Business Analytics

Group 12 Project Report

Enrico Cipolla Cipolla 3139480
Fabio Giovanni Stefana 3121555
Yang Xiaohan 3126780
Jakub Nahorny 3255614

Table of Contents

Introduction

In an era defined by rapid technological advancements and evolving business landscapes, the process of transitioning from academia to the professional realm poses significant challenges for university students and recent graduates. As they set their sights on the job market, many individuals find themselves grappling with a lack of clarity regarding the specific skills and experiences demanded by prospective employers. Recognizing this pervasive issue, our team has embarked on an initiative to harness the power of data analytics to demystify this transition, enabling job seekers to refine their profiles and align with industry standards effortlessly.

Our project is dedicated to simplifying the journey for individuals on the cusp of graduation. We understand the common struggle of sending out numerous job applications without certainty about hitting the mark.

Through our algorithm, applicants can clarify their target positions, submit their credentials, and receive a thorough analysis of their compatibility with potential roles. This process aims to dispel the ambiguities of job hunting, paving a more focused and strategic path to employment.

Of course, even though our product is focused on students and people early in their careers, but it’s not limited to such demographics, it could be useful to anybody, and down the line, it could end up as a tool for recruiters.

Product Description

Our product is an AI-powered software that helps to elevate careers, it is designed to turn the professional aspirations of students and young professionals into achievable goals.

In order to better explain the workings of our product we will illustrate the process that a new user goes through when first approaching our product, this process is divided into several steps:

1. Personalized Career Mapping: A customer starts their journey by uploading his CV and/or sharing key personal details, such as skills, education, work experience, and certifications. Our platform is built to understand and analyze a customer’s unique professional profile.
2. Skill Assessment & Enhancement: Through a series of targeted tests, we assess various skills, providing a comprehensive evaluation that's both insightful and actionable. This would be done through GPT-generated quizzes that assess correctly the level of the candidate.
3. Targeted Job Matching: The customer inputs the jobs that he is interested in applying to - be it Data Scientist, Data Engineer, Machine Learning Engineer, etc. Our intelligent system, fueled by the latest data from top job sites like LinkedIn, Glassdoor, and Indeed, as well as direct feeds from employers' career pages, will guide the customer toward the roles they are best suited for.
4. AI-Driven Insights: Receive a detailed analysis from our AI model, including:
    a. Skill Improvement Guide: Specific recommendations for enhancing those skills (an example of a recommendation would be “improve proficiency in Python, better understand how the Keras library works, etc.).

b. Skill Gap Identification: Highlighting missing skills, that the customer needs to acquire in order to have a significant improvement in their alignment with the requirements of the positions he is interested in (an example of a recommendation would be “Learn the Julia programming language, 83% of our customers who have learned Julia have gotten a job as an ML Engineer”).
c. Education Level Assessment: An evaluation of the customer’s education level and its alignment with job requirements.
d. Educational Growth Plan: Suggestions on how a customer may elevate their educational qualifications.
e. Experience Appraisal: An analysis of the customer’s professional experience and its relevance to their desired role.
f. Competitive Edge Analysis: Our AI model also provides the customer with a data-driven estimation of their likelihood of securing the job (based on our data on previous users), and it also offers a comparative analysis, showing how the customer's profile stacks up against other candidates interested in the same position.

Comparison with LinkedIn

This platform stands apart from LinkedIn in some essential concepts. First of all, It conducts active skill assessments through targeted tests, providing a more diagnostic evaluation of abilities compared to LinkedIn's self-reported skill endorsements. LinkedIn offers the possibility to take quizzes to assess competencies, but this is done to signal specific skills to employers. Instead, with our product, quizzes enable to understand the real level of the candidate against the competition through AI-generated quizzes, which is something that LinkedIn does not offer at the moment. Moreover, it provides a competitive edge analysis using historical user data to estimate job acquisition likelihood.
In general, the product positions itself as a comprehensive career advancement tool, offering a mix of personalized career advice, skill assessment, and strategic job matching, functioning more like a career coach than a professional networking site.

Business Model

After thorough analysis, we decided that the subscription model would be our most important and simultaneously beneficial revenue stream. Different tiers of subscription plans will provide access to different levels of features, insights, and information. It would be possible to purchase it on our online platform.

The next step could be partnerships and collaborations with such institutions as schools or universities to offer our platform as an indispensable service to the students looking for a job or adequate career path. At the same time, partnerships with MOOC providers, like Coursera or EdX, are crucial. In this way, once people are aware of what they are lacking on, they can enroll in high-quality online courses to tackle this problem in their profiles.

