

# Mismatch between jobs and skills in the EU

João Lopes<sup>1</sup>, Marco Moura<sup>2</sup>, Sónia Quaresma<sup>3</sup>

<sup>1</sup>*Instituto Nacional de Estatística, [joao.lopes@ine.pt](mailto:joao.lopes@ine.pt);*

<sup>2</sup>*Instituto Nacional de Estatística, [marco.moura@ine.pt](mailto:marco.moura@ine.pt);*

<sup>3</sup>*Instituto Nacional de Estatística, [sonia.quaresma@ine.pt](mailto:sonia.quaresma@ine.pt).*

---

**Abstract:** We analysed the impact of Labour Market Attractiveness on pressing policy questions, namely, the mismatch between jobs and skills at regional level, and labour market mobility. For this purpose, we used clustering techniques, multivariate regression analyses and weighted network correlation analyses. Our approach provides a proof-of-concept of a tool available for common citizens and enterprises seeking for labour market information, and for policy makers aiming to reduce the skills mismatch.

**Key-words:** Labour Market Attractiveness, Labour Market Mobility, Skills Demand, Skills Supply.

---

The European Big Data Hackathon took place in Brussels from 13 to 15 March 2017, in parallel with the New Techniques and Technologies for Statistics (NTTS) 2017. This event, organised by the European Commission (Eurostat), gathered 22 teams from 21 European countries to compete for the best data product combining official statistics and big data to support policy makers in one pressing policy question Europe is facing. The question, revealed the day prior to the start of the competition, was related to European Commission (EC)’s priority “Jobs Growth and Investment”, namely, “How would you support the design of policies for reducing mismatch between jobs and skills at regional level in the EU through the use of data”. The data product proposed was required to be supported by relevant data, statistical analysis and visualization. Teams were also invited to use some datasets provided (including European Employment Services (EURES) data from the EC on jobseekers and on job vacancies), and additional publicly available data sources with international applicability (e.g. Eurostat online database).

The development of our data product was focused on three main ideas: 1) combine official statistics data from Eurostat at NUTS2 level (i.e. solid data sets known for being well-structured, clean and accurate, but also characterized by a morose collect and release process) with real-time unstructured “big data”; 2) explore the notion of Labour Market Attractiveness as an important factor in the mismatch between skills demand and supply, as well as in the potential to reduce this skills mismatch, in labour market mobility, and in ordinary emigration; 3) create a flexible, interactive and user-friendly product that allows for customization of the answers in order to be used either by policy makers or by both citizens searching for help on jobseeking and enterprises looking for particular labour market characteristics.

The definition of Labour Market Attractiveness has to be considered carefully, thus, our approach should be seen as a first-step towards a more mature definition. We considered 17 variables from 6 Eurostat data sets, namely “reg\_demo” for demographics data, “earn” for earnings structure data, “edtr” for data on education and training, “ilc” for life conditions information, “employ” for employment/unemployment data, and “na10” for national accounts data. These variables were broken by several categorical levels (e.g. “age groups”, “level of education”, “qualifications”, “occupations”) originating more than 70 variables. Several data mining techniques were then considered to analyse the data. Using the Labour Market Attractiveness set, we calculated the distance between regions and visualize those using networks, we further clustered the regions using Partition Around Medoids (PAM) method on those distances creating a categorical variable with grouping information. This variable, along with variables on skills mismatch, labour market mobility, and emigration were used separately as dependent variables on multivariate linear and non-linear regression analyses using the Labour Market Attractiveness set as independent variables. We further constructed eigen variables from the considered set and performed Weighted Correlation Network Analysis (WCNA) on the dependent variables. Analyses were performed at country-level and at NUTS1- and NUTS2-level.

In our work we assumed two major simplifications in the construction of the skills mismatch indicator, however, these simplifications do not affect our product in terms of proof-of-concept and can be dropped in later developments. The first one was to use previously cleaned and treated data on job vacancies and education attainment from the Eurostat’s “labour” and “edtr” data sets, respectively. Instead, a better approach would be to use the freshly collected EURES data provided, but the use of this data would have two caveats: a) the cleaning of the data requires a considerable expertise on the subject; b) the normalizing of the data, using for example marginal calibration techniques, requires several demographic data at the required regional level in order to successfully capture the populations considered. The second simplification was to use an *ad hoc* mapping between qualifications (classified using ISCED-F 13) and the cross between occupations (defined using ISCO-08) and economic activity (defined using NACE Rev. 2). Nonetheless, a formal mapping will be released in mid-2017 by European Skills, Qualifications and Occupations (ESCO) from the EC.

We used exclusively Microsoft tools to store the data (SQL Server 2008 Express), to load and parse the data and to establish communications with a cloud service (SQL Server Management Studio 2016 and Azure Cloud), and to develop dashboards and reports available in Desktops and mobile devices (Power BI). All statistical analyses were performed in R using libraries “cluster”, “glmulti”, “Hmisc”, “MASS”, “nnet”, “sna” and “WGCNA”.