Data processing and analysis in Python language - Project

Individuals' level of digital skills

Author: Michele Guderzo

Warsaw, Academic Year 2021/22

Contents

1 Introduction

1.1 Background

Almost half of the jobs currently carried out in the world will, within a few years, be performed wholly or partly by machines. Millions of jobs will be lost, millions more will be created, but it is highly probable that the new ones will be more skilled, while those lost will be mostly low-skilled jobs. Moreover, there will be (and already are) professions that we do not yet know well or that we are not even able to conceive of. This is why it will be increasingly important to work on training, cultivating and constantly updating the so-called **digital skills**.

Digital skills were included among the eight key competences in the 2006 *Recommendation of the European Parliament and of the Council* and are considered essential for personal fulfilment and development, active citizenship, social inclusion and employment. The Recommendation defines digital competence as "the confident and critical use of Information Society Technology (IST) for work, leisure and communication". It is underpinned by basic skills in ICT (Information and Communication Technologies): the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet.

There are many ways in which digital skills can be categorised, each highlighting some specific characteristics over others. In 2017, AgID, the Agency for Digital Italy, defined the roadmap for monitoring and enhancing digital skills, dividing them into 3 levels: *basic digital skills* (users, including administrative staff), *specialist skills* (ICT), *e-leadership skills* (management).
Basic digital skills are the ability to use information technologies for work, leisure and communication with familiarity and critical thinking. They are useful skills for all citizens to be able to participate in the information and knowledge society and exercise their digital citizenship rights.
Specialised digital skills concern ICT professionals and future professionals and are in demand in both the private and public sectors. The development of appropriate ICT specialist skills is a crucial condition for digital growth, in the online public services sector, the evolution of manufacturing products, as well as for the efficiency and evolution of services.
Finally, e-leadership, or e-business, skills are the ability to make the best use of digital technologies within any type of organisation and to introduce digital innovation in the specific market sector in which one operates. Digital skills are closely integrated with the leader's transversal skills and with sector-specific competencies. They also include "digital skills for work", which must be possessed by all workers, as all jobs require the ability to use digital technologies.

Also, digital skills belong in general to two macro categories: *Digital Hard Skills* and *Digital Soft Skills*. Digital Hard Skills are the basic, specific, technical digital competences that define a professional figure. They can be acquired at school, at university, with masters and advanced courses, but also in the workplace, and they are quantifiable.
On the other hand, Digital Soft Skills are the transversal skills which concern people's relationships and behaviour in any work context, enabling them to use new digital tools effectively. Digital Soft Skills are not learned at school or at work and are difficult to quantify.

With an ever-increasing dominance, the importance of digital skills affects every sector. However, in Europe as in the rest of the world, there is still a gap between the supply and demand of digital skills, according to estimates by the European Commission: only 3.5% of university students attend an ICT degree course, and 1 in 3 workers lack basic digital skills. This shortage is reflected in business performance: 4 out of 10 companies reported a drop in productivity and customer retention due to a lack of digital skills. In fact, in the European Union countries, it was estimated that in 2016 around 37% of the workforce had insufficient digital skills to carry out their jobs; within this percentage, workers who have no digital skills at all, as they do not use the Internet, also fall. They make up 11% of the EU workforce, but this figure is over 25% for countries like Portugal, Bulgaria, Romania and Italy. Today, we are still a long way from achieving significant results in terms of increasing e-skills uptake, with only little improvements compared to several years ago.

1.2 Aim

This project, based on the level of people's digital skills, aims to study and analyse the situation of individuals in this area, paying particular attention to the year and the geographical area in which the data was collected. I am going to import, plot and analyze the data to look for some recurrence or trends among the data, in order to have a clear view of the situation and of the hypothetical development of the digital skills of individuals. Specifically, I will focus more on the condition of individuals around the world, and then shift the attention to countries closer to me, such as Italy and Poland. Eventually, I will draw the conclusions of my results, comparing them with each other.

2 Dataset

2.1 Data description

The dataset that was used for this project was found on the Eurostat.eu website, in the "digital economy and society" section of "science, technology, digital society".
The dataset consists of 860 observations and 7 variables which, specifically, are as follows:

- *TIME*: quantitative variable indicating the year in which the data were collected;
- *GEO*: qualitative variable that represents the geographical area;
- *INDIC_IS*: qualitative variable that indicates which skill sets individuals possess;
- *IND_TYPE*: qualitative variable which refers to the type of individuals;
- *UNIT*: qualitative variable that contains the string "Percentage of individuals";
- *Value*: quantitative variable indicating the percentage of individual;
- *Flag and Footnotes*: qualitative variable that represents footnotes (represented with letters).

