing the total from 68 to 31.

3.1 Preliminary Feature Analysis

To effectively remove highly correlated variables, **Pillar-related information** was included in the dataset. This allowed for a more precise analysis of correlations within each Pillar, ensuring decisions were based on the relevance of the indices to their specific macro area. The objective was to **avoid spurious correlations** and maintain the **integrity of the dataset** by **removing** only **redundant** or **non-essential variables**.

The dataset was first transformed from wide format to long format, incorporating Pillar information. During correlation analysis, it was observed that some **Pillars** contained **national-level** data that **did not vary across regions**. These Pillars were deemed **irrelevant for regional-level analysis** and were subsequently **excluded**. The table *Figure 9* provides a summary of the number of indicators removed within each Pillar. As a result of this process, 4 Pillars comprising a total of 20 indicators were completely removed. Additionally, 1 indicator was removed from the Innovation Pillar, reducing the total number of indicators from 68 to 47.

	Pillar	Indicator
0	Basic education	3
1	Innovation	1
2	Institutions – national	10
3	Macroeconomic stability	5
4	Technological readiness – national	2

Figure 9: Number of eliminated indicators per pillar

Following this step, **correlation matrices** were calculated for **each Pillar**. For each matrix, **indices** with a **correlation higher than 0.9** were **removed**.

As an example, the *Institutions – regional* Pillar's correlation matrix is presented *Figure 10*. The Quality and Accountability variable was excluded due to its strong correlation with Corruption and Impartiality, reflecting a relationship consistent with theoretical expectations and common sense.

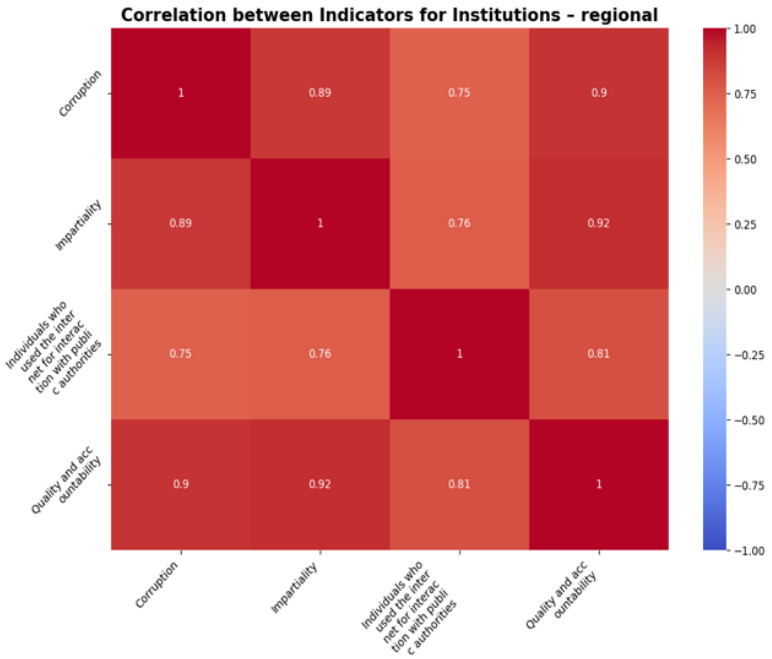

Figure 10: Correlation Matrix of Institutions – regional Pillar

3.2 Screening Based on the Coefficient of Variation (CV)

After a preliminary feature analysis based on correlations, the dataset contains 40 indicators. To further refine the set of usable indicators, it's applied a **screening based on** the **coefficient of variation** (CV). Indicators with a **CV below 0.15** were **excluded** as they were considered **insufficiently discriminative** for clustering purposes.

The bar chart above (*Figure 11*) illustrates the CVs of the indicators, and the exclusion threshold set at 0.15.

