Exploiting EU integration indicators at infra-national level: Which regions are comparable?

Pilot Report: Targeted Technical Support to Implementation of Action 'Facilitating Evidence-Based Integration Policies in Cities'

ANNE-LINDE JOKI
Service carried under guidance of Migration Policy Group

Page Break

1. Introduction and objective

Scholars and policy makers have long focused predominantly on the national level when analysing migration and integration policy. While national-scale outcomes reflect and inform the development of national integration policies and enable cross-national comparisons, but they fail to adequately capture spatial differences within countries. This being said, it is increasingly recognised that while we often think of immigrants as moving from one country to another, really they arrive from a particular place and settle in a particular community.

As a key step towards better availability of comparable integration on urban-regional level and as outcome of the Urban Agenda Action on Facilitating evidence-based integration policies in cities, EUROSTAT has been testing the possibility of publication, to the widest possible extent, of the existing EU integration indicators on NUTS-2 level and by 'degree of urbanisation' (cities, towns and suburbs, rural areas). The feasibility testing has resulted to the recent publication of new indicators for most classic and robust indicators as part of the Eurostat migrant integration database (employment regional series). Activity rate, employment rate, unemployment rate are now available to be disaggregated by country of birth and country of citizenship at regional level (NUTS-2) and by degree of urbanisation (cities, towns and suburbs, rural areas).

The second phase of the data feasibility regarding a new regional education series resulted in the publication of the infra-national statistics for educational attainment and young people neither in employment nor in education or training (NEET) that are now available to be disaggregated by country of birth and country of citizenship at regional level (NUTS-2) and by degree of urbanisation. Since the publication of the Options Report of the Action's Stakeholder Working Group EUROSTAT has continued feasibility testing, which has resulted in the publication of LFS-based demographic data on regional level.

The analyses and results presented in this paper are a first step to showcasing the newly available comparative data on infranational level, in making meaningful comparisons in education and labour market integration outcomes across cities and regions. The overall aim of this exercise is to understand how EU regions (NUTS-2) differ concerning integration outcomes of migrants, and how certain characteristics influence the migrants' integration outcomes at the regional level.

One of the main objectives of this research is the creation of a classification of NUTS2 regions, which reflects the heterogeneity of NUTS2 (1) Integration outcomes and (2) characteristics in the EU. The rationale behind this is that the clusters provide optimal groups of similar units allowing NUTS2 regions within clusters to learn more easily from those that are more similar. Closer similarity may show regions how to achieve the changes they seek in the most efficient way. The NUTS-2 region clusters are 'taxonomized' by combining information about the clustering variables with the dynamics of the clustering variables across different clusters. Finally, regression analysis is performed in order to explore which regional characteristics best explain integration outcome differences between the NUTS-2 regions and clusters.

All integration outcome variables and regional characteristics included in the analysis are described in Section 2 with an assessment of the availability. In Section 3 the methodological approaches utilised for the analyses are presented. Section 4 examines the results for each approach and some concluding remarks are provided in Section 5.

2. Data and variables

Because of the difficulties in defining what integration means, extensive effort has been expended on defining what immigrant integration might look like. Specifically, the focus has been on ways of measuring integration, with an emphasis on quantification of educational and employment outcomes. The main data source for comparable educational attainment and employment statistics, is the EU labour force survey (LFS). The LFS is a large quarterly sample survey that covers the resident population aged 15 and above in private households. Migrant indicators are calculated for two broad groups: the foreign population determined by place of birth and the foreign population determined by citizenship. While in this this report we, at times, highlight the findings for the latter analyses have been run consistently for both broad groups.

All indicators are considered at NUTS2 level in line with Regulation (EU) 2016/2066 amending annexes to NUTS Regulation 1059/2003, meaning 281 NUTS3 regions are included in the analysis. It should also be noted that some EU Member States have a relatively small population and may therefore not be subdivided at some (or even all) of the different levels of the NUTS classification. For example, five of the Member States — Estonia, Cyprus, Latvia, Luxembourg and Malta — are each composed of a single NUTS level 2 region according to the 2016 version of the NUTS classification.

