

## **UniMi Laboratory Project with OSN Data**

- Online Lab (Zoom)
- Event for: team (3)
- Event open to everyone, there is no limit on the number of participants
- Stage / Thesis / Internship opportunities provided by Mediolanum

### **TITLE**

Skills of the future analysis

### **DESCRIPTION**

This analysis focuses on identifying the necessary skills for STEM (Science, Technology, Engineering, and Mathematics) profiles to succeed in a recruitment process. It involves the use of a database from a headhunting company that specializes in finding STEM talent for Italian companies in Italy or abroad. The database contains candidate information, job information, company information, and recruitment progress. All personal information has been anonymized.

Open Search Network is the first headhunting company specialized in Italian STEM profiles globally. Since its establishment in 2013, it has successfully placed over 500 candidates and collected more than 24k CVs. Based on this experience and data, the company aims to identify the right set of skills for STEM profiles to succeed in a recruitment process and determine the career path to become a CDO (Chief Data Officer) of the future.

If you're up for the challenge, you will have the opportunity to analyze a unique and niche database of a headhunting company that focuses on STEM opportunities in the Italian market. The database can provide various types of analysis.

There are several types of data science analyses that could be performed using the data in the database. Some potential examples include:

- **Descriptive Analysis:** This type of analysis would provide a summary of the data in the database, such as the number of candidates, open positions, and Italian companies seeking STEM talent. Such an analysis would help the headhunting company better understand overall trends and patterns in the data.
- **Predictive Analysis:** This type of analysis would aim to predict future outcomes based on the data in the database. For example, the headhunting company could use predictive analysis to forecast the number of candidates they are likely to receive for a particular role, or the probability of a candidate accepting a job offer.
- **Prescriptive Analysis:** This type of analysis would focus on identifying the best course of action based on the data in the database. For example, the headhunting company could use prescriptive analysis to determine the optimal recruitment strategy for filling a particular role, or identifying the best-fit candidate for a given position.
- **Text Mining and Sentiment Analysis:** This type of analysis would involve extracting insights from unstructured data, such as candidate resumes or job descriptions. Text mining and sentiment analysis could help the headhunting company identify trends in the language used by candidates or companies, which could inform recruitment strategies and provide insights into the job market.

Overall, the most appropriate type of data science analysis would depend on the specific goals and objectives of the headhunting company. The analyses mentioned above are just a few examples of the types of insights that could be derived from the data in this type of database.

Other analysis:

- **Skills analysis:** You could analyze the skills listed on the CVs of the 600 candidates with STEM skills to identify the most in-demand and commonly used skills in the STEM job market. This information can be useful for creating job profiles and identifying required skills for open positions.
- **Frequent skills analysis among non-inserted candidates:** You could examine the CVs of the 28,000 non-inserted candidates to identify the most frequently occurring skills. This information can be useful for identifying job market trends and the skills required for open positions.
- **STEM candidate availability analysis:** You could analyze the availability of candidates with STEM skills compared to the total number of candidates in the database. This information can be useful for identifying the availability of STEM candidates in the job market and for evaluating competition for candidates with STEM skills.
- **Required skills analysis for open positions:** You could analyze the skills required for open positions and compare them with the skills of the 600 candidates with STEM skills to identify matches and discrepancies. This information can be useful for identifying the skills required for open positions and identifying candidates with the necessary skills.
- **Origin analysis:** This analysis could identify the universities, companies, and geographic regions where candidates come from. This will allow you to evaluate the geographic origin and educational background of candidates, as well as to identify any clusters of specific skills or industries that may be of interest for the open positions.
- **Job transition analysis:** This analysis could identify patterns of job movement and career changes among the candidates in the database. This way, you can identify the most common positions or industries where candidates worked before applying for current positions, as well as the average duration of such positions.
- **Regression analysis:** This type of analysis could be used to model the relationship between different variables in the CV data, such as education level, work experience, and job performance. For example, you could build a regression model to estimate how job performance is affected by a candidate's education level and work experience.
- **Cluster analysis:** This type of analysis could be used to identify different groups of candidates based on their characteristics, such as education, work experience, and skills. This could be useful for identifying patterns or trends in the data, as well as for segmenting the candidate pool for different job positions.
- **Time series analysis:** This type of analysis could be used to examine how the characteristics of candidates change over time, such as how the education level or work experience of candidates has evolved over the past decade. This could be useful for identifying long-term trends in the labor market for STEM jobs.
- **Network analysis:** This type of analysis could be used to examine the relationships between different variables in the CV data, such as the connections between different universities and industries. This could be useful for understanding the broader context in which STEM jobs are located and for identifying potential opportunities for collaboration or partnership between different organizations.
- **Identify the most in-demand skills and experiences:** Using data analysis, you could identify the skills and experiences that are most in demand for a particular job or industry. This could help