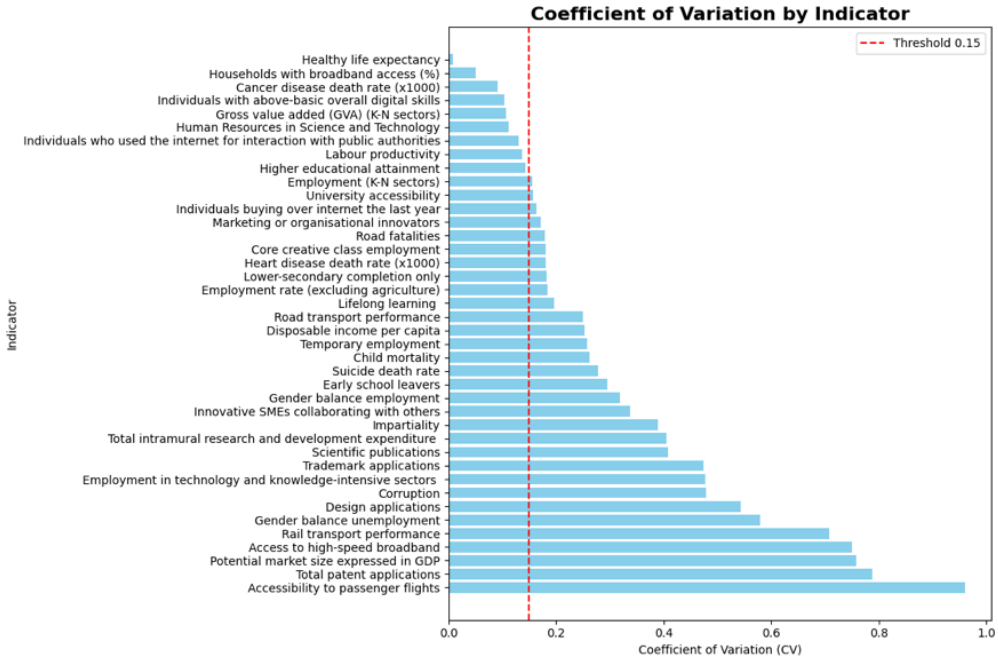

Figure 11: Coefficient of Variation for each indicator.

The following examples, illustrated at *Figure 12,* highlight how the **CV affects the ability to discriminate between regions**. For instance, the indicator Accessibility to passenger flights has a CV of 0.96, while Employment in K-N sector has a CV of 0.151, just above the threshold. Finally, the indicator Healthy life expectancy has a CV of 0.008, making it less useful for clustering due to the observed low variability.

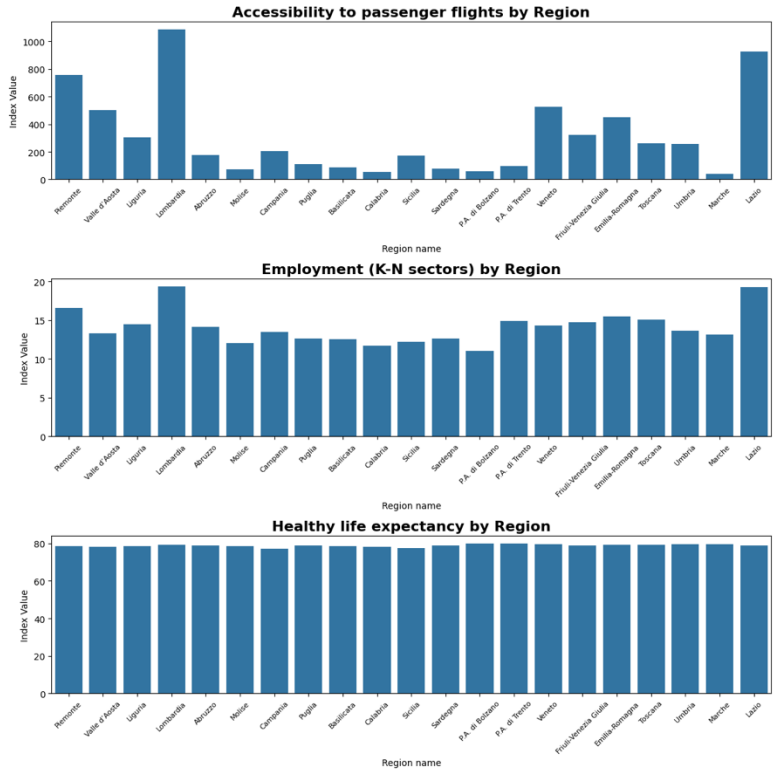

Figure 12: Coefficient of Variation (CV) for selected indicators.

3.3 Imputation of Missing Values and Data Standardization

To address the missing values in the dataset, two missing values were imputed using the **KNNImputer** method[5], as described in the scikit-learn documentation.
Subsequently, the dataset was **standardized** to comply with criteria of the selected clustering process.