That exactly shows our target customer groups: students and early graduates, seeking guidance on career paths and skill development as well as educational institutions and corporates.

Two key resources for our business are AI Technology and Data. The first of them requires investing in AI development and algorithms to provide accurate skill assessments and career recommendations. The second one demands access to updated and diverse job market data from platforms like LinkedIn. In addition, every candidate that uses our platform improves the effectiveness of our algorithm, as it counts as an extra data point to assess accurately the competition for a given job position. All the data points would be anonymously collected. Given how important this element is, a powerful go-to-market strategy would be to offer, for the first few months from the birth of this project, our product for free, in order to expand our dataset.

Cost structure

The biggest costs will be generated by the two mentioned key resources: investment in ongoing AI and platform development costs associated with accessing and analysing data from various job sites and employers. In particular, the latter one is particularly urgent.

Consecutive expenses are related to marketing efforts and forming partnerships to expand the user base.

Future Directions

Not only do we have to extend the platform's reach but also enhance the AI algorithms.
The first path should be concerned with getting to as many customers as possible by incorporating successive job markets from various countries and specialized versions of the platform. It can be also achieved by simplifying and making it more available by creating a mobile app and teaching modules that could help in using our platform most efficiently.
The second path should be pursued by persistently enhancing AI algorithms and continually updating machine learning models based on user interactions and the changing job market. Integrating advanced Natural Language Processing capabilities to allow users to interact conversationally with the platform might be game-changing on the market and our outstanding feature.
The final destination would be to develop an integrated ecosystem to connect candidates and companies, where candidates can have instant access to tailored assessment on their profile and what and how they can improve it, and employees can browse the job market landscape to find the ideal candidate for a specific job position.

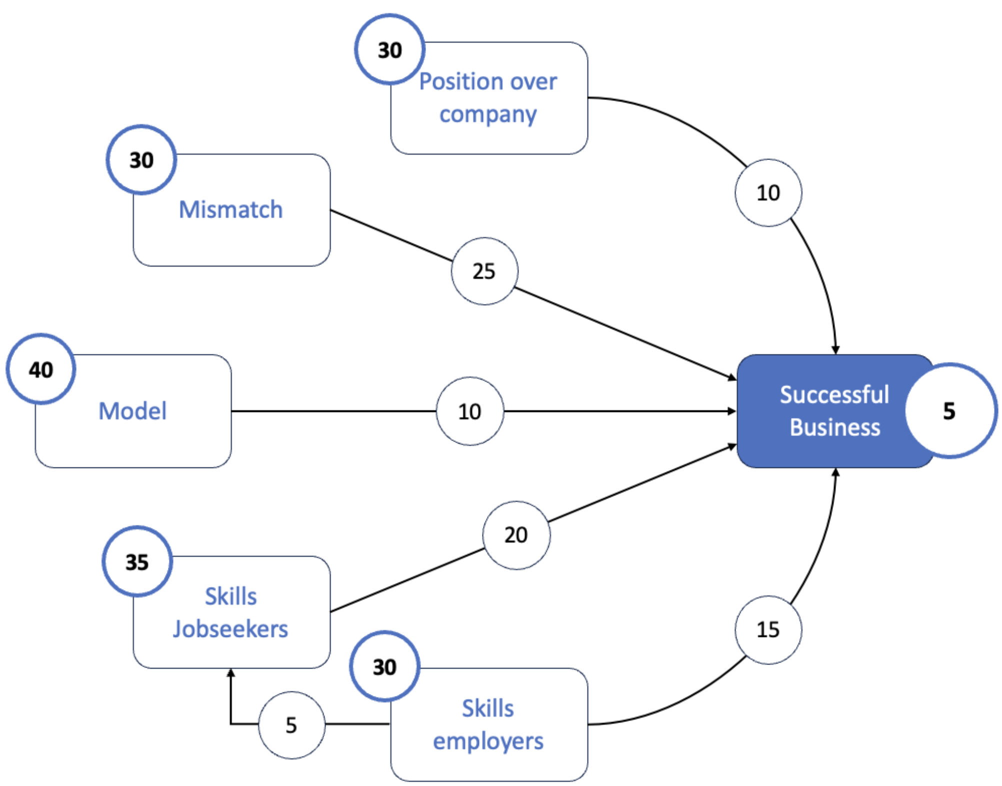
Decision-making framework:

As university students ourselves, we cannot just throw ourselves into this entrepreneurial journey, we need to evaluate this business decision very well. The decision we face is not trivial, it is a “low frequency, high impact decision” with far-reaching implications. Thus, we are applying the framework of Camusso et al. to meticulously evaluate this business venture.

