

**Table of content**

[**General introduction 3**](#_gjdgxs)

[**Chapter 0: Project overview 4**](#_30j0zll)

[Introduction 4](#_1fob9te)

[1. Presentation of the project 4](#_3znysh7)

[1.1. Project context 4](#_2et92p0)

[1.2. Problematic 4](#_tyjcwt)

[1.3. Objective 4](#_3dy6vkm)

[2. Methodology adapted 5](#_1t3h5sf)

[2.1. IBM master plan 5](#_4d34og8)

[2.2. IBM master plan phases 5](#_2s8eyo1)

[Conclusion 7](#_3rdcrjn)

[**Chapter 1: Business Understanding 8**](#_26in1rg)

[Introduction 8](#_lnxbz9)

[1. Business objectives 8](#_35nkun2)

[1.1. Determination of business objectives 8](#_1ksv4uv)

[1.2. Data Science objectives 9](#_44sinio)

[1.3. Used Tools 9](#_2jxsxqh)

[Conclusion 9](#_z337ya)

[**Chapter 2: Analytic Approach 10**](#_3j2qqm3)

[Introduction 10](#_1y810tw)

[1. Analytic Approach 10](#_4i7ojhp)

[Conclusion 10](#)

[**Chapter 3: Data Requirement 11**](#_2xcytpi)

[Introduction 11](#_1ci93xb)

[1. Data Requirement 11](#_3whwml4)

[Conclusion 11](#_2bn6wsx)

[**Chapter 4: Data Collection 12**](#_qsh70q)

[Introduction 12](#)

[1. Data Collection Steps 12](#_3as4poj)

[1.1. Google Forms 12](#_1pxezwc)

[1.2. Web Scraping 13](#_49x2ik5)

[Conclusion 13](#)

[**Chapter 5: Data Understanding 14**](#_2p2csry)

[Introduction 14](#)

[1. Data Understanding 14](#_147n2zr)

[Conclusion 15](#)

[**Chapter 6: Data Preparation 16**](#_3o7alnk)

[Introduction 16](#)

[1. Removing empty experiences 16](#)

[2. Data translation 16](#)

[3. Extraction of graduation year from ESPRIT 17](#_23ckvvd)

[4. NLP Preprocessing 18](#_ihv636)

[5. Extracting ngrams 19](#_32hioqz)

[**Chapitre 7 : Data Modeling and Evaluation 20**](#_1hmsyys)

[Introduction 20](#_41mghml)

[1. Model: KeyBert 20](#_2grqrue)

[2. Model: FuzzyMatching 22](#_vx1227)

[2.1. Filtering the extracted KeyBert keywords 22](#_3fwokq0)

[Conclusion 23](#_1v1yuxt)

[**Chapter 8: Deployment and Feedbacks 24**](#_4f1mdlm)

[Introduction 24](#_2u6wntf)

[1. Common skills in both curriculum and extracted competencies 24](#_897x79u8rfpi)

[2. Missing skills in the curriculum 25](#_3tbugp1)

[Conclusion 26](#_28h4qwu)

[**General Conclusion 27**](#_nmf14n)

[**References 28**](#_1u7ay1s6fmk6)

**Table of figures**

Figure 1 : IBM master plan Phase5

Figure 2 : Libraries and tools used9

Figure 3 : Questions and answers of Google Forms12

Figure 4 : Example of the scraped results13

Figure 5 : Extracted 3-gram keywords19

Figure 6 : Extracted 2-gram keywords19

Figure 7 : KeyBert model figure20

Figure 8 : Extracted skills with Keybert21

Figure 9 : Fuzzy Matching figure22

Figure 10 : List of common skills and their percentages24

Figure 11 : List of missing skills and their percentages25

# General introduction

In today's rapidly changing job market, it is essential for students to possess the necessary skills and knowledge to succeed in their chosen careers. However, traditional education models may not always reflect the latest skills and competencies required in the job market. As a result, there is often a mismatch between what employers are looking for and what graduates can offer.

To address this issue, the Competency-based Curriculum Platform project was initiated.   
This innovative educational approach focuses on competency-based learning (CBL) which ensures that students are able to focus on the most important subjects or modules that are needed in real-life work and demonstrate their understanding through assessments and real-world applications of the competencies they have acquired.

The Competency-based Curriculum Platform project aims to automate the process of collecting data, extracting competencies, and updating IT school curriculums to ensure that they accurately reflect the latest skills and competencies required in the job market. This approach ensures that the curricula are up-to-date and relevant, and that graduates have the skills and knowledge they need to succeed in their careers.

By doing so, the platform will help schools to provide students with a more practical and relevant education experience and ensure that graduates have the necessary skills and knowledge to succeed in their chosen careers.

Overall, the Competency-based Curriculum Platform project is an innovative approach to modernizing the traditional educational model and providing students with a more practical and relevant education experience.

# Chapter 0: Project overview

## Introduction

This chapter gives a brief introduction to the project's objectives, scope, and context. It outlines the key issues and constraints of the project and the methods that will be used to achieve its desired outcomes. This chapter is important because it provides a clear understanding of the project and sets the foundation for the rest of the project documentation.

## Presentation of the project

In this part of the report, we provide an overview of the project context, problem and objective.

### Project context

The competency-based curriculum platform is designed for the education sector, specifically to address the gap between the skills taught in school and those required in the real-world job market. It aims to modernize the traditional educational model and provide students with a more practical and relevant education experience.

### Problematic

The current manual process of updating the curriculums for each specialization can be time-consuming, inefficient, and may not reflect the latest skills and competencies required in the job market. As a result, students who graduate from these programs may not be equipped with the necessary skills to succeed in their careers, leading to a skills mismatch between what employers are looking for and what graduates can offer.

### Objective

The overall objective of the competency-based curriculum platform project is to automate the process of collecting data, extracting competencies, and updating the IT school curriculums ensuring that they accurately reflect the real-world job market. By doing this, the competency-based curriculum platform will help the school to ensure that its IT curriculums are up-to-date and relevant, and that its graduates have the skills and knowledge they need to succeed in their careers.

## Methodology adapted

### IBM master plan

IBM's master plan is a comprehensive strategy that has guided the company's operations for decades. It emphasizes innovation, strategic partnerships, talent development, and corporate responsibility as key drivers of long-term success in the technology industry.

Moreover, IBM's master plan also serves as the foundation for a data science process model that can be used to guide the development and deployment of data-driven solutions. This process model emphasizes the importance of collaboration, experimentation, and continuous improvement in creating data-driven solutions that deliver real business value.

In summary, IBM's master plan offers a wealth of insights and best practices that can help organizations to thrive in the technology industry. By studying and adapting IBM's approach, businesses can create a clear and comprehensive roadmap for their own success, leveraging IBM's experience and expertise in areas such as innovation, strategic partnerships, talent development, and corporate responsibility.