4. Clustering and Results Analysis

The **Mean Shift** algorithm was employed for the clustering analysis, which identifies clusters based on data density. This method was chosen because **it automatically determines the optimal number of clusters**, eliminating the need to specify it a priori. Consequently, the analysis remains unbiased by prior knowledge, such as the visualization of regions based on their RCI values that may suggests the presence

of specific clusters, ensuring greater objectivity in the clustering process. Two clustering approaches were applied: one considering all indices and another grouping the indices by Pillar.

To evaluate the **quality of the clusters**, the **Silhouette Score** was calculated for each data point (Italian region). The graph on the right (*Figure 13)* shows the silhouette score of each region, with the regions colored by cluster. This score measures how well-separated each point is from points in neighboring clusters, providing an **assessment** of both the **internal cohesion** and the **separation of the clusters**.

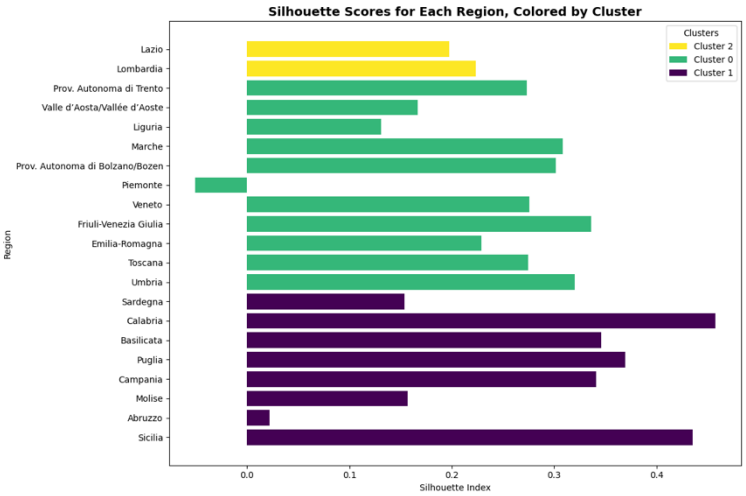

Figure 13: Silhouette score for regions grouped by cluster.

Observation of *Figure 13* suggests that the **clustering algorithm performed well overall**, except for the regions **Piemonte** and **Abruzzo**, which **appear to be mismatched** with their assigned clusters.
Possible explanations for this discrepancy will be provided in the discussion of *Figure 15* (par. 4.1), which illustrates how the regions would have been clustered if a separate clustering had been conducted for each pillar.

Before examining how our clustering algorithm has divided Italy, it is important to note that the **color assignment for clusters** is not arbitrary but is **based on a score**. The cluster with the highest score (representing the "best" values) is colored yellow, the lowest score is in purple, and the intermediate cluster is in green, complying with the *virids* color scale, already used previously for representing RCI values.

The **score** was derived as a weighted sum of the indices considered, using the average of the values of each index within each cluster. The **weight** assigned to **each index** reflects **its impact on the quality of the clustering**: it was determined as the absolute value of the difference between the overall silhouette of the clustering and the silhouette calculated excluding a specific index, normalized by the sum of the impacts of all indices, thus obtaining a standardized measure. **Each weight** was then **multiplied by** the **sign of the correlation** between each index and the RCI 2.0 - 2022, to account for whether the relationship between the index and the competitiveness score is direct or inverse.

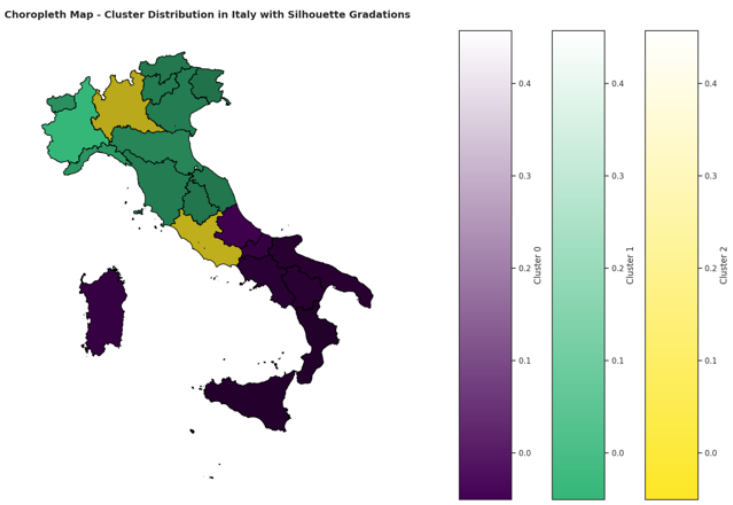

Figure 14: Choropleth map of Italy showing cluster distribution based on all indices.