We begin by identifying the key attributes that will shape our product's success, these include the relative importance of job positions over companies, the potential mismatch between perceived and actual job requirements, our ability to construct a robust model, the comparative value of hard versus soft skills, and the volume of job seekers in the market. With these factors in mind, we have crafted a causal map to visualize their impact on our business's prospects.

1. Position over company: When looking for a job do people tend to look for a specific job position and they are not really concerned with what company they might end up with, or do they tend to look for any position within their realm of interest but for a specific company? We believe that job positions tend to be more important. Let this attribute denote the fact that for most jobseekers position is more important than company.
2. Mismatch: Is there a mismatch between what people think are the requirements for certain job positions and what the actual requirements are? Let this attribute denote the fact that there is a significant enough mismatch of knowledge.
3. Model: Can we build the model? Let this attribute denote the fact that we can successfully build a stable and working model.
4. Skills matter for jobseekers: Do jobseekers consider skills as an important factor that may help them distinguish themselves from others in the recruiting process? Let this attribute denote the fact that jobseekers believe skills to be an important factor.
5. Skills matter for employers: Do employers consider skills to be an important factor when comparing candidates? Let this attribute denote the fact that employers do rely on skills when comparing candidates.

With this understanding, we invite you to delve into the nuances of our causal map:



In order to properly analyze our causal map, we need to isolate each link and explain how we theorize its functioning. We will start from the top and explain each link one by one.

1. Position over Company:
   We believe that if most people think that positions are more important than the company offering the position this will have a positive effect on the success of our business. We believe this to be true for several reasons: firstly, because our product focuses on job positions rather than companies, therefore, a bigger interest in job positions would translate into more demand for our product, secondly, an emphasis on position might reflect a more dynamic job market where individuals are more adaptable and likely to switch roles, once again increasing the demand for our product.
2. Mismatch:
   We believe that the existence of a mismatch in knowledge is an essential element for our product to be successful if it were not true there would be little to no demand for our product.
3. Model:
   Our ability to produce a strong model is another essential element, maybe even the most important one. The ability to build a model is crucial not just for initial success, but for scalability. A robust model can adapt to changing market trends, maintaining relevance.
4. Skills matter for jobseekers :
   We believe that if jobseekers prioritize skills in their career progression, they are more likely to seek tools that help them understand and improve these skills, also our AI-driven insights into skill gaps and improvement areas will be more valued by users who understand the importance of skills.
5. Skills matter for employers:
   We believe that if employers prioritize skills in their hiring decisions, our product becomes a critical tool for job seekers to align their profiles with these expectations. This alignment can significantly boost the success rates of users applying for jobs, thus proving the effectiveness of our platform. On top of that the fact that skills matter for employers is also going to affect the perception of skills for jobseekers, as they would like to fulfill employers' expectations.

Moreover, we would like to test whether to focus on soft or hard skill in our system, according to which one of the two is more effective in landing a job, thus in the workplace.
In the causal map we came up with the probabilities based on our experience, also considering the fact that we have a low-risk tolerance we decided to stay on the safe side.
Regarding the probabilities of each attribute, given the fact that we have no evidence, yet we've chosen to err on the side of caution, but nevertheless since we believe in our capabilities to deploy an effective algorithm, we decided to put a higher probability (40%) on this attribute.
Looking at the causal links we expect that Mismatch and Skills Jobseekers will have the biggest impact on our probability of success, this is because these directly influence the demand for our product. Following closely, these two we find that Skills Employers has a pretty significant impact on our probability of success, but not as much as the two mentioned above. Lastly Position over Company and Model have the least impact on our probability of success.

With our causal structure and probabilities, the expected value of our theory assuming its true is 31%.
We are around 50% sure that our theory is true, and we believe the expected value of our project if the theory was not confirmed to be around 20%. So we can conclude that our overall expected probability of success is around 26%.

$$V = \omega V_{\Theta} + (1-\omega) V_{\hat{\Theta}}$$
$$\omega = 0.31$$
$$V_{\Theta} = 0.5$$
$$V_{\hat{\Theta}} = 0.2$$
$$V = 0.26$$