2.1. Integration indicators

The Integration Indicators included in the report are five of the official education and employment 'Zaragoza' integration indicators. The Zaragoza indicators have been widely used to identify successes or challenges in the process of immigrant integration at the national level. These outcomes and indicators have been chosen to allow for comparability across EU member states.

The integration indicators are operationalised by calculating comparisons between national and foreign populations (gap), with the both groups subdivided into people who are citizens of another EU Member State and non-EU citizens.

The age categories applied in this report are in line with those employed by the EC for dissemination of Integration Indicators.

1. Activity rate

Activity rate is defined as the percentage of the population in a given age group who are economically active. According to the definitions of the International Labour Organisation (ILO) people are classified as employed, unemployed and economically inactive for the purposes of labour market statistics. The economically active population (also called labour force) is the sum of employed and unemployed persons. Inactive persons are those who, during the reference week, were neither employed nor unemployed.
Age: 20-64 Year: 2018 Data source: EU Labour Force Survey

2. Employment rate

The employment rate is calculated by dividing the number of persons aged 20 to 64 in employment by the total population of the same age group. Employed population consists of those persons who during the reference week did any work for pay or profit for at least one hour, or were not working but had jobs from which they were temporarily absent.
Age: 20-64 Year: 2018 Data source: EU Labour Force Survey

3. Unemployment rate

Unemployment rate represents unemployed persons as a percentage of the labour force.
The labour force is the total number of people employed and unemployed. Unemployed persons comprise persons aged 15 to 74 who were: a. without work during the reference week, b. currently available for work, i.e. were available for paid employment or self-employment before the end of the two weeks following the reference week, c. actively seeking work, i.e. had taken specific steps in the four weeks period ending with the reference week to seek paid employment or self-employment or who found a job to start later, i.e. within a period of, at most, three months. Age: 15-74 Year: 2018 Data source: EU Labour Force Survey

4. NEET (Young people neither in employment nor in education and training)

NEET corresponds to the percentage of the population of a given age group and sex who is not employed and not involved in further education or training. The numerator of the indicator refers to persons who meet the following two conditions: (a) they are not employed (i.e. unemployed or inactive according to the International Labour Organisation definition) and (b) they have not received any education or training (i.e. neither formal nor non-formal) in the four weeks preceding the survey. The denominator in the total population consists of the same age group and sex, excluding the respondents who have not answered the question 'participation in regular (formal) education and training'.
Age: 15-34 (deviate from EC applied age group 15-24 due to data unavailability)
Year: 2018 Data EU Labour Force Survey

5. Share of tertiary educated

The indicator is defined as the percentage of the population aged 30-34 who have successfully completed tertiary studies (e.g. university, higher technical institution, etc.). This educational attainment refers to ISCED (International Standard Classification of Education) 2011 level 5-8 for data from 2014 onwards and to ISCED 1997 level 5-6 for data up to 2013. Age: 30-34 Year: 2018 Data source: EU Labour Force Survey

2.2. NUTS 2 descriptive variables

2.2.1 Regional typology

NUTS2 regions have been classified into Predominantly Urban, Intermediate and Predominantly Rural to take into account geographical differences among them. The OECD regional typology is applied and it is based on criteria of population density.
Definition: 1. The first step of the methodology consists in classifying each NUTS 3 as rural if their population density is below 150 inhabitants per square kilometre.
2.The second step consists in aggregating this lower level (NUTS 3) into NUTS2 regions and classifying the latter as “predominantly urban”, “intermediate” and “predominantly rural” using the percentage of population living in rural lower levelunits (local units with a population density below 150 inhabitants per square kilometre). NUTS 2 regions are then classified as:

- Predominantly Urban (PU), if the share of population living in rural local units is below 15%;
- Intermediate (IN), if the share of population living in rural local units is between 15% and 50%;
- Predominantly Rural (PR), if the share of population living in rural local units is higher than 50%

Results from this classification are presented in Table 1

Table 1. NUTS2 classification by urban/rural predominance

Classification	**Frequency (number of NUTS2 regions)**	**Percentage**
1: Predominantly urban	99	35.2
2: Intermediate Region	45	16.0
3: Predominantly rural	137	48.8
Total	**281**	