Now, it is possible to analyze the clusters obtained from the algorithm used.
The map of Italy, divided according to the clusters obtained using all indices, is shown in *Figure 14*. The **color intensity** of the clusters **represents the silhouette score** of each region: lighter colors indicate regions with lower silhouette scores, and therefore a worse assignment to the cluster.

Observing the map of Italy, a clear distinction between the North and South emerges. As previously indicated by the RCI 2.0 - 2022, the **Central and Northern regions tend to be more competitive than those in the South**, where the color intensifies the further south one goes, reflecting worsening conditions. Interestingly, something not immediately apparent from the RCI, our clustering algorithm has grouped **Lazio** and **Lombardia together**. To understand the similarities between these regions, it's necessary to analyze the graph showing the clustering by pillar (*Figure 15).*
Further observations and analyses are made in the following paragraph (par 4.1).

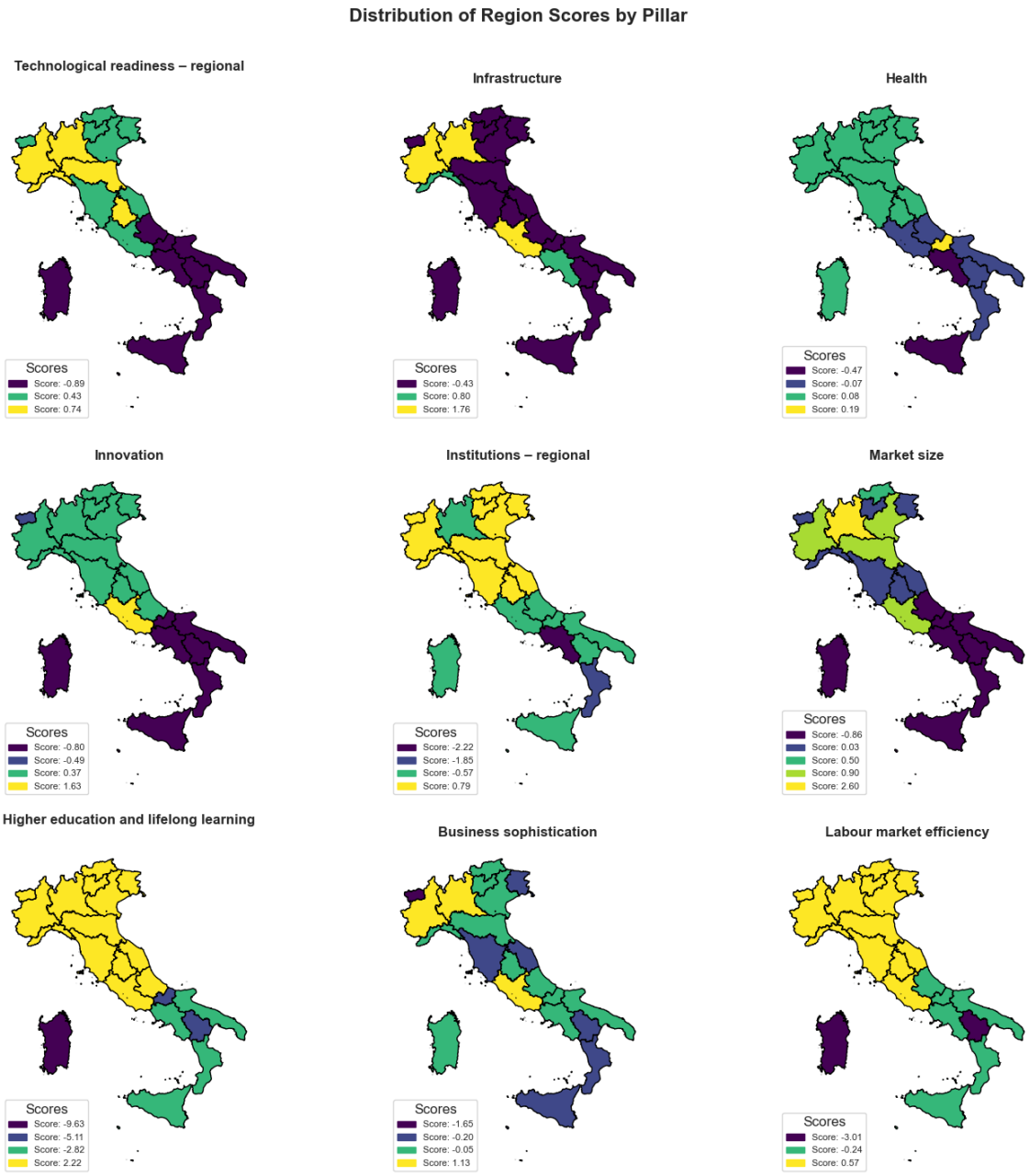

Figure 15: Choropleth map of Italy showing cluster distribution based on pillars.