In the next section I will import the dataset and start the processing procedure.

2.2 Data processing

As mentioned in the previous section, the original dataset contains 860 observations and 7 variables. However, once it was visualised, it was realised that it could be optimised and made leaner. First of all, I realised that the *IND_TYPE* variable merely returns the phrase "All individuals" for each observation. In my opinion, this decision does not bring any concrete knowledge and only makes the dataset heavier. It was therefore decided to remove this variable, obviously taking into account that this type of study was conducted on the totality of individuals, and not on a particular niche.
A similar consideration is made for the variable *UNIT*. This variable only carries the string "Percentage of individuals" for each observation. The real number concerning the percentage of individuals with specific digital skills is already represented by the variable *Value*. It is my belief that it is not useful to have a variable whose purpose is to "define" what is already defined by another variable, and so I proceed to remove *UNIT* from the dataset.
Moreover, the last variable *Flag and Footnotes* is confusing. It reports letters that are often associated, especially in the first few hundred observations, with '0' values in the *Value* variable. In addition, for the vast majority of observations the field of this variable is not filled in. As things stand, this variable is not useful and is therefore removed.

It was also noticed that the *Value* variable sometimes contains ":" instead of a number. Since this character cannot be interpreted, it is treated as NA and the corresponding rows are deleted from the dataset. Furthermore, rows relating of the *Euro area* variable have also been removed, as we already have information on several years of the European Union. I have therefore considered the Euro area observarions to be of secondary importance.
Lastly, rows corresponding to the *European Union* variable which are not part of the reference period (2013-2020) have also been deleted.

After the actual cleaning of the dataset, it is checked whether there are any missing values (and eliminated in case there are), hypothetical duplicates were removed and columns were renamed to make it easier to write the code. The new columns names are *year*, *geo*, *skills* and *value*.
As a final step, indexes were reset because they changed after dropping rows in *value*, *Euro area* and *European Union*.
After the whole data cleaning process, the final dataset consists of 691 observations and 4 variables.

2.3 Data analysis

Once the data have been processed, the first step in the analysis work is data visualization. The reason for that is because visualizing data often gives an intuitive understanding of the data and it helps to see patterns otherwise hard to see.

First of all, I decided to focus on individuals who possess at least basic digital skills, thus considering, at the same time, both people who possess basic skills and those who possess high skills. Of all the levels of digital skills possessed by individuals, I consider the one I chose to be the most interesting, since nowadays digital skills are increasingly in demand and are spreading also among the more manual jobs. Subsequently, in order to get an overview of the phenomenon and its development over the years, I decided to focus on the group of individuals who possess, globally, basic or higher digital skills from all over the world. A bar graph was then drawn in order to have a clearer and more immediate view of the data considered.

Figure 1 shows that, from 2015 to 2017, the average percentage of individuals possessing at least basic digital skills was on the rise, before falling in 2019. This goes against my initial idea that, generally, the progressive development of technology would be associated with increasingly higher digital skills.

Afterwards, I thought it would be interesting to propose a related study about the skills possessed by

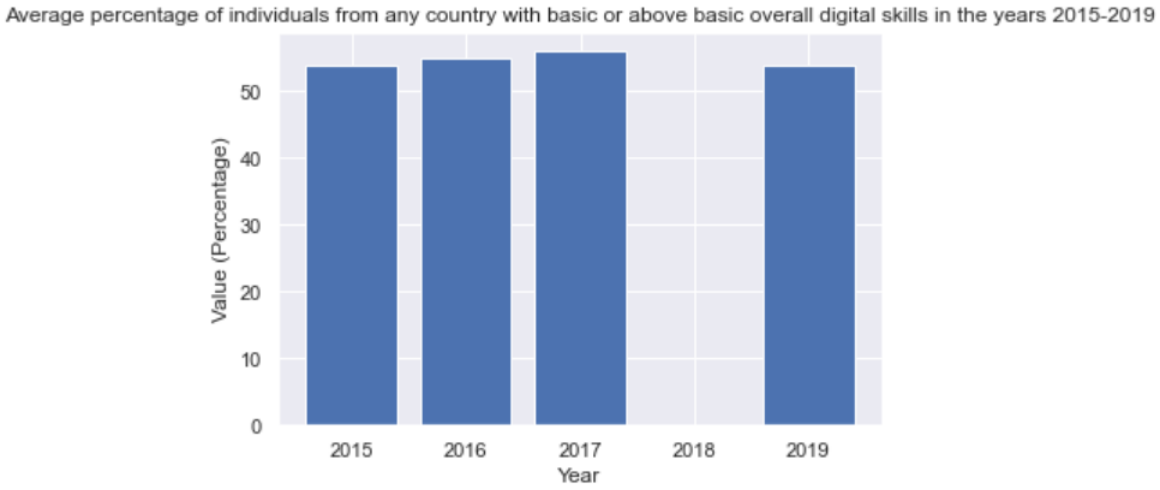

Figure 1: Average percentage of individuals from any country with basic or above basic overall digital skills in the years 2015-2019