2.2.2 Regional gross domestic product (PPS per inhabitant in % of the EU28 average)

GDP (gross domestic product) is an indicator of the output of a region. It reflects the total value of all goods and services produced less the value of goods and services used for intermediate consumption in their production. Expressing GDP in PPS (purchasing power standards) eliminates differences in price levels between countries. Calculations on a per inhabitant basis allow for the comparison of economies and regions significantly different in absolute size. GDP per inhabitant in PPS is the key variable for determining the eligibility of NUTS 2 regions in the framework of the European Union's structural policy. Year: 2016-2017

2.2.3 Net migration

Crude rate of net migration including statistical adjustment is the ratio of the net migration including statistical adjustment during the year to the average population in that year. The value is expressed per 1000 inhabitants. The crude rate of net migration is equal to the difference between the crude rate of population change and the crude rate of natural change (that is, net migration is considered as the part of population change not attributable to births and deaths). It is calculated in this way because immigration or emigration flows are either not available or the figures are not reliable.
Year: Average 2016-2017 (2017: UKM7, UKM8, UKM9, IE04, IE05, IE06) Data gaps: none

2.2.4 Population size

Population on 1 January should be based on concept of usual resident population, i.e. the number of inhabitants of a given area on 1 January of the year in question (or, in some cases, on 31 December of the previous year). The population figures can be based on data from the most recent census adjusted by the components of population change produced since the last census, or based on population registers.
Year: average 2016-2018 Data gaps: none

2.2.5 Foreign-born population

This indicator is measured as a percentage of population. The foreign-born population covers all people who have ever migrated from their country of birth to their current country of residence. The foreign-born population captured in this indicator include people born abroad as nationals of their current country of residence.
Year: Average 2016-2018 (Exceptions: PL22 2017-2018, PL42 2018, PL61 2018, PL71 2017-2018, PL81 2017, RO32 2017-2018).
*Note: This indicator is calculated at hand of demographic data for the age group 15-64. The reason for this is that population data disaggregated by Country of Birth at NUTS2 is only available for the age-group specified.

2.2.6 Regional Competitiveness Index (RCI)

Definition: The EU Regional Competitiveness Index (RCI) is the first composite indicator which provides a synthetic picture of territorial competitiveness for each of the NUTS 2 regions of the 28 EU Member States. The definition of competitiveness used by the EC for RCI ('the ability of a region to offer an attractive and sustainable environment for firms and residents to live and work') takes into account both business success and personal well-being.

Operationalisation: The RCI is based on the methodology developed by the World Economic Forum. The indicators are followed within 11 pillars that describe both inputs and outputs of territorial competitiveness. The 11 pillars are grouped into three sub-indices, which are basic (five pillars), efficiency (three pillars), and innovative (three pillars) factors of competitiveness.

Year: RCI values are published at three-year intervals however, it should be noted that a number of indicators differ across RCI editions 2010, 2013, 2016 and 2019. For the purpose of this research we use data published in 2019.
Data gaps: none

3. Methodology

Cluster Analysis (CA) covers a rather wide collection of statistical methods that can be used to assign cases, i.e. records or units (here, NUTS2 regions), to groups (clusters) that are mutually exclusive. Group members will share some properties in common, so that the degree of associations is strong between cases of the same clusters and weak between cases of different clusters. Each cluster thus describes the grouping to which its members belong. The resultant classification can then provide some insights and help for the interpretation of integration outcomes because it may reveal associations between regional characteristics and integration outcomes. The clusters themselves may, in turn, contribute to the definition of classification of regions or even suggest models with which to describe a grouping of regions.

While there is no such thing as a single correct classification, we aim to find the 'optimal classification', which is defined by the fitness tests. Ultimately we wish to discriminate between variables that are not used to implement the CA and if those discriminations are useful, then results from the CA are useful. The first step is to decide which clustering variables will be included in the analysis; one should avoid using an abundance of clustering variables, as this increases the probability that they are no longer dissimilar. If there is a high degree of collinearity between clustering variables, they cannot have the power to discriminate between groups. In this paper we will explore two separate clustering models, as described in the section below.