4.1 Results Analysis

Looking at *Figure* 15, it was possible to make an overview analysis of the results obtained for each different pillar.

- **Technological Readiness:** Regarding technological readiness, the data reveal a division that **closely aligns with the RCI 2.0 - 2022**, identifying **three macro-groups** that generally **correspond to** the geographical divisions between **Northern, Central, and Southern Italy**. The score associated with the clusters highlights a greater similarity between the regions of Northern and Central Italy, with a clear distinction from the Southern regions, which show a significantly lower level of technological development.

  Supporting these observations, an analysis conducted by the European Commission with the report "*Broadband Coverage Europe 2021*"[6] shows that the **digital divide between Northern and Southern Italy remains a significant issue**, with notable differences in access to high-speed Internet and the quality of digital infrastructure. According to this analysis, Italy has good fast broadband (NGA) coverage available to 97% of households, but only 88.4% of households in rural areas have access to this technology. This gap is even more pronounced in the South, where in 2021, the rate of Internet access in households was 77.6%, compared to 83.3% in Central-Northern Italy.

  Moreover, the Istat report "*Territorial disparities in the NRRP: Ten goals for the South*" (mentioned inside "*Broadband Coverage Europe 2021")* highlights that 60% of residents in the South struggle to access high-speed Internet connections. This issue is exacerbated in rural and peripheral areas, where the lack of adequate infrastructure limits high-speed connectivity.

  The **disparity in Internet access** not only limits economic and social opportunities for citizens in the South but also **affects other essential areas such as education and access to digital services**, reinforcing a form of technological exclusion that requires urgent and targeted interventions.

- **Infrastructure**: From an infrastructural perspective, **Italy exhibits generally low competitiveness**, with the exception of the regions of Lombardia, Piemonte, and Lazio, which show superior performance. The infrastructure score index reveals significant discrepancies between different regional groups.

  According to an analysis by Il Corriere della Sera[7], Italy's infrastructural delay resulted in losses of 77 billion euros in exports in 2022, equivalent to 15% of the total national export value, negatively impacting key sectors of Made in Italy, such as fashion and agri-food. Italy ranks 17th globally for infrastructure quality, far from world leaders such as Singapore and major European countries. In particular, the maritime transport system is notably deficient, ranking 24th, while the rail and air networks fail to reach top positions. Italian logistics is dominated by road transport, exceeding the European average. Over the past decade, infrastructure investments have shown a yearly decrease; however, an increase is expected thanks to the National Recovery and Resilience Plan (NRRP), which will allocate 20% of the 190 billion euros planned for infrastructure projects, aiming to reduce the gap and stimulate sector development.

- **Health** - **Higher Education, Training,** and **Lifelong Learning** - **Labor Market Efficiency:** The **health gap** also raises alarm. There do not appear to be middle grounds, no intermediate clusters or isolated cases: a **clear divide between Northern and Southern Italy emerges distinctly** from the graph. The report "*Un Paese, due cure. I divari Nord-Sud nel diritto alla salute*" by SVIMEZ, presented by Save the Children[8], provides a troubling snapshot of inequalities in the Italian healthcare system. The data reveal that Southern regions are severely disadvantaged compared to the North and Central regions. The situation is particularly critical when considering indicators such as **infant mortality**, with **Sicily and Calabria recording rates nearly double those of Central-Northern regions** like Tuscany.

  Healthcare poverty is an **increasing problem**, with approximately 6.1% of Italian families struggling to cover healthcare expenses, but the situation is worse in the South, where the percentage rises to 8%, double that of the Northeast. Life expectancy is also lower in the South, with a gap of about 1.5 years compared to the North.

  The **lack of prevention and care services** is evident in cancer screening programs, with only 58% of women in the South participating in these checks, compared to 80% in the North. This gap drives many **Southern citizens to seek healthcare in Central and Northern regions**, especially for more severe conditions. In 2022, 44% of healthcare migrants came from the South, while only 0.1% of patients from Central-Northern regions moved to the South.