### IBM master plan phases

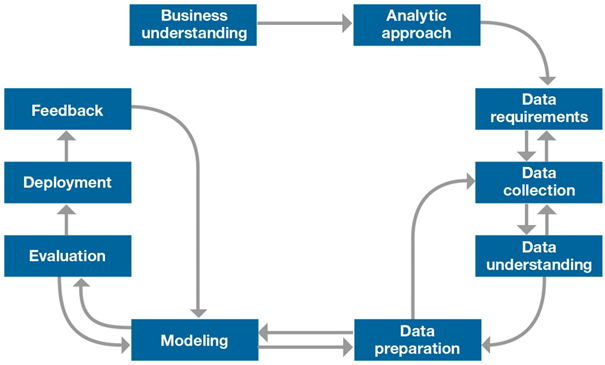


Figure 1 : IBM master plan Phases

#### Business understanding

In this phase, the data science team works with stakeholders to understand the business problem or opportunity that the project is intended to address. This involves defining the scope of the project, identifying key success metrics, and establishing a clear understanding of the target audience and their needs.

#### Analytic approach

The Analytic Approach phase focuses on how we can use the data in order to answer the client’s problems.

#### Data requirements

The Data Requirements phase focuses on understanding and acknowledging the data content, formats and sources for initial data collection.

#### Data collection

This phase involves identifying and acquiring the data needed for the project. This includes identifying relevant data sources, determining the best methods for collecting the data, and ensuring that the data is of sufficient quality and quantity to support the project.

#### Data understanding

In this phase, the data science team explores and evaluates the data to gain a better understanding of its characteristics and relationships. This includes conducting descriptive statistics, data visualization, and data profiling to identify patterns and trends in the data.

#### Data preparation

This phase involves cleaning, transforming, and formatting the data to ensure that it is ready for analysis. This includes handling missing data, correcting errors, and dealing with outliers and anomalies.

#### Data Modeling and Evaluation

In this phase, the data science team develops and tests models that can be used to address the business problem or opportunity. This includes selecting the appropriate modeling technique, training the model and evaluating the model's performance against the success metrics established in the Business Understanding phase.

#### Deployment and Feedbacks

The final phase of IBM's data science process model involves deploying the model into production and integrating it into the business processes of the organization. This includes developing an API or other interface for accessing the model, testing and validating it in a production environment, and developing a plan for ongoing maintenance and support of the model, while focusing on getting re-evaluation from the customer point of view.

## Conclusion

In conclusion, IBM's data science process model provides a comprehensive framework for organizations to undertake data-driven projects in a structured and effective manner. The model includes seven distinct phases, starting with Business Understanding, followed by Data Collection, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Each phase builds upon the previous one, and the overall goal is to develop and deploy predictive models that can help organizations address business problems and opportunities. By following this process model, organizations can increase their chances of success, minimize risk, and ensure that their data science efforts are properly aligned with their business objectives.

API :Application Programming Interface. In the context of APIs, the word Application refers to any software with a distinct function.

# Chapter 1: Business Understanding

## Introduction

In this section, we will gather requirements, define the scope of the project, and identify key stakeholders. We will establish clear project goals and objectives that are aligned with business needs and priorities.

## Business objectives

The Business Objectives section establishes clear, measurable goals for the project that align with the needs and priorities of the business. We'll work with stakeholders to define objectives that guide the project and measure its success.

### Determination of business objectives

Business goals describe goals in business terms, while data mining goals describe project goals in technical terms. Therefore, these are our main goals:

#### Identifying skills

We suggest using a form to gather the most common activities and categorize them according to skills or competencies.

#### Automate the process of updating curriculums

This process used to be manually done by a group of teachers. The automation of collecting data, extracting competencies, and updating IT school curriculums will be more efficient, gaining time and resources.

#### Matching the job market skills with current curriculums

This phase consists in comparing the skills required in the job market with current curriculums to identify outdated skills.

#### Updating the previous study plan

This process is the main goal of our project. By the final data, we will be able to make an improvement to the study plan.

## Data Science objectives

Data Science objectives describe goals in technical terms.  
For this project, our main data science goals are:

* Collect and preprocess LinkedIn data through scraping and cleaning
* Identify relevant skills and competencies using NLP techniques
* Matching skills in the job market with current university curriculums
* Generate insights and dashboards using visualisation tools
* Creating an easy-to-use User Interface using Django framework

## Used Tools



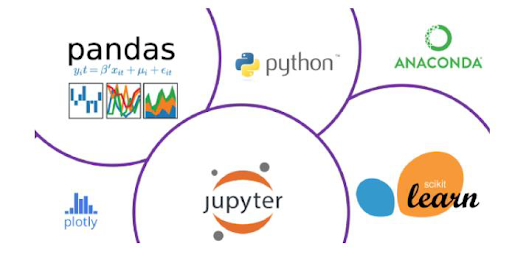




Figure 2 : Libraries and tools used

## 

## 

## Conclusion

In this chapter, we focused on our project business goals. The next chapter covers the third phase of IBM master plan, analytic approach.

**NLP**:Natural language processing **Django**: Tool for **high-level Python web framework**

# Chapter 2: Analytic Approach

## Introduction

In this chapter, we will be answering the question of how we can use the data in order to answer the client’s problems.

## Analytic Approach

The analytic approach is a crucial phase that helps to determine the appropriate algorithm(s) for the data science project. Depending on the type of question being asked, different models such as predictive, descriptive, or classification may be necessary. In the case of our project, various insights can be derived from the data by conducting descriptive analysis, such as treating each Curriculum separately for better results. Additionally, competencies can be classified to address the project's main objective of updating the study plan.

To ensure the success of the project, a series of steps and activities were carefully managed to avoid delays and development problems. By following a structured and organized approach, the project team was able to achieve their goals and deliver a practical and relevant education experience for students.

## Conclusion

In summary, the analytic approach is a critical phase in the data science process that helps determine the appropriate algorithms for a given project.Various insights can be derived from the data by conducting descriptive analysis to update the study plan. By following a structured approach, the project team was able to achieve their goals and deliver a practical and relevant education experience for students.

# Chapter 3: Data Requirement

## Introduction

In this chapter, we will identify the necessary data content,types, formats and sources for initial data collection.

## Data Requirement

The Competency-based Curriculum Platform project requires text data from two sources - Linkedin profiles and alumni forms. The linked profiles will provide information on past professional experience, while the alumni forms will provide additional insights into graduate careers. This data is crucial for extracting competencies and updating IT school curriculums to ensure they reflect the latest job market requirements.

By using this data, the project aims to provide students with a more practical and relevant education experience, and bridge the gap between employer requirements and graduate skills.

## Conclusion

The project needs text data from linked profiles and alumni forms to update the IT school curriculums and bridge the gap between employer requirements and graduate skills. This innovative approach has the potential to create a more competent workforce and meet the demands of a rapidly changing job market.