3.1. Cluster analysis of integration outcomes across NUTS2 regions

On the one hand we wish to develop a taxonomy of regions at the hand of four integration outcome indicators, as described in the previous section. The ambition is to investigate whether integration outcomes reveal a clear pattern of differences between regions. This amounts to the questions whether it is possible to distinguish groups of regions that have comparable patters of integration outcomes.

3.2 Cluster analysis of regional characteristics

In the same vein, we set out to group regions at the hand regional level characteristics that empirical work has identified as significant factors influencing migrant integration outcomes. Closer similarity may show regions how to achieve the changes they seek in the most efficient way. The NUTS-2 region clusters are 'taxonomized' by combining information about the clustering variables with the dynamics of the clustering variables across different clusters.

3.3 Differences in integration outcomes based on regional characteristics

At the hand of the second clustering model (NUTS 2 regions clustered by regional descriptive characteristics) a one-way between subjects analysis of variance (ANOVA) is conducted, in order to compare the effect of Cluster on the five integration outcome variables at NUTS2 regional level. By conducting this test we are able to determine whether the groups of regions differ significantly in terms of the average score on each of the integration outcomes considered in this report.

4. Results

4.1 Regional disparities and similarities in integration outcomes

As an exploratory analysis, we run a cluster analysis on the basis of the selected migrant integration outcome indicators. NUTS2 regions are clustered at the hand of TCN- & EU-national integration outcomes (on the basis of citizenship). Cases are excluded listwise, on the basis of missing data. While this has a significant effect on the sample size (N=58), the assumption that data is missing at random and thereby to cluster the complete set of NUTS2 regions, on the basis of the data that is available, cannot be made.

The analysis has been replicated at the hand of TCN-born & EU-born integration outcomes (on the basis of country of birth). The results of the cluster analysis are comparable to those obtained by country of citizenship. Keeping in mind that this is an exploratory analysis in order to explore the data, in an effort to avoid duplication this section will be limited to the finding at the hand of the integration outcomes for EU28 nationals and TCNs.

The set of variables that we consider includes five quantitative variables: Activity gap, Employment gap, Unemployment gap, NEET gap and Tertiary education attainment gap.

4.1.1 Selection of Clustering variables

A preliminary analysis considering the relationships among the variables is considered, with the aim of having a first glance at potential patterns in the data. The correlation between the Employment Gap and Activity Gap measures is very high for both TCNs (.915**) as for EU28-nationals (.850**). The correlation between these two variables is high enough to cause problems in the cluster analysis. Considering issues of data availability and the higher number of valid cases for the measure of Activity Gap, an initial decision is made to retain Activity Gap in the clustering analysis and to exclude ‘Employment gap’.

However, the results of the first model reveal that that ‘Activity gap’ does not significantly differentiate the membership of NUTS2 regions in a 2-cluster model (see Annex 2). For this reason, we will include ‘Employment rate’ as a clustering variable for the clustering models that follow

4.1.1.2 Validation of a 3 Cluster Model

Several models are run in order to obtain the most meaningful number of clusters, at hand of the available integration outcomes, as specified above. Evaluation of the model specifications pleads for 3 meaningful clusters of NUTS2 regions. In the following section, we report the output obtained with K-Means clustering, which accounts for 3 clusters.

Table 2 shows final cluster centers: representing the average in each cluster. At first glance we observe that, generally speaking, the largest gaps between foreign nationals (EU & TCN) and nationals are found in Cluster 3 and that the largest gaps between EU28-nationals and TCNs are found in cluster 1. In Table 3, the number of cases in each cluster is reported. Cluster 1 and 2 are roughly the same size, while cluster 3 is includes double the number of NUTS2 regions. While one traditionally strives to obtain equally sized clusters, it should be noted that the majority of NUTS2 regions are excluded from analysis due to missing data. Furthermore, a small cluster does not need to mean that it is not meaningful but potentially that it is underrepresented.