  Similarly to the healthcare sector, both the **labor market and education** also **show a significant gap**. The labor market in Southern Italy remains problematic, with a considerable disparity compared to the North. The National Recovery and Resilience Plan has allocated 40% of the resources to Southern regions to address this gap. Since 2000, the South has experienced a significant decline in youth employment, with about three fewer employed individuals per ten compared to Central-Northern regions. The economic crisis between 2008 and 2013 worsened the situation, and the pandemic further hindered recovery. Non-standard employment and undeclared work are more prevalent in the South, and youth employment rates are much lower compared to Northern and Central Italy. The Plan aims to reduce these disparities, but success will depend on the effectiveness of the proposed reforms[9].

  Additionally, the **educational gap between Northern and Southern Italy** is marked, with the South showing severe setbacks in educational levels. In 2020, 32.8% of Southern Italians aged 25 to 49 had only a lower secondary education, compared to 24.5% in Central-Northern Italy, while only 22.6% had a tertiary education degree, compared to 27.6% in Central-Northern Italy. Students from the South have lower skills across all subjects, with a significant gap that widens as education progresses. In 2021-22, 42.7% of Southern high school seniors exhibited very weak math skills, compared to 28.3% nationally and 15% in the Northeast. Only 6.7% achieve a very good level in math, compared to 14.9% in Italy and 22.6% in the Northeast[10].

- **Institutions – Regional:** The distribution of regional clusters in Italy, based on the level of institutions, highlights the **structural differences between the North and the South of the country**. Southern regions, such as Calabria and Basilicata, exhibit lower scores due to specific issues

that negatively impact their institutional scores. In contrast, Northern regions generally present a better institutional level, although some areas, such as Lombardy, show significant exceptions.

The trend in institutional scores can be explained through the various indicators that comprise the Pillar, which assesses the level of corruption, impartiality, accessibility, and development of institutions. To understand how the levels of these indicators have influenced regional clusters, the analysis based on the publication by the Bank of Italy has been useful[11].

The document highlights that **corruption is a significant problem** in Southern Italy. Southern regions, such as Calabria and Basilicata, are heavily penalized by the presence of organized crime, which negatively impacts the quality and impartiality of institutions ("La criminalità organizzata in Italia: un'analisi economica" p. 63[12]). In contrast, although Northern regions are not completely free from corruption, they enjoy a relatively better situation.

Regarding **institutional development**, which also includes the degree of digitalization of public institutions, **Northern regions have made greater investments** in modernizing and computerizing public services, facilitating online interactions between citizens and authorities. In the South, the adoption of digital technologies is slower, and limited Internet access in regions such as Calabria and Basilicata reflect a lower utilization of online services ("La digitalizzazione in Italia: evidenze da un nuovo indicatore regionale," p. 44[13]).

- **Innovation:** The analysis conducted on regional clusters, based on the Pillar Innovation which collects indicators related to the importance attributed to the research sector, highlights a significant gap between Northern and Southern Italy, with **Southern regions lagging significantly behind those in the North**. In particular, the article "*Ricerca e Sviluppo in Italia | Anni 2020-2022*", published by ISTAT[14], reports that Southern Italy experienced a generalized decline in R&D spending in 2020, with a reduction of -5.0% compared to the previous year. Central-Southern regions, such as Molise, Marche, and Basilicata, suffered the most severe losses, with declines exceeding -20% in Molise and Marche and -21.5% in Basilicata. Tuscany and Campania also saw significant decreases, over -10%.

  In contrast, Northern Italy showed a more stable situation, although the North-East experienced a decline in R&D spending of -5.5%. However, some Northern regions, such as the Province of Bolzano-Bozen and the Aosta Valley, have shown improving performance.

  Among the Italian regions, **Lazio stands out positively**, as demonstrated by the clustering analysis. Lazio, along with Piedmont and Emilia-Romagna, has shown **strong performance in terms of the incidence of R&D spending on GDP**. Specifically, Lazio has seen a significant increase compared to 2019, countering the prevailing negative trend in many other regions. This increase is notable compared to the performance of other Northern regions, such as Veneto and Lombardy, which, despite their historical leadership in R&D, have positioned themselves below the national average with modest growth levels.