people in my home country, Italy, and the same skills possessed by people in Poland, where my Erasmus experience is taking place. I therefore grouped the data for these two geographical areas and drew a single line plot, superimposing one graph on the other. In this way, the comparison and any difference between the two countries is immediately apparent.

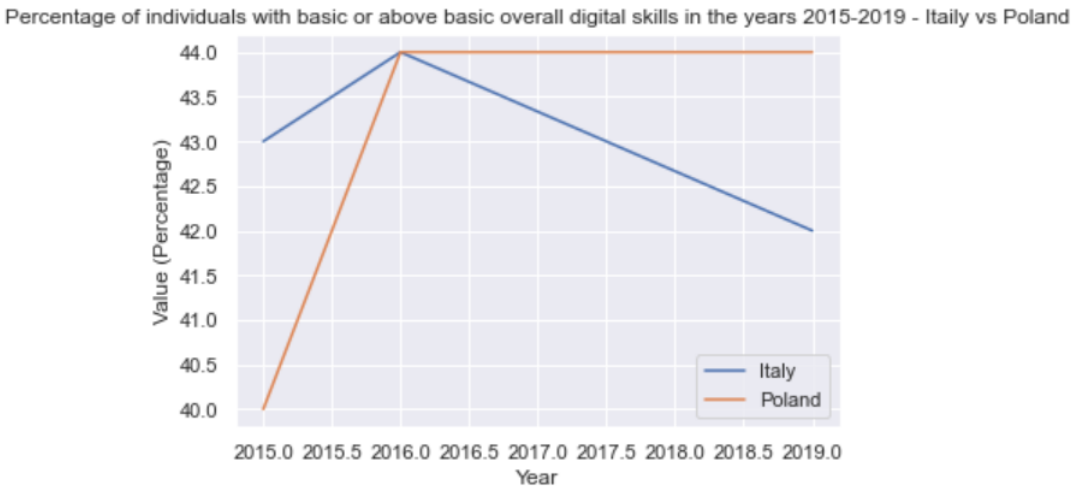

Figure 2: Percentage of individuals with basic or above basic overall digital skills in the years 2015-2019 - Italy vs Poland

This plot comparison shows something interesting. From the Figure 2 it can be seen that the two countries show different, almost opposite developments. Italy shows initially, in the year 2015, higher digital skills which increase in 2016 but decrease significantly in 2019. On the contrary, Poland presents a rather lower percentage in 2015 than Italy. However, this percentage increases significantly in 2016 (from 40% to 44%), and stays at this level also in the year 2019.

Finally, I compared the results of the global dataset and the datasets for Italy and Poland, also using the European percentage for a very similar period (2013-2020) as a reference.
It can be seen that the European percentage and the global average for individuals with at least basic digital skills are very similar (the European Union with 55%, the global average with 54.75%), also the European percentage is slightly higher. The situation is different for Italy and Poland. With a percentage of 43 and 42.67 respectively, the two countries do not reflect the global and European average, placing them significantly below.
What stands out here is the percentage of Poland. While it had, overall, higher digitally-skilled individuals than Italy in 2019, the percentage recorded in 2015 lowers its average by a large margin and, therefore, the latter is lower than Italy's, albeit by a small margin.

3 Conclusions

As mentioned in the first section, this project aimed to analyze trends in the digital skills possessed by individuals over a period from 2015 to 2019.
To summarize, the most important differences were seen when comparing the datasets from Poland and Italy with the results found for the global dataset and the European Union. The two countries showed rather lower percentages than their global and European counterparts. This result is in line with what was my initial idea, but the causes are not known. In addition, Italy and Poland are part of the European Union and therefore participate in its statistics. Probably, the reason why the percentage of the European Union is so high, relatively speaking, is to be found in other countries such as Scandinavia, but in general the northern European countries.
Thus, it would be interesting to look into this area with the aim of understanding which are the countries where individuals have the highest digital skills, not only in Europe, but worldwide.