Table 2. Final Cluster Centres

	Cluster		
	1	2	3
NEET_gap_EU28NAT	4.85	1.79	12.23
NEET_gap_TCNNAT	15.63	8.48	18.44
Unemp_gap_EU28NAT	1.93	-.54	6.02
Unemp_gap_TCNNAT	11.68	2.32	10.37
Tert_gap_EU28NAT	-5	-2	12
Tert_gap_TCNNAT	11	-7	15
Empl_gap_EU28NAT	1.25	-5.57	2.56
Empl_gap_TCNNAT	21.46	4.66	11.11

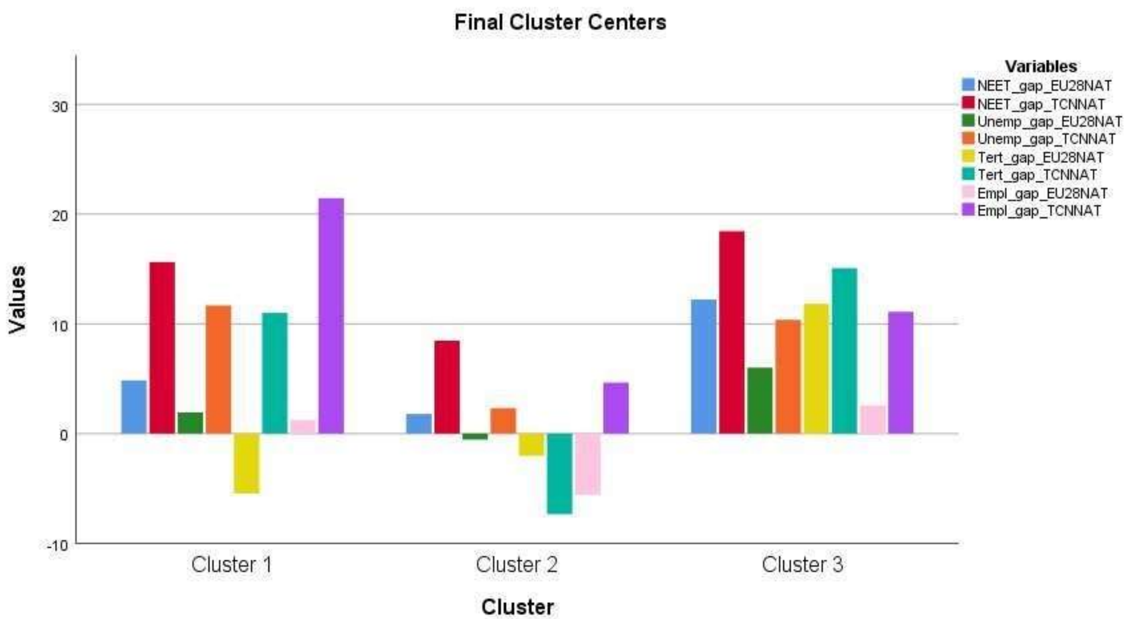
Table 3. Number of cases in each cluster

Cluster	1	15.000
	2	12.000
	3	31.000
Valid		58.000
Missing		223.000

4.1.1.3 Cluster Taxonomy

In the section below, features of the three clusters are discussed by combining information about the clustering variables with the dynamics of the clustering variables across the different clusters (Figure 1). we set out to detect patters across the Integration Indicators in order to characterise the similarities between the NUTS2 regions clustered together and the differences between NUTS2 regions clustered into different groupings.

Figure 1. K-means Integration Outcomes, 3 Cluster Comparison



Cluster 1 is characterised by both EU28 nationals and TCNs finding themselves in a disadvantaged position to natives when it comes to the Employment, Unemployment and NEET. While we observe a significant positive tertiary education attainment gap between TCNs and natives, EU28 nationals are more often higher educated than the natives. While the general trend for TCNs and EU28 nationals is the same across the NUTS2 regions in Cluster 1, it can be concluded that TCNs are significantly more disadvantaged than EU28 nationals, with gaps between EU28 nations and natives being much smaller. While interactions between variables fall outside the scope of this paper, we observe that on average, the share of tertiary educated EU28 nations is higher than that of natives, possibly contributing to closing the employment gap between EU28 nationals and natives in these regions.