# Chapter 4: Data Collection

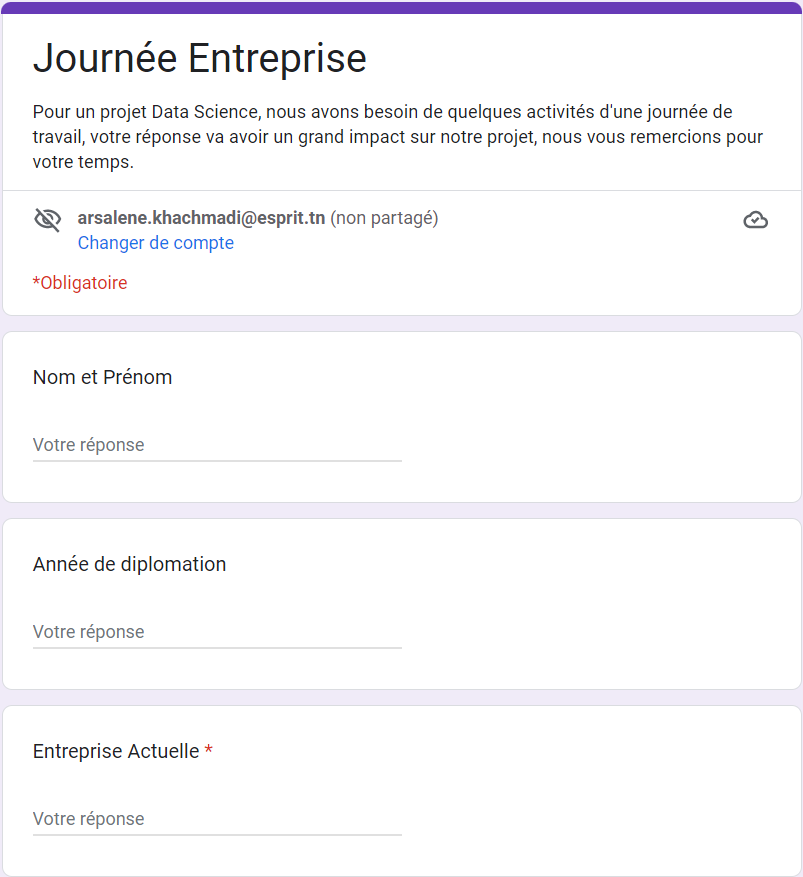
## Introduction

In this chapter, we will identify the data content, format and sources required for our data collection.

## Data Collection Steps

In this part, we have used various tools to collect our data, because it is necessary to collect data with quality, to do so we have used two data sources.

### Google Forms



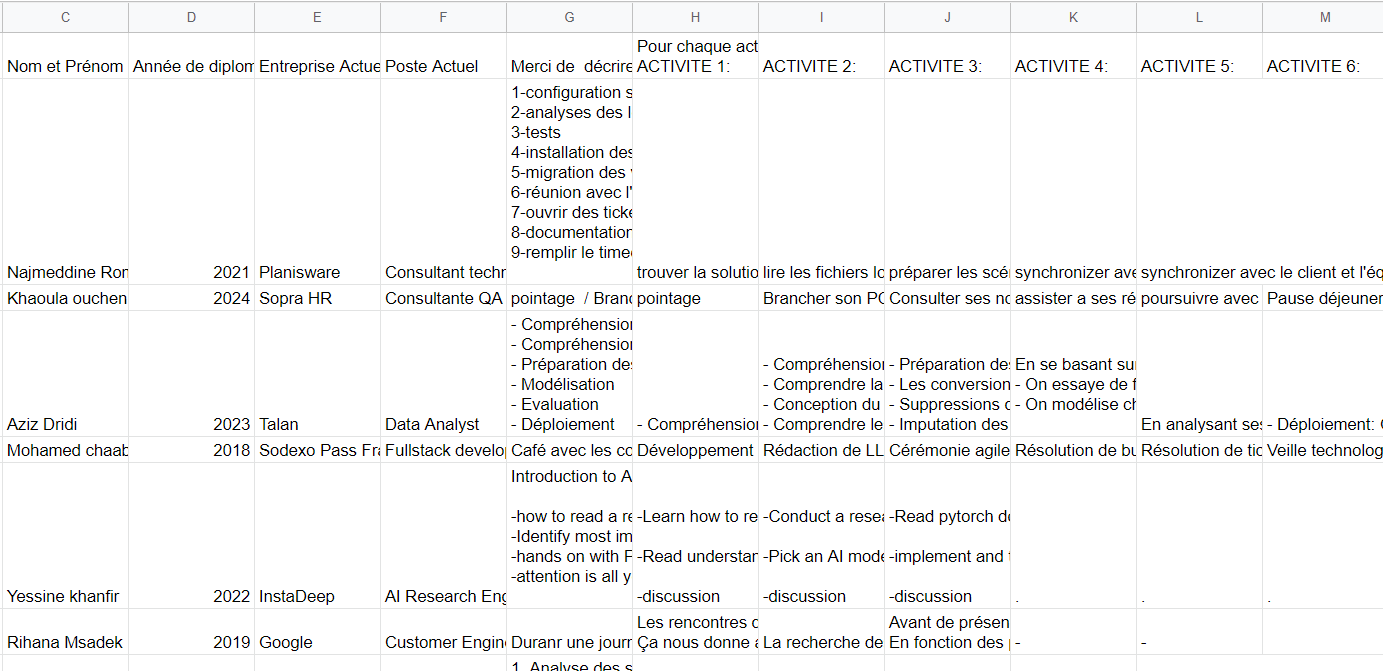
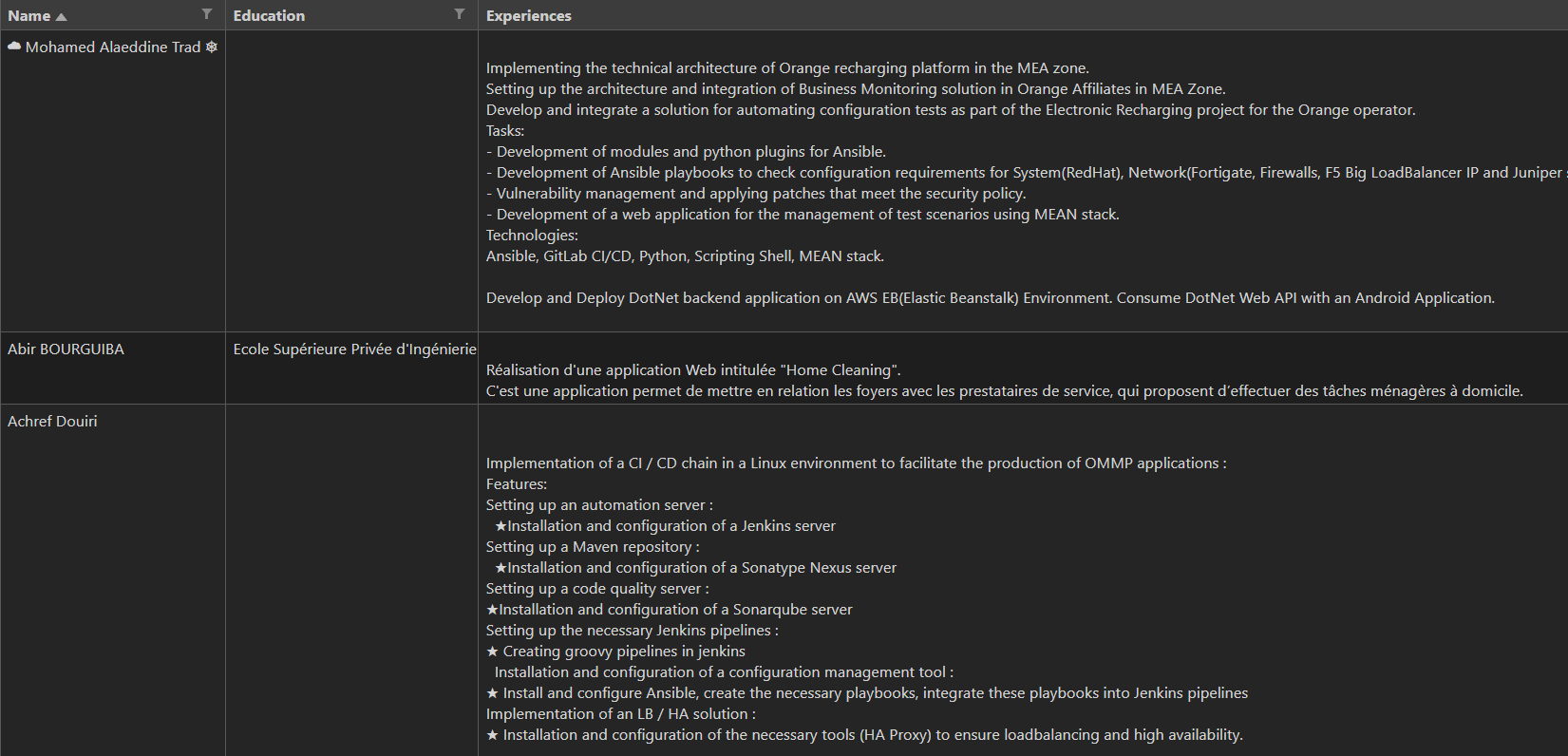