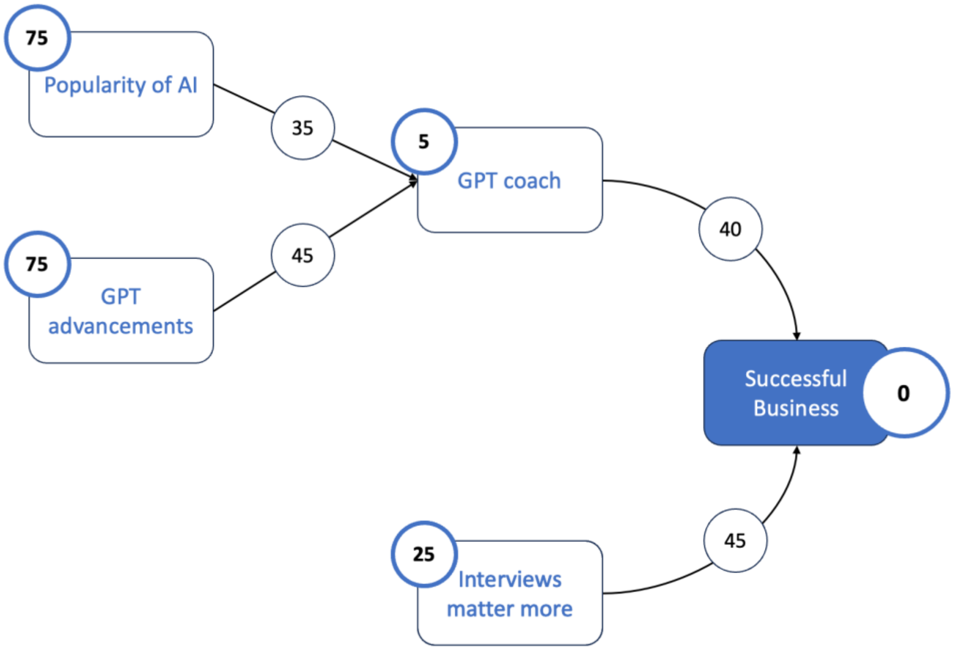
Alternative Theory

Along with that, we decided to also consider an alternative theory, a GPT-powered coach for interviews. The concept of a ChatGPT-powered coach for interview training is grounded in the recognition that securing a job may depend more on interview performance than on actual skill levels. This coach would leverage ChatGPT's AI capabilities to simulate various interview scenarios, providing users with a range of questions typically asked in job interviews. The AI would not only assess the content of the answers but also offer feedback on aspects like clarity of communication, confidence, and appropriateness of responses.

This approach is particularly valuable as it allows for personalized practice sessions where users can refine their interview techniques in a stress-free environment. The AI can adapt to different industries and roles, offering targeted practice that reflects the specific requirements of various job positions. Furthermore, such a system could provide tips on body language, answer structuring, and how to effectively convey one's experiences and skills, all of which are crucial for making a strong impression in an actual interview setting.

The likelihood of a ChatGPT-powered interview coach being successful is moderately high, around 40%, considering the popularity of similar AI-driven tools in various sectors. For instance, in language learning, AI-powered applications have been widely adopted due to their interactive and personalized learning experiences. ChatGPT's ability to simulate natural conversations has proven invaluable in helping users practice and improve their language skills. Additionally, in sectors like customer service, ChatGPT has been instrumental in providing automated but human-like support, enhancing user experience and efficiency.

This is the causal map we came up with, now we will take a good look at all the different attributes:



1. Popularity of AI: this attribute refers to the fact that we think that in the future AI is going to be a widespread tool, given the current state of things we are pretty sure this will be the case, so we decided to put a high probability of 75%
2. GPT advancements: this attribute refers to the fact that the GPT ecosystem of tools will further develop allowing us to create an even better model, after seeing the recent advancements with chat-GPT 4, and the recent OpenAI developer conference we are confident that there will be many more advancements. That being said we decided to put a high probability of 75%.
3. GPT coach: this attribute refers to the fact that we can build an effective GPT coach, this is influenced greatly by both Popularity of AI and GPT advancements. Furthermore, this attribute has a base probability of 5%, we are not very sure we could build the model without the preceding attributes being true.
4. Interviews matter more: this attribute refers to the fact that in the selection process recruiters don't actually evaluate skills, but rather the ability of the jobseeker to be a good interviewee.

With this theory, the probability of success assuming our theory its true is 37%. We are less sure about this theory than our main one, so we considered a 35% probability that our theory is true. In this case though if our theory were to be wrong, we are expecting a bit higher probability of success, for that reason we considered 25%. This brings the overall probability of success of this theory to 29%.

$$V = \omega V_{\Theta} + (1-\omega)V_{\hat{\Theta}}$$
$$\omega = 0.37$$
$$V_{\Theta} = 0.35$$
$$V_{\hat{\Theta}} = 0.25$$
$$V = 0.29$$

Considering that our initial theory has a lower overall probability of success, that we think that it may have a greater upside potential, and that it perfectly lends itself to be tested with a survey; we decided to test our initial hypothesis.