NUTS-2 regions of cluster 1 include, among others: Vienna, Brussels, Antwerp, Oberbayern, Berlin, Catalonia, North Holland, South Holland, Stockholm, South Sweden

Cluster 2 is characterised by outcomes on the integration indicators much closer to that of natives. On average, foreign nationals are much less disadvantaged as compared to Clusters 1 and 3. On some measures foreign nationals even outperform natives. Furthermore, in contrast to what we observe in Cluster 1, the gap between EU28 nationals and TCN is also much smaller in this cluster. On average, EU28 nationals living in NUTS2 regions assigned to Cluster 2 are more advantaged in terms of employment and unemployment, as compared to natives. Furthermore, the analysis reveals that on average foreign nationals in Cluster 2 are higher educated than natives, possibly explaining the narrowing of them employment and unemployment gap.

NUTS-2 regions of cluster 2 include, among others: Cyprus, Canary Islands, Southern Ireland, East and Midland Ireland, Luxembourg, Malta, West Midlands, Inner London-East, Outer London East and North East & West and North West

EU28 national and TCNs in **Cluster 3** display a significant NEET gap, employment gap, unemployment gap and tertiary level education gap as compared to natives. However, compared to Cluster 1 where EU28 national outperform TCNs, in Cluster 3 we observe comparable integration outcomes for EU28 nationals

and TCN. EU28 nationals and TCN are, so to say, equally 'disadvantaged' as compared to national citizens in these regions.
NUTS-2 regions of cluster 3 include, among others: Hainaut, Karlsruhe, Darmstadt, Köln, Attica, Galicia, Basque Community, Madrid, Valencian Community, Helsinki-Uusimaa, Île de France, Piemonte, Valle d'Aosta, Liguria, Lombardia

4.2 Cluster analysis based on NUTS2 regional characteristics: K-means analysis

The set of variables that we consider includes five quantitative variables (GDP in PPS, Net migration, population size, share of foreign born and RCI) and one qualitative variables (rural/urban typology). The traditional cluster analysis is only feasible with quantitative variables, since they are based on the calculation of a distance matrix.

A rough solution to obtain a preliminary analysis of our variables is to recode and to interpret the qualitative variables as an ordinal variable:
- Proxy of urban index: 0=rural; 1= intermediate; 2=urban
Clearly, results can only give an approximation of the complexity of the situation. An initial 2-cluster model including the complete set of clustering variables reveals 'Total Population' does not significantly differentiate between the Clusters. 'Total Population' is therefore excluded from the subsequent analysis. Additional analyses indicate a 2 Cluster model to be the best fit.

In the following tables, we report the output obtained with K-Means clustering, which accounts for 2 clusters. This algorithm assigns cases to clusters based on the smallest amount of distance between the cluster mean and each case. This is an iterative process that stops once the cluster means do not significantly change in successive steps. The output of K-Means is provided in the following tables.

Table 5. Final Cluster Centers

Clustering variables	Cluster 1	Cluster 2
Z_ShareFB	.48420	-.70768
Z_Net Migration	.59412	-.75352
Z_GDP_PPS	.55611	-.73282
Z_RCI_2019	.65285	-.82660
Urban_Rural	1.27	.35

The final cluster centers are computed as the mean for each variable within each final cluster. The final cluster centers reflect the characteristics of the typical case for each cluster. The absolute value of the z-score tells us how many standard deviations it is from the mean. If a z-score is equal to 0, it is on the mean. A negative z-score reveals the raw score is below the mean average.
In Table 6, the number of cases in each cluster is reported. The clusters are relatively equal in size, which supports the validity of the cluster structure.