- **Market Size:** The distribution of Market Size scores across Italian regions highlights significant **economic disparities that align with long-standing geographical divides**. The Market Size

index, which takes into account factors such as disposable income, household income in purchasing power standards (PPCS), GDP, and population, reveals a clear hierarchy that reflects the uneven development of Italy's regional economies.

At the top of the scale, regions with the highest scores tend to be located in the North, where industrialization and economic diversification have historically fostered higher levels of GDP per capita and disposable income[15]. These regions benefit from a strong manufacturing base, advanced services, and significant infrastructure investments (as demonstrated by the analysis of the other pillars), which collectively contribute to their larger market size. The population density in these areas also enhances their market potential, further increasing their economic output.

As we move down the scale, Central regions exhibit slightly lower market size scores. These regions typically have solid economic foundations, with strong service sectors and well-developed industries, but they face challenges such as slower economic growth compared to the Northern regions. While their disposable income and GDP per capita remain relatively high, these regions do not reach the same economic power as their Northern counterparts.

At the lower end of the spectrum, Southern regions struggle with significantly lower market size scores. This is largely due to lower levels of GDP per capita, weaker infrastructure, and a less developed industrial base. The South is also marked by higher unemployment rates, lower disposable incomes, and more limited economic opportunities. Population size, while not necessarily smaller, does not compensate for the economic challenges these regions face. As a result, the potential market size is much smaller in comparison to Northern and Central regions.

- **Business sophistication:** The "Business Sophistication" pillar — which refers to the ability of companies to adopt advanced and innovative management and business strategies — is the **only regional indicator that**, with the exception of leading regions such as Lombardy, Piedmont, and Lazio, **does not show a marked disparity between Northern and Southern Italy**. In fact, business sophistication clusters are distributed quite evenly across the country, demonstrating a relative uniformity in the ability of companies to adopt and implement sophisticated and modern solutions.

The various pillars considered ultimately represent **different facets of a common issue**, highlighting how the Italian situation is strongly influenced by **regional disparities**. As illustrated in Figure 14, the discrepancy between the North and South is evident, with several factors contributing in varying degrees to the creation of this structural divide. Returning to the results of our initial clustering, it becomes clear why **Piemonte** recorded a relatively l**ow silhouette score**: in several cases, it is **grouped in the same cluster as regions like Lombardy and Lazio**, reflecting a certain similarity in overall performance, which is quite strong.

The most surprising element, however, is represented by **Lazio**. Despite not appearing among the leading regions based on the RCI index, it **demonstrated** greater **competitiveness** than Piemonte, Veneto, and other areas. This result aligns with the data related to specific pillars such as innovation and infrastructure, in which Lazio stands out. It confirms that, in certain aspects, central regions can exhibit a level of development exceeding that predicted by traditional evaluation models. This underscores the **complexity**

of Italy's regional dynamics, where performance can vary significantly depending on the pillar considered, demonstrating the need for a more nuanced analysis that goes beyond mere geographical divisions.

5. Conclusions

The analysis presented in this report aimed to address the central questions outlined in the introduction: how do **regional differences in Italy influence varying levels of competitiveness**, and what are the **key factors in defining regional competitiveness** as encapsulated by the Regional Competitiveness Index (RCI)?

Through a comprehensive examination of the RCI and related indicators, this analysis has confirmed the **presence of substantial territorial disparities across Italian regions**, reflecting a long-standing North-South divide.
The primary factors contributing to these disparities are related to technological readiness, infrastructure quality, healthcare, education, labor market efficiency, and institutional strength. Northern and Central regions consistently outperform Southern regions across most pillars, particularly in technological infrastructure, economic development, and access to essential services such as healthcare and education. These differences are reflected in the overall RCI scores, with Northern regions such as **Lombardia**, **Emilia-Romagna**, and **Veneto** emerging as **leaders**, while **Calabria** and **Sicilia lag** significantly **behind**.

The clustering analysis further elucidated the geographical division, demonstrating that regions in Northern and Central Italy exhibit higher levels of competitiveness across multiple dimensions, whereas Southern regions struggle due to weaker infrastructure, poorer health services, and limited innovation capacity. A notable finding was the unexpected **strength of Lazio**, which showed **competitiveness in innovation and infrastructure**. This result suggests that Lazio might have been anticipated to cluster with other Northern and Central regions rather than being grouped with Lombardia, which is significantly distanced from the rest of Italy according to the RCI.