Figure 3 : Questions and answers of Google Forms

First of all, we have prepared a survey to collect the needed data from former ESPRIT students. We asked them to recite ten work activities.

### Web Scraping

To collect data from LinkedIn profiles, the Competency-based Curriculum Platform project utilized the Selenium and BeautifulSoup libraries to scrape the text data from experience descriptions. This automated approach allowed for efficient data collection and extraction of relevant information, such as skills and competencies acquired in previous roles, which were then used to update the IT school curriculums. By using this approach, the project was able to obtain large amounts of data in a relatively short period of time, facilitating the achievement of its goals.

Figure 4 : Example of the scraping results

## Conclusion

In this part, we have identified the different stages of the data collection phase.

**Selenium , BeautifulSoup** : python libraries used for creating an automated process (bot)

# Chapter 5: Data Understanding

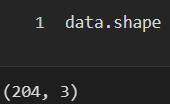
## Introduction

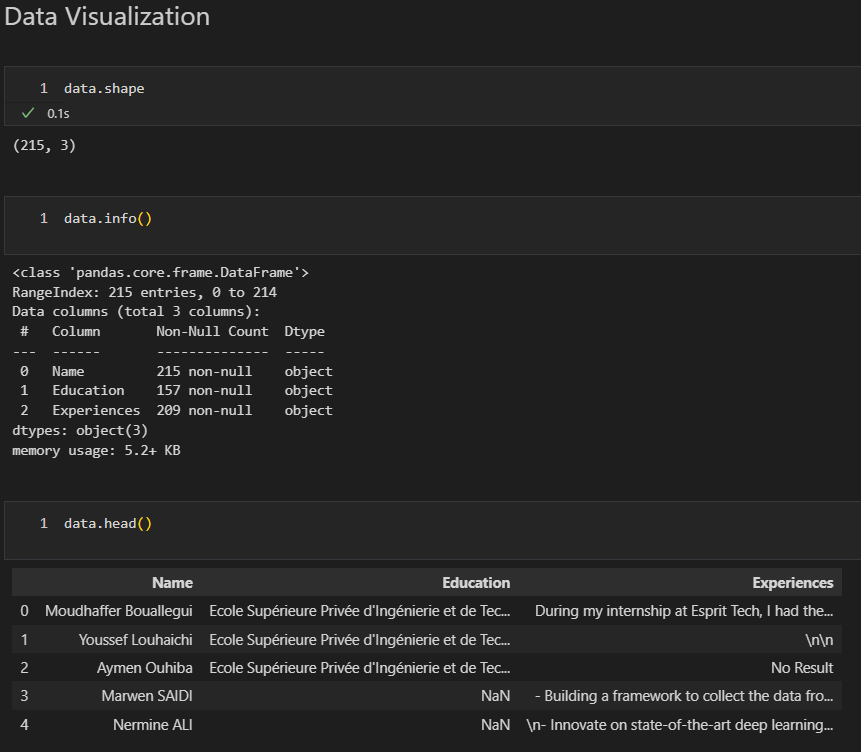
Data Understanding is a crucial step in the IBM Master Plan methodology. By loading and visualizing the data, we gain a comprehensive understanding of its components and prepare for the preprocessing phase. We'll explore the data to identify patterns, relationships, and anomalies while evaluating its quality for project requirements.

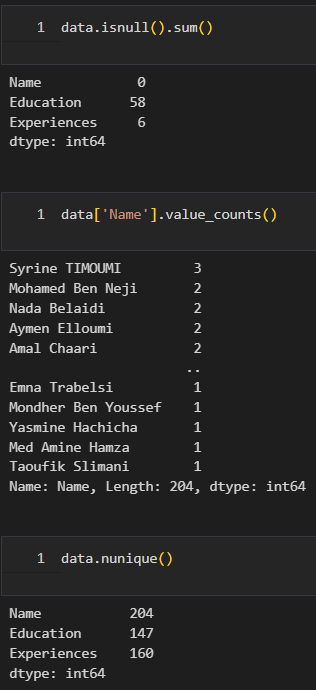
## Data Understanding

We have scraped our data from LinkedIn profiles using BeautifulSoup and Selenium.

Then we have used a dataset for each specialization, each one contains approximately 200 experiences:







## Conclusion

In this chapter, we overviewed the data understanding phase, visualizing missing data, and providing a general idea of the data issues. This data will be cleaned as it needs during the data preparation phase, which will be the subject of the next chapter.

# Chapter 6: Data Preparation

## Introduction

In this step, we will prepare the data using several techniques that we will be discussing in this chapter.

## Removing empty experiences

After collecting the data, we observed that many experiences are missing, so to ensure the quality of our data we have to remove these experiences.