Table 6. Number of cases in each cluster

Cluster	1	157
	2 124	
Valid 281		

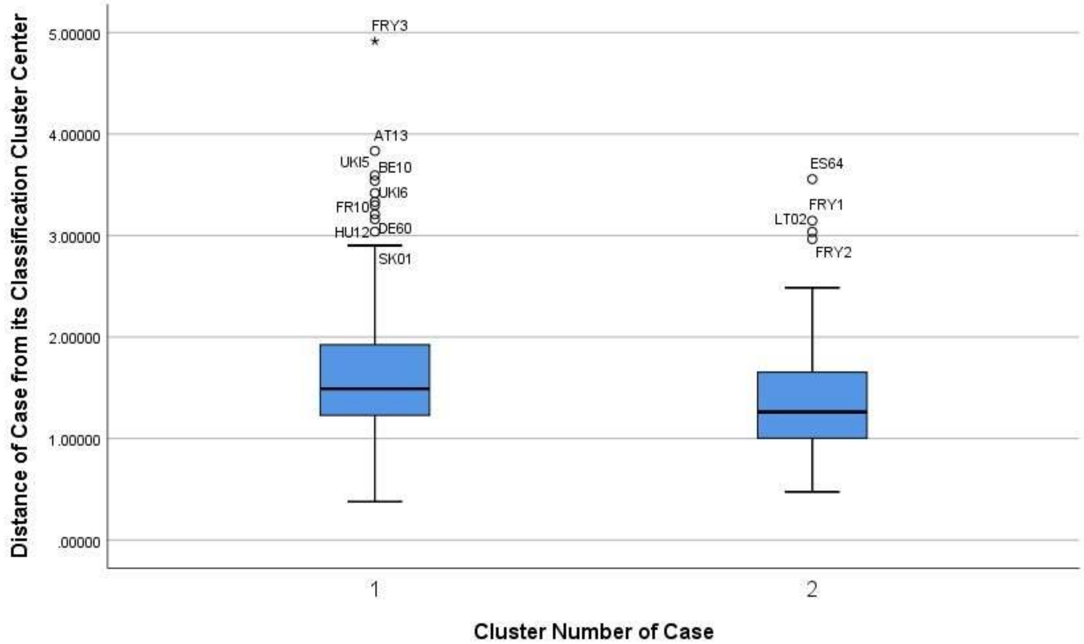
Kruskal-Wallis test confirms that all variables have significant power to discriminate between the clusters (Table 7). An additional inspection of the fit of the 2-cluster model can be made by inspecting the distances between the final cluster centres (Figure 2). The diagnostic plot reveals several outliers within each cluster. Furthermore, there is a lot of variability in cluster 1. This suggests that although a 2-cluster model is satisfactory, additional meaningful groupings may be missed in the current cluster model.

Test Statistics[a,b]

	Share of foreign-born	Net migration	GDP	RCI	Urban/Rural typology
Kruskal-Wallis H	103.029	149.094	151.189	158.318	70.920
df	1	1	1	1	1
Asymp. Sig.	.000	.000	.000	.000	.000

a. Kruskal Wallis Test
b. Grouping Variable: Cluster Number of Case

Figure 2. Plot of Distances from Cluster Center by Cluster Membership



4.3 Detecting differences in integration outcomes between clusters

A one-way between subjects analysis of variance (ANOVA) was conducted to compare the effect of Cluster on activity gap, employment gap, unemployment gap, NEET gap and the share of tertiary educated NUTS2 regional level. The analysis was conducted for foreign-born (TCN-born and EU28-born) and foreign nationals (TCN and EU28 national) separately.

In the case of TCN-born, the analyses reveal a significant effect of cluster membership on Activity gap, Employment gap and NEET gap of TCN-born. No significant effect of cluster membership is found on the level Tertiary educational attainment gap of TCN-born. Generally speaking, the NEET and Unemployment gap is significantly lower for TCN-born living in NUTS 2. Employment and Activity rate of TCN-born are more similar to those of the native-born in cluster 2. There is no significant difference in Tertiary attainment gap between TCN-born and native-born in NUTS 2 regions in cluster 1 and cluster 2.
On the whole, these results suggest that NUTS2 regions with similar characteristic have comparable integration outcomes.

5. Discussion of results

To date, the national scale has been taken as the starting point for assessing immigrant integration outcomes, and an emphasis has been made on differences and similarities between countries. As a consequence of this activity, immigrant integration outcomes—particularly as shown by large-scale data sets—are primarily measured at the national scale and understood as a consequence of national characteristics, policies and practices.

The analyses and results presented in this paper are a first effort to exploit the newly available Eurostat data on infranational level, in line with the Partnership's stated overall goal of 'better data'. Using the newly available NUTS2 data, we show how a focus on place and scale provides a more nuanced understanding of immigrant integration outcomes and of the process of integration. By doing so, we have been able to highlight the data's potential for assessing cities' integration outcomes in a comparative way and their usefulness for data practitioners. The Zaragoza indicators have been widely used to identify successes or challenges in the process of immigrant integration at the national level. In this paper we applied different methods in order to classify NUTS2 regions by integration outcomes on the one hand and NUTS2 regional characteristics on the other.