Our test starts with analyzing current literature on what are 2 critical attributes of our theory, Skill Mismatch & importance of skills for employers, that are particularly difficult to test. Then, we develop a survey to collect other relevant data that would be useful for our research, before taking our final decision.

Preliminary analysis

While some causes might be difficult to assess through already existing theories - possibly because they did not provide an incentive to develop one; others find a relatively solid grounding in already established literature.
More specifically, we can start by evaluating whether there actually is skill mismatch between the labor supply and labor demand, and how significant is its impact. The expression skill mismatch is generally used with a broad meaning, and can be distinguished in vertical mismatch, horizontal mismatch, skill gap, skill shortage and skill obsolescence (McGuinness et al., 2018). Since horizontal mismatch refers to the differences dictated by the specialization in different fields of study, we won't be expanding on the subject since it won't be relevant for the purpose of our project.

While measurements based on education level can be obtained in an easier manner, skill based metrics often provide a more precise evaluation of the discrepancy (OECD, 2013, and Desjardins & Rubenson, 2011).

Skill gaps and skill shortages are usually analyzed in an aggregated manner, from the labor demand perspective. Skill gaps occur when employers believe that workers do not have the ability to correctly fulfill their role - oftentimes the issue is worked around by allocating more employees to the task, or extending working hours. On the other hand, a skill shortage occurs when there are vacant positions due to the shortage of competent applicants.

Although most of the literature on the subject of skill mismatch tend to focus on the over-education/ over-skilling side, researches such as Verhaest & Omey (2012) tend to prove that human capital deficit also have relevant effects on the market.

Though most skill mismatch cases can be correlated with the purpose of our project, skill shortages are probably the ones that most closely match - since the aim of our product is to point out areas where users are lacking in. An issue encountered by the literature addressing skill shortage can be identified in the presence of friction in the labor market. More specifically the wages and working conditions offered might disincentivize qualified individuals from applying for otherwise suiting positions. Since the literature is mainly based on employer surveys, frictions can be mistakenly accounted as the effect of skill shortage. Nevertheless, even adjusting for the friction, researches such as Cedefop (2015) found that skill shortage still has a significant effect on labor supply.

Overall current literature, although conflicting about the magnitude, seems to agree that skill mismatches, independently of their nature, have a relevant and negative effect on companies' productivity, employees' salary, and thus on the overall economy. We can then conclude this is, in fact, a problem that our product can hope to tackle in an effective way.
To better understand skill mismatch both the supply side and the demand side need to be taken into consideration, since there are already existing studies based on the demand side of skill shortages, we plan to further investigate the supply side in our survey.

Another subject that has a large pre-existing literature background is the impact of hard and soft skills on the success in getting a job. It is proven by several sources that employers highly value a candidate's skill set since it plays a crucial role in their performance (WEF, 2020). While it is certain that both kinds of skills positively affect the employability of a candidate, we need to assess which has a more significant impact.

Over the past years the acquisition of soft skills has become increasingly more relevant, to the point that it is currently difficult to tell how they compare with respect to job specific knowledge. It is to be noted that oftentimes both types of skills are required concurrently, as noted in Andrews & Higson (2008). Moreover, broad employer surveys, across sectors - or even within the same sector, are difficult to interpret since different positions require a different set of specific hard skills, while soft skills tend to overlap across positions, further complicating the comparison between the two.

Although employers have been giving serious attention to soft skills, in sectors that require more technical skills, such as the energy sector, a significant correlation with productivity is still to be proved (Lyu & Liu, 2021).
Brüning & Mangeol (2020) have observed an emerging steady trend in sectors such as Information and Communication Technology (ITC), employers are requiring higher education credentials and skill proficiency in occupations, including medium and low skilled ones. Additionally, skill-biased technological improvements are leading to larger wage increases in high-skill positions.

On a different note, still according to Brüning & Mangeol, since industries such as ITC are constantly growing and evolving, when evaluating skill demand, we need to accentuate the relevance of its dynamic nature. While on one hand this might imply the significance of transferable skills, on the other hand it highlights the necessity for timely and precise information regarding the demand side, which aligns with the purpose of our product.

Overall it is undeniable that both hard and soft skills have a significant impact on the success in getting a job and on the wage patterns (Deming & Kahn, 2018). However, for the purpose of our project, we still need to investigate which factor has the higher explanatory power.

Survey

After having developed our causal map, estimated our probability of success and evaluated existing literature, we wanted to test our theories with an experiment. The goal of this experiment was to examine our causal links to determine their actual relevance and, if applicable, assess their strength in comparison to our theoretical assumptions.