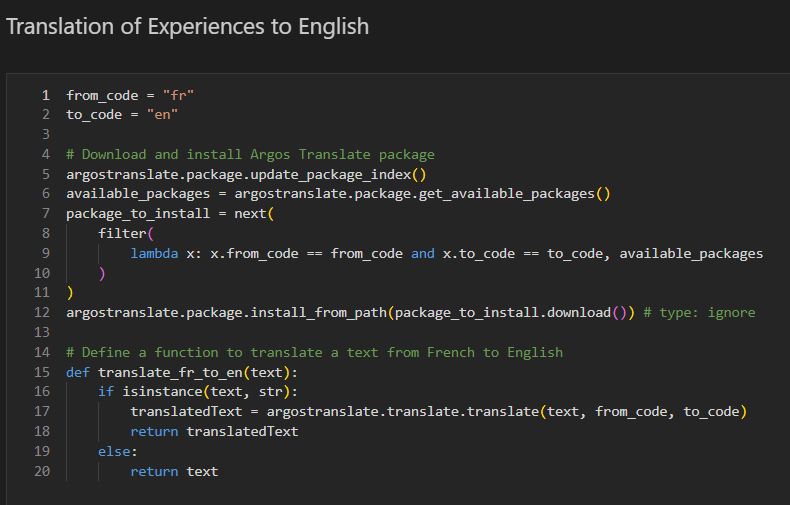


## Data translation

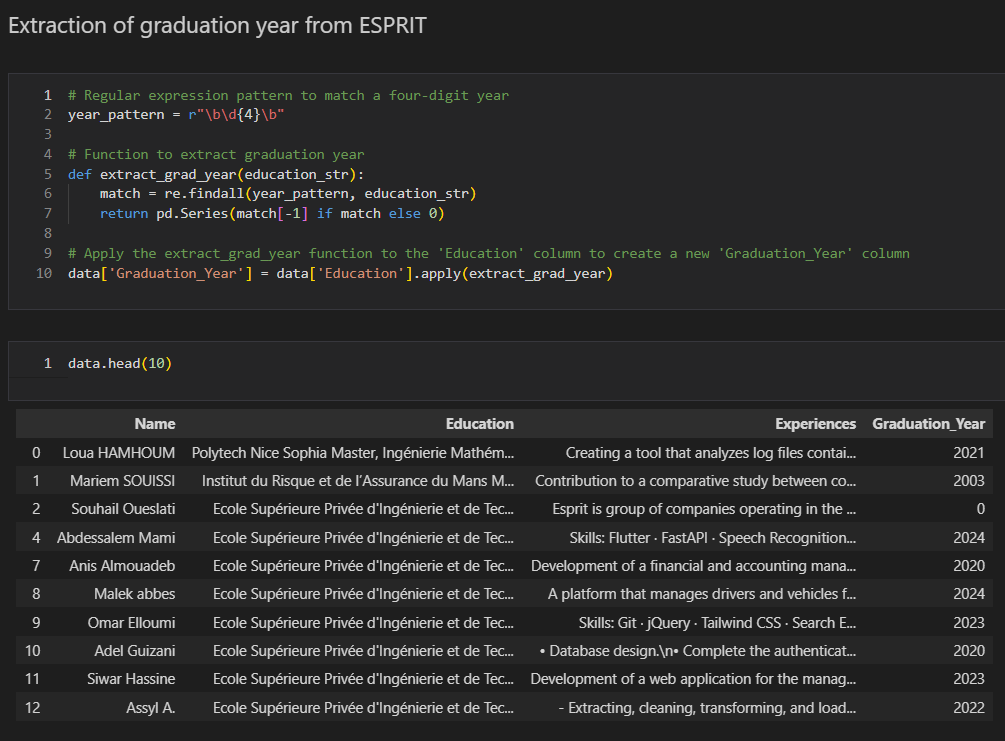
Another important step is to ensure all experiences are in a common language (English).

To accomplish this, we used Argos Translation package from Python:

**Argos :** a tool that translates text **Python:** a programming language

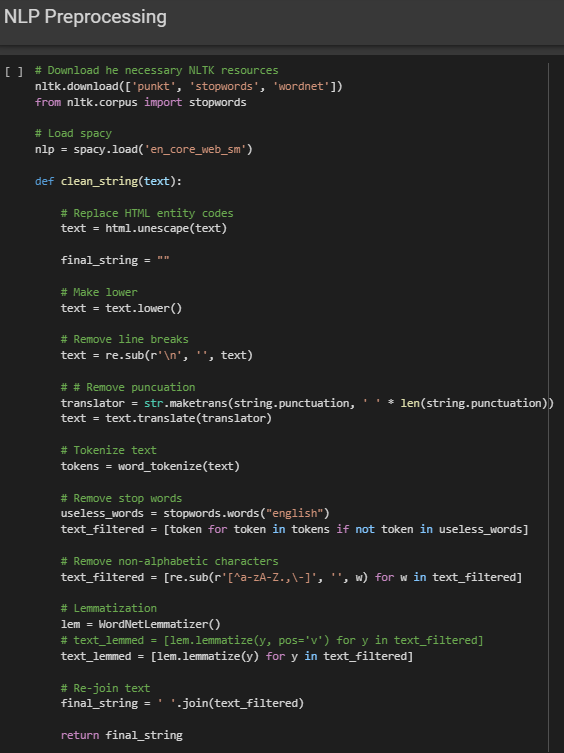


## Extraction of graduation year from ESPRIT

We might need the graduation year from ESPRIT if we want to filter the experiences to keep only previous alumni:

## NLP Preprocessing

NLP preprocessing involves cleaning and transforming raw text data into a format that can be analyzed by NLP algorithms. The preprocessing steps typically include tasks such as removing punctuation and special characters, converting all text to lowercase, tokenization (breaking text into individual words or phrases), removing stop words, stemming or lemmatizing words to their root forms, and encoding text into a numerical format suitable for machine learning models.

****

## Extracting ngrams

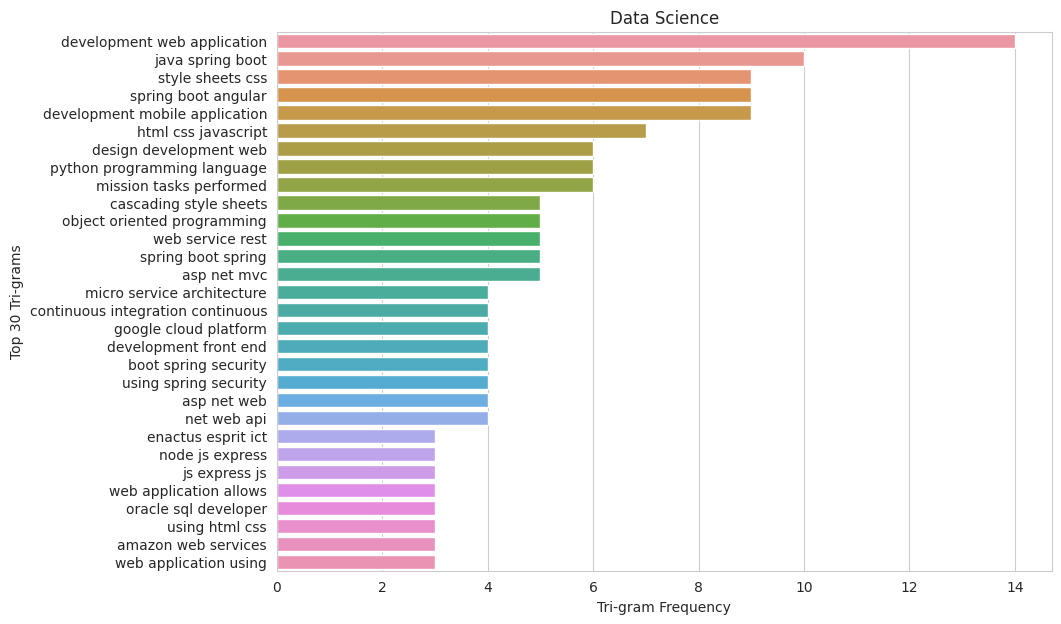
****

Figure 5 : Extracted 3-gram keywords

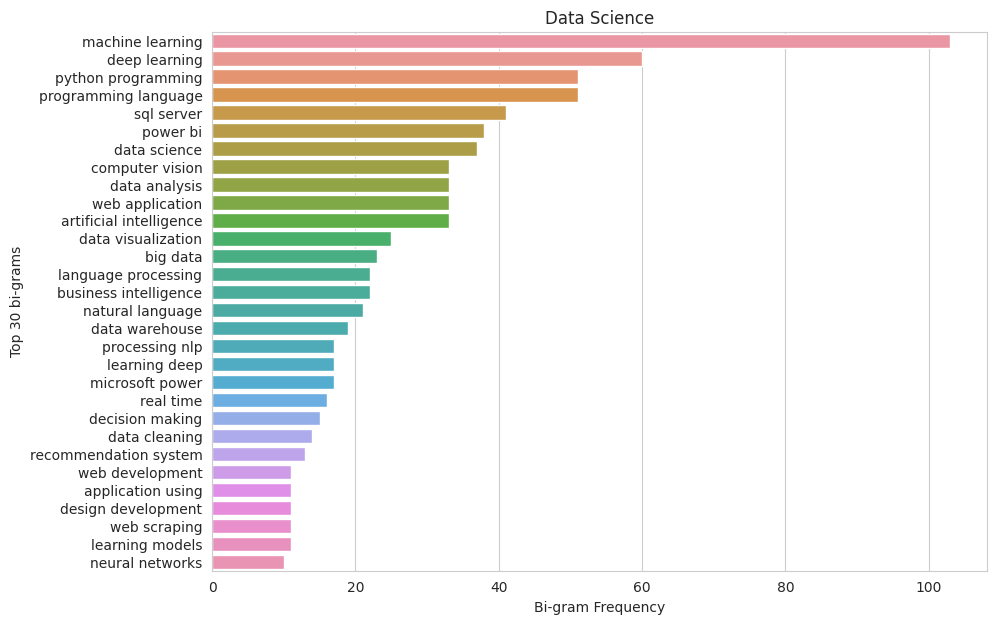
****

Figure 6 : Extracted 2-gram keywords

# Chapitre 7 : Data Modeling and Evaluation

## Introduction

After completing the data preparation phase, the next step was to build a model that could extract competencies from the textual data collected from LinkedIn and alumni forms. KeyBERT, a state-of-the-art unsupervised method for keyword extraction, was used for this purpose. The extracted competencies were then compared to a dictionary using fuzzy matching to improve accuracy. This approach helped to identify the relevant competencies and skills that were required, which were then used to update the study plan.

The evaluation process assessed the accuracy and reliability of the extracted competencies by comparing them with curriculum-based skills.

## Model: KeyBert

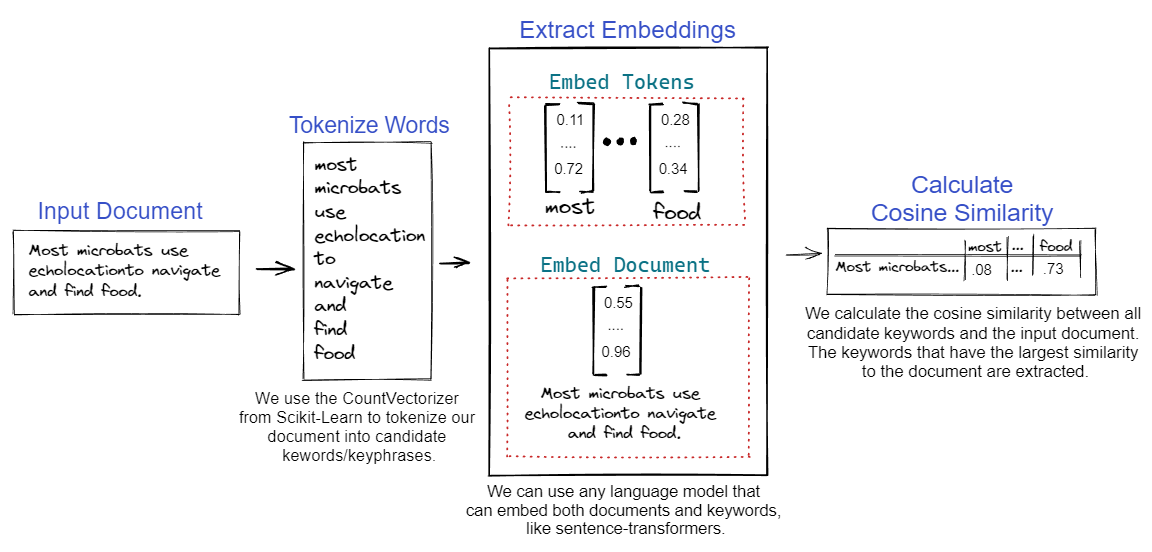
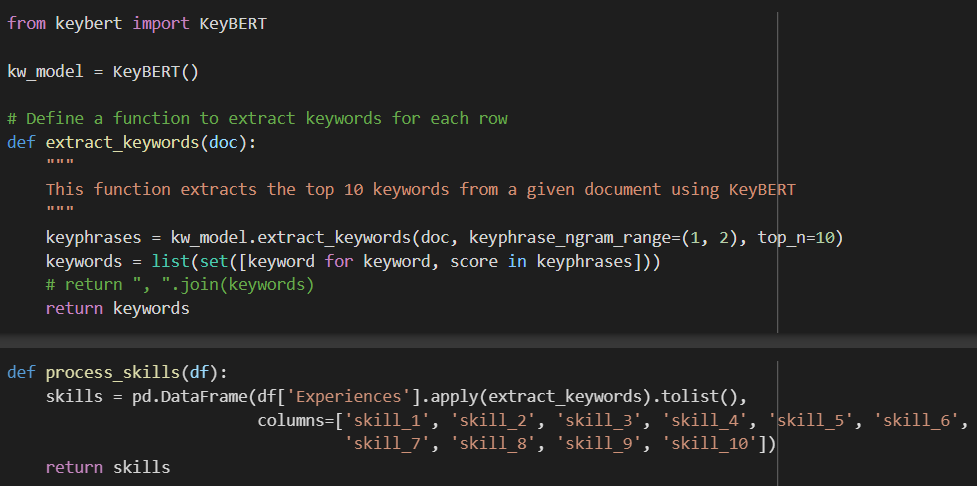
****

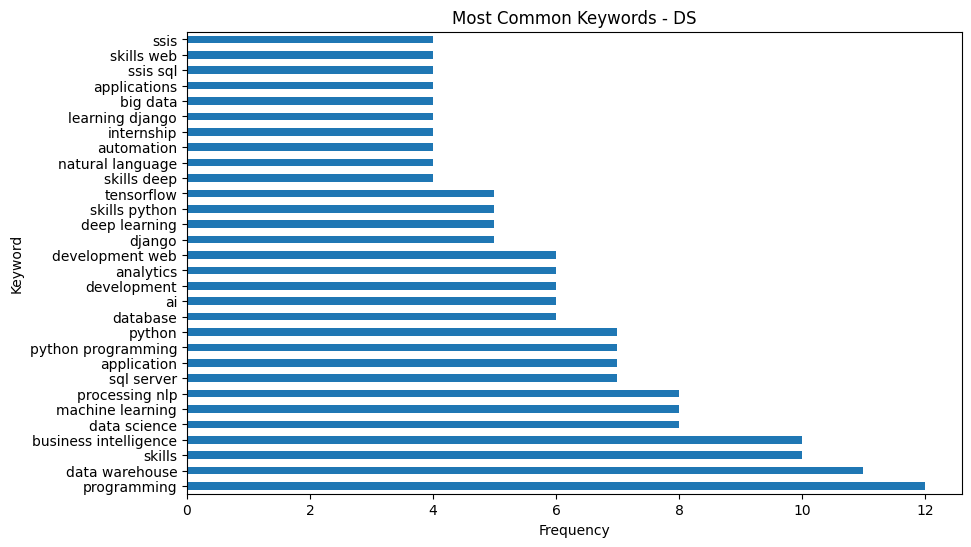
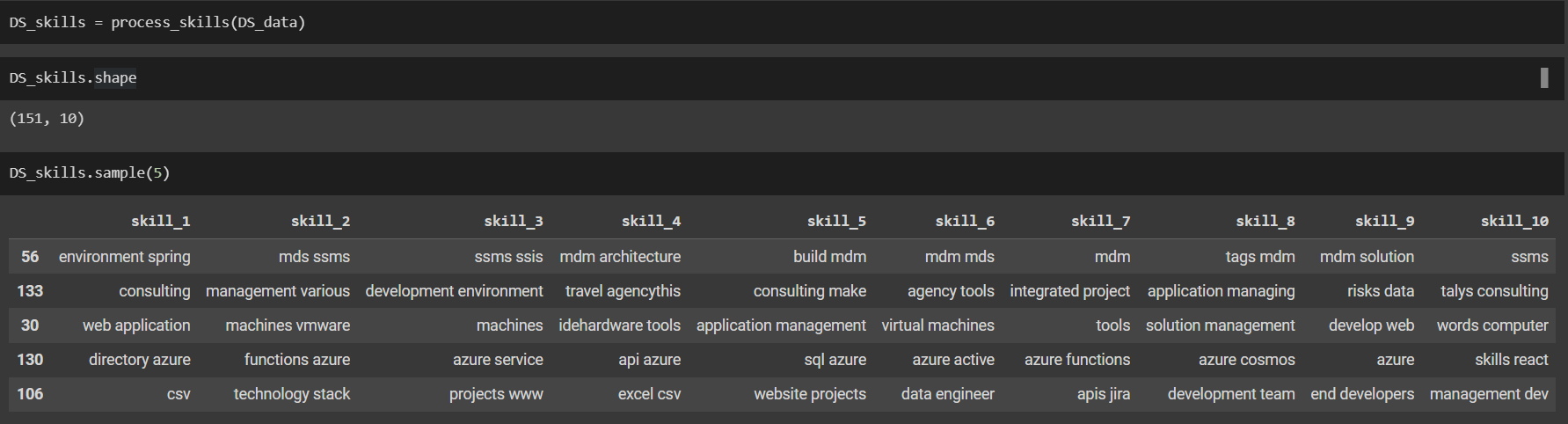
Figure 7 : KeyBert model figure

In order to ensure that the Competency-based Curriculum Platform project accurately reflects the competencies required in the job market, we utilized Keybert, a natural language processing library for keyword extraction. Keybert was used to extract relevant keywords from the text data collected from LinkedIn profiles' experiences descriptions.

**Keybert :** keyword extraction technique **Fuzzy Matching :** a technique that helps identify two elements that are approximately similar .



The format of the extracted keywords is as the following:

Figure 8 : Extracted skills with Keybert

We also visualized the most important keywords extracted by KeyBert:

## Model: FuzzyMatching

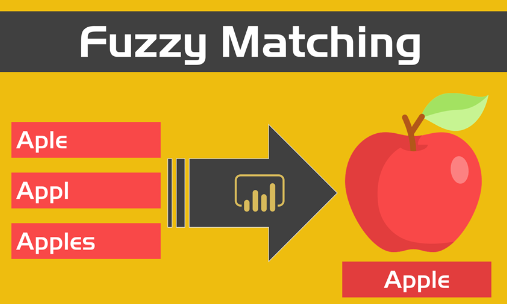
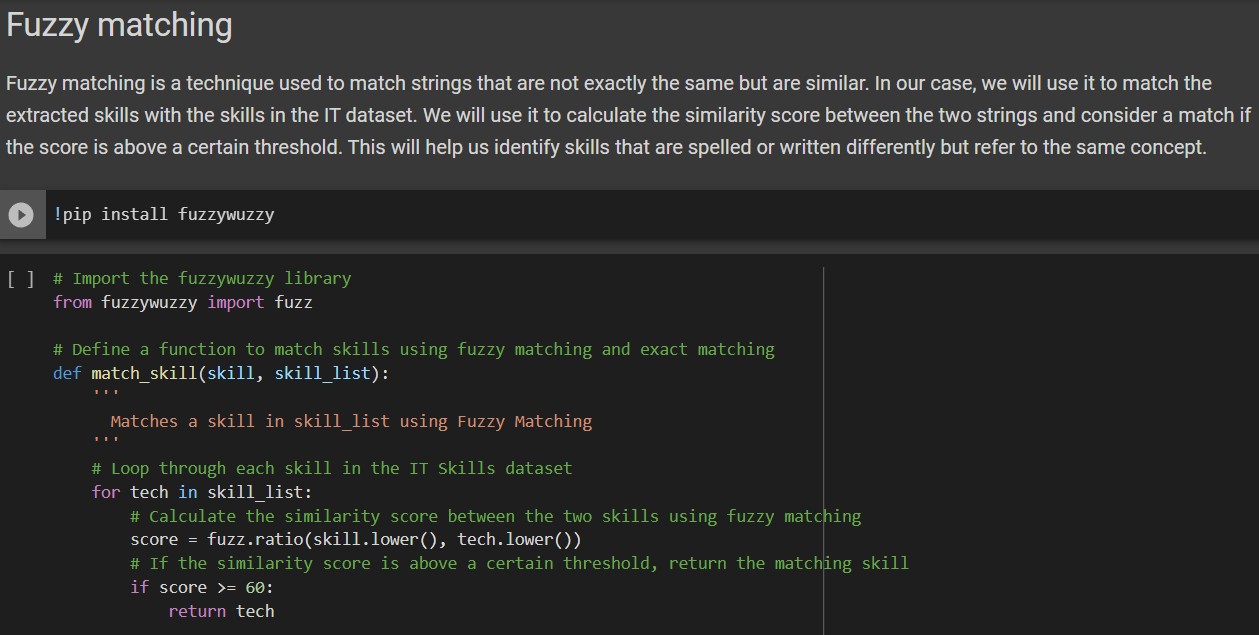
****

Figure 9 : Fuzzy Matching figure

### 2.1. Filtering the extracted KeyBert keywords

To further improve the accuracy of the extracted keywords, we also used fuzzy matching, which is a process of comparing strings and determining how similar they are to each other.   
 In this case, the extracted keywords were compared to a technologies dictionary using fuzzy matching to ensure that they really are relevant competencies.



Resulting dataset:



## Conclusion

Using automation tools such as Keybert and Fuzzy Matching can significantly reduce the time and effort required to extract and analyze competencies from LinkedIn profiles. These tools not only make the process more efficient but also improve the accuracy of the extracted data.

By using these techniques, educational institutions can ensure that their study plans are updated with the latest and most relevant competencies required in the job market.

# Chapter 8: Deployment and Feedbacks

## Introduction

The Deployment and Feedbacks phase involves deploying the model into production and integrating it into the business processes of the organization. This includes developing an API or other interface for accessing the model, testing and validating the model in a production environment, and developing a plan for ongoing maintenance and support of the model, while focusing on getting re-evaluation from the customer point of view.

### Common skills in both curriculum and extracted competencies

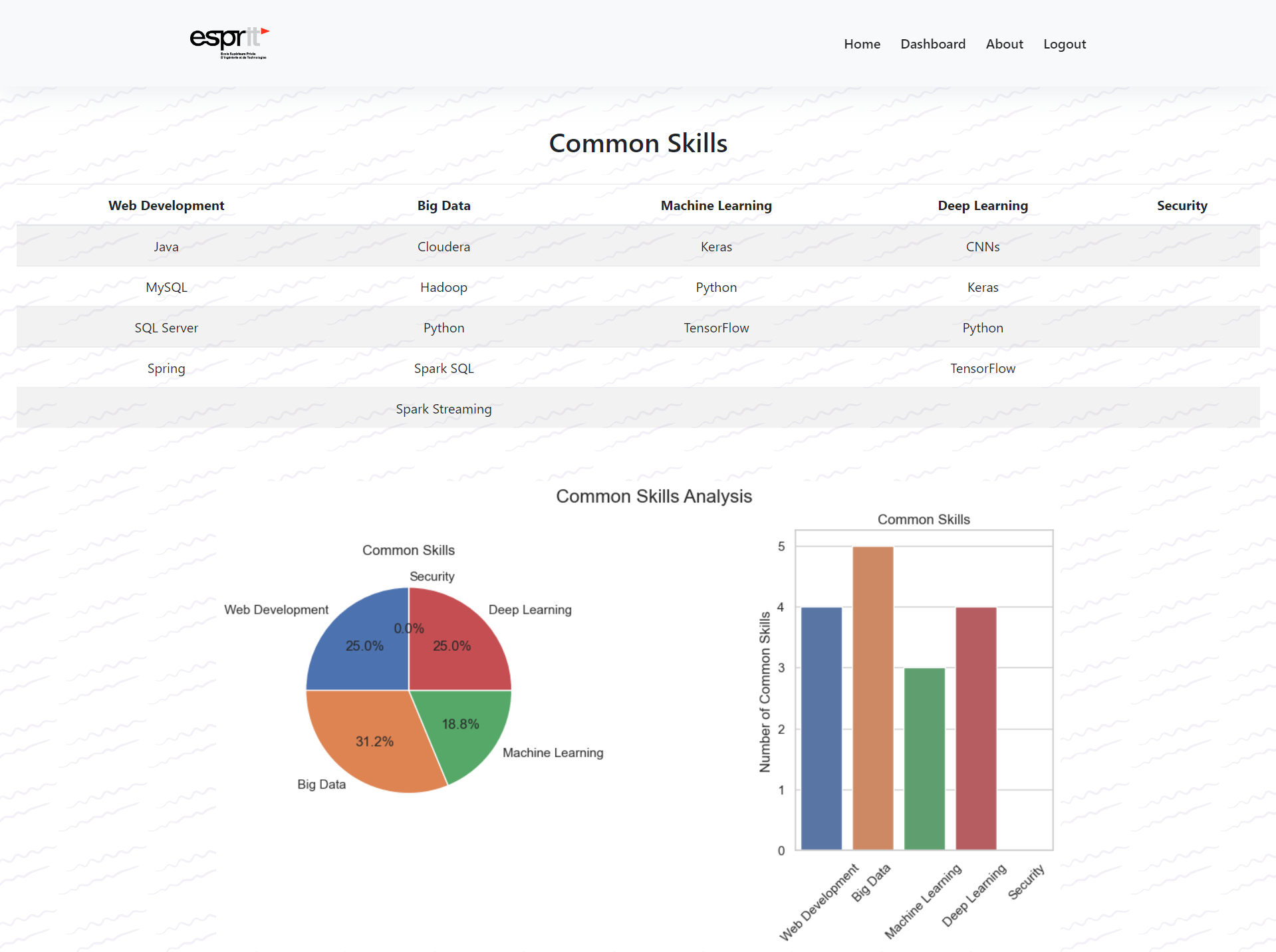


Figure 10 : List of common skills and their percentages

### Missing skills in the curriculum

Figure 11 : List of missing skills and their percentages

## Conclusion

The Deployment and Feedbacks phase in our project brings us closer to achieving our goal of providing students with a more practical and relevant education experience. By successfully deploying the platform and actively collecting feedback, we can refine and enhance its features to better bridge the gap between educational curriculums and real-world job market requirements.

Through an iterative approach, we are ensuring that the platform remains responsive to the evolving needs of the industry, ultimately equipping graduates with the skills they need to succeed in their careers.

# General Conclusion

# 

In conclusion, the Competency-based Curriculum Platform project is an innovative and promising solution to address the mismatch between traditional education and the constantly evolving job market. By focusing on competency-based learning and automating the process of updating school curriculums, this project aims to equip students with the most up-to-date skills and competencies that are required by the job market.

The project's approach ensures that graduates have the skills and knowledge they need to succeed in their chosen careers, which will not only benefit the students but also the employers who are looking for skilled workers. With the Competency-based Curriculum Platform, schools can provide students with a more practical and relevant education experience that better prepares them for the challenges of the job market.

Overall, the Competency-based Curriculum Platform project has the potential to revolutionize the education sector by bringing it closer to the needs of the job market and creating a more skilled and competent workforce for the future.

# References

[https://maartengr.github.io/KeyBERT](https://maartengr.github.io/KeyBERT/)

<https://github.com/lukebarousse/Job_Analysis>

[https://argos-translate.readthedocs.io](https://argos-translate.readthedocs.io/)

<https://pypi.org/project/fuzzywuzzy>

<https://pub.towardsai.net/understanding-semantic-analysis-using-python-nlp-f48016422677>

<https://medium.com/@dcraven02/how-easy-it-is-to-scrape-linkedin-profiles-lets-find-out-ce82685d0a91>

[https://www.nltk.org](https://www.nltk.org/)

[https://docs.djangoproject.com](https://docs.djangoproject.com/)

[https://www.selenium.dev/documentation](https://www.selenium.dev/documentation/)

[https://www.crummy.com/software/BeautifulSoup/bs4/doc](https://www.crummy.com/software/BeautifulSoup/bs4/doc/)

[https://seaborn.pydata.org](https://seaborn.pydata.org/)