The findings suggest that if the assessment of immigrant integration outcomes continues to rely primarily on data aggregated to the national scale, we will miss important spatial differences and likely not adjust integration policies and practices to better support specific regional needs. The analysis reveals that NUTS2 regions can be clustered in a meaningful way, at the hand of the labour market and educational situation of foreign citizens residing in the region. Three clusters emerge: one is formed of NUTS2 regions where EU28 foreign citizens fair significantly better compared to TCN, a second Cluster emerges which is made up for regions where employment and education outcomes of foreigners are much closer to those of natives, and a last cluster of regions in which EU28 citizens and TCN are equally disadvantaged, as compared to natives (with exception of employment rate).

This being said, the model in question is estimated at hand of only 58 of the 281 NUTS2 regions, due to gaps in data availability across the integration outcome indicators. In contrast to what may be expected (on the basis of availability of national level integration indicators) the large number of missing cases from the analysis is due to data gaps on a handful of indicators disaggregated for EU foreign nationals. The most problematic indicators are NEET rate and Unemployment rate, with only 84 and 70 valid cases respectively.

While the two most extreme cases are highlighted above, it should also be noted that significant data gaps are observed for TCN, likewise. While the key advantage of LFS data is that they come from a survey which is highly harmonised and optimised for comparability, there are some limitations when considering the coverage of the LFS for various populations based on citizenship, as the survey was designed to target the whole resident population and not specific subgroups. Given that the LFS is a sample survey, it is possible that some of the results presented for labour market characteristics analysed by citizenship are unrepresentative or of low reliability, especially in EU Member States with small populations of foreign citizens. It is paramount that large-scale data sets are designed to provide the means for a spatialized investigation of immigrant integration outcomes, as this aspect is often overlooked in the presentation of publicly available data. Although data on immigrants are often made available in a variety of ways, it is rarely provided at a regional scale. A concerted effort is needed to insist on the importance of making data on immigrant integration publicly available at a range of spatial scales.

Turning to NUTS2 regional descriptive data at the hand of which factors enabling successful migrant integration can be explore, data availability is much less of an issue. An exception is the disaggregation possibilities of NUTS2 demographic data. While the intention had been to account for the total share of the foreign-born population in each NUTS2 region in our multiple regression models, we were only able to reliably calculate this for a sub-group (age 15-64) of the population.

A point that deserves and requires additional attention is a feasible operationalisation of a ruralurban typology for NUTS2 regions. Various European countries have different ways of classifying urban and rural areas. As a consequence, these classifications are specific to the countries concerned and therefore not strictly comparable across countries. Based on the OECD regional typology,

Eurostat has developed a rural-urban typology for NUTS 3 regions to cover all countries of the European Union. However, Eurostat does not publish an urban-rural typology at NUTS 2 level, while there is a need of such data at NUTS 2 level (think of regional policy instruments). On average, urban residents have better access to education, health care and transportation than rural populations and thereby urban-rural differences are relevant for integration outcomes. While we agree with Eurostat's argumentation that an identical application of an urban-rural typology at NUTS 2 could hide significant differences at a low regional level (also due to the large variations in NUTS 2 region size – think of EU MS consisting of a single classification) nevertheless, an urban-rural typology of NUTS 2 regions would be useful and would limit data users constructing and applying proxy measures that do not undergo an assessment of validity.

Future considerations

The assessment of immigrant integration outcomes at the NUTS2 scale provides an alternative insight into the process of integration and highlights important function of showing spatial differentiation in integration outcomes using existing data sets. This paper illustrates the key difference that place makes in understanding the process of immigrant integration. While we begin to approach the question of integration using a more nuanced spatial perspective, it is important to acknowledge and explore the interplay between various scales and different dimensions of place, which in turn create or challenge barriers to inclusion. Additionally, the national integration policies in Europe differ vastly between the EU countries, as well as over time. Therefore, not only the national legal framework of integration policy, but also the current political and cultural debate on the integration of migrants at different levels should be included. Having established marked differences in immigrant integration outcomes across NUTS2 regions, future work should aim to highlight the different national and regional barriers to enabling integration.