Toolkit 1: FUNDAMENTALS OF DATA SCIENCE

What is data science?

According to the Cambridge Dictionary[1], data science refers to the employment of scientific methods to extract information from a large amount of data. It is commonly defined as data analysis to provide insights for businesses although it can be used for different purposes, such as improving policymaking. It combines statistics, mathematics, and technology to provide insights based on complex understanding data.

Data science step-by-step

The first step is data collection. Data comes in different formats. It might be primary, when it is directly collected by the researcher, or secondary, when exported from other sources. Primary data refers to answers to surveys conducted by the researcher, for example. The second step refers to cleaning the data. Data is used organised into databases, i.e., collections of structured data. Databases might be designed for different purposes than the research question. In this sense, data cleaning refers to excluding variables and missing information that are not useful for the project. Finally, data analysis implies to extract information from the data. This can be done either quantitively or qualitatively: employing statistics to estimate probabilities or investigating cases deeper. After analysing data, it is time to communicate it! After all, knowledge is only real when shared. To understand more each stage, access our toolkits on data collection, cleaning, and analysis.

Tools for Data Science

We have a great number of computational that might help data scientists. Python is one of the most used programming languages in this field due to its convenient libraries and readability. R is another programming language that might be used for data science. There are many other languages and software, such as SQL, Java, CC++ and Power BI. Overall, Python and R are employed for all the stages above, whereas SQL deals with relational databases and Power BI is a great tool for visualisation. Many platforms offer free and paid courses on these languages and tools.

References and further readings:

Cambridge Dictionary | English Dictionary, Translations & Thesaurus
George, N. *Practical Data Science with Python.* Birmingham: Packt Publishing, 2021.

[1] DATA SCIENCE | English meaning - Cambridge Dictionary.