you improve the candidate screening process and identify candidates with the most suitable profile.

- **Identify candidates with the best profile:** Using data analysis, you could identify candidates with the most suitable profile for a particular job or industry. This could help you improve the candidate selection process and identify the most promising candidates.
- **Predict employment duration:** Using data analysis, you could predict the length of employment for a candidate in a particular job or industry. This could help you identify candidates who are more likely to stay long-term in a company or identify any turnover issues.
- **Diversity and inclusion analysis:** Using data analysis, you could identify any inequalities or inclusion issues in the candidate selection process. This could help you improve diversity and inclusion within the company.

**DURATION:** 20 total hours

Total hours for operational work = 16

Hours for data presentation and results = 4 (2 + 2)

### TIME SCHEDULE

- 27/03/2023, 2 hours, 10:30-12:30, Lab Presentation, Dataset & Task
- 03/04/2023, 2 hours, 10:30-12:30, Lab
  - Task: Data Cleaning, Data Preprocessing, Data Reading, Data Merging
  - Output: Database with a reduced number of tables and/or joined based on a primary key
- 05/04/2023, 2 hours, 14:30-16:30, Lab
  - Task: Data Analysis, Data Extraction, Skill Extraction, Job Profile Extraction
  - Output: Vocabulary of all the different skills present in the CVs, Vocabulary of all the professional profiles present in job postings.
- 11/04/2023, 2 hours, 14:30-16:30, Lab
  - Task: Data Analysis, Data Extraction, Skill Extraction, Job Profile Extraction
- 12/04/2023, 2 hours, 14:30-16:30, Lab
  - Task: Data Analysis, Data Extraction, Skill Extraction, Job Profile Extraction
- 17/04/2023, 3 hours, 14:30-17:30, Lab
  - Task: Data Exploration [\*]
- 19/04/2023, 2 hours, 14:30-16:30, Lab
  - Task: Data Exploration
- 4 hours in Asynchronous Teaching Mode
- 03/05/2023, 2 hours, 14:30-16:30, Lab
  - 2 hours, Result Presentation

During the Data Exploration sessions starting from 17/04/23, you will have the opportunity to use your intuition and experience to decide which type of analysis would give you the best insights.

All analyses must be completed with a presentation of the results using a dashboard or a data visualization tool of your choice (we suggest using Einblick platform [www.einblick.ai](http://www.einblick.ai) for computational analysis & visualization).

To participate in this laboratory, you need to create a team of three students. For the Result Presentation on 10/05/2023, all members of your team need to participate during the pitch.

This laboratory will be conducted online on Zoom.

Online learning requires a higher degree of individual autonomy and responsibility than face-to-face learning. You will be expected to follow the instructions provided online and work independently, while we will be available to offer support as needed.

To participate in the online laboratory session, you will need access to a computer or mobile device with a stable internet connection. You must have your webcam connected to allow us to verify your presence throughout the laboratory.

Please make sure to test your equipment and familiarize yourself with the online platform before the session starts.

I look forward to working with you all online!

**ELIGIBLE STUDENTS :**

Open to everyone

Preferred Skills:

- Basic knowledge of R/Python
- Basic knowledge of SQL and Relational DB
- Interest in innovation and HR Analytics