Regarding the indicators contributing to the profiling of Italian regional clusters, the analysis narrowed down from the initial 68 indicators to 31 that effectively capture the salient characteristics of Italian regions. The primary indicators identified include:

- **Institutions – Regional**: Corruption, Impartiality;
- **Infrastructure**: Accessibility to passenger flights, Rail transport performance, Road transport performance;
- **Health**: Child mortality, heart disease death rate, Road fatalities, Suicide death rate;
- **Higher Education and Lifelong Learning**: Early school leavers, Lifelong learning, Lower-secondary completion only, University accessibility;
- **Labour Market Efficiency**: Employment rate (excluding agriculture), Gender balance in employment, Gender balance in unemployment, Temporary employment;
- **Market Size**: Disposable income per capita, Potential market size expressed in GDP;

- **Technological Readiness – Regional**: Access to high-speed broadband, Individuals purchasing online in the last year;
- **Business Sophistication**: Employment in K-N sectors, Innovative SMEs collaborating with others, Marketing or organizational innovators;
- **Innovation**: Core creative class employment, Design applications, Scientific publications, Total intramural research and development expenditure, Total patent applications, Trademark applications, Employment in technology and knowledge-intensive sectors.

While the **RCI provides a broad overview of regional competitiveness in Italy**, it does **not fully capture the specific dynamics** and **challenges faced by individual regions**. The index offers a general picture but lacks the granularity needed to understand the unique issues and opportunities within each region. For instance, while the RCI may indicate that Southern regions are less competitive overall, it does not specify the precise areas where intervention is most needed or the specific drivers of competitiveness within each region.

Therefore, while the RCI is a valuable tool for assessing regional competitiveness on a macro scale, it is essential to complement this with more detailed, region-specific analyses. Understanding the nuances of each region's performance and challenges will enable more targeted and effective policy interventions. This nuanced approach can help identify key areas for improvement and ensure that efforts are directed towards addressing the most pressing issues in each region, rather than relying solely on aggregate scores that may obscure underlying disparities.

In conclusion, the **regional competitiveness landscape in Italy is shaped by multiple**, interacting factors that vary in significance across the country. The **overall clustering** largely **reflects** the **deep-rooted North-South divide**, but **individual pillars reveal complexities** that defy simple geographical explanations. Addressing these disparities requires an integrated, multi-dimensional approach that focuses on both enhancing competitiveness in underperforming regions and fostering sustained growth in areas that are already performing well.

6. References

1. https://www.treccani.it/enciclopedia/competitivita_(Enciclopedia-Italiana)/#
2. https://ec.europa.eu/regional_policy/information-sources/maps/regional-competitiveness_en
3. https://ec.europa.eu/regional_policy/assets/regional-competitiveness/index.html#/
4. https://www.econopoly.ilsole24ore.com/2023/05/08/distanza-nord-sud-italia/?refresh_ce=1
5. https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html
6. https://www.forumpa.it/pa-digitale/servizi-digitali/digital-divide-laccesso-a-internet-al-sud-non-per-tutti-il-ruolo-degli-operatori-di-mercato/
7. https://www.corriere.it/economia/aziende/22_dicembre_16/infrastrutture-ritardo-dell-italia-costa-commercio-77-miliardi-all-anno-6bebedb2-7d2d-11ed-93d0-fd9373385b22.shtml
8. https://www.savethechildren.it/blog-notizie/i-divari-tra-nord-e-sud-nel-diritto-alla-salute

9.https://www.corriere.it/economia/consumi/cards/divario-nord-sud-pil-all-istruzione-10-fattori-chiave-che-bloccano-mezzogiorno/gap-istruzione-occupazione.shtml).
10.https://www.corriere.it/economia/consumi/cards/divario-nord-sud-pil-all-istruzione-sanita-10-fattori-che-bloccano-mezzogiorno/istruzione-competenze.shtml
11.https://www.bancaditalia.it/pubblicazioni/collana-seminari-convegni/2022-0025/n-25_mezzogiorno_appendice.pdf).
12. n-25_mezzogiorno_append
13. n-25_mezzogiorno_append,
14. https://www.istat.it/it/files/2022/09/Report-Ricerca-e-sviluppo-2020-2022.pdf
15. https://www.statista.com/statistics/793266/gdp-in-italy-by-region/
16. Materials:
https://drive.google.com/drive/folders/178SREoVB8Vdvuq0evd8yC7hTYk9IYuu8?usp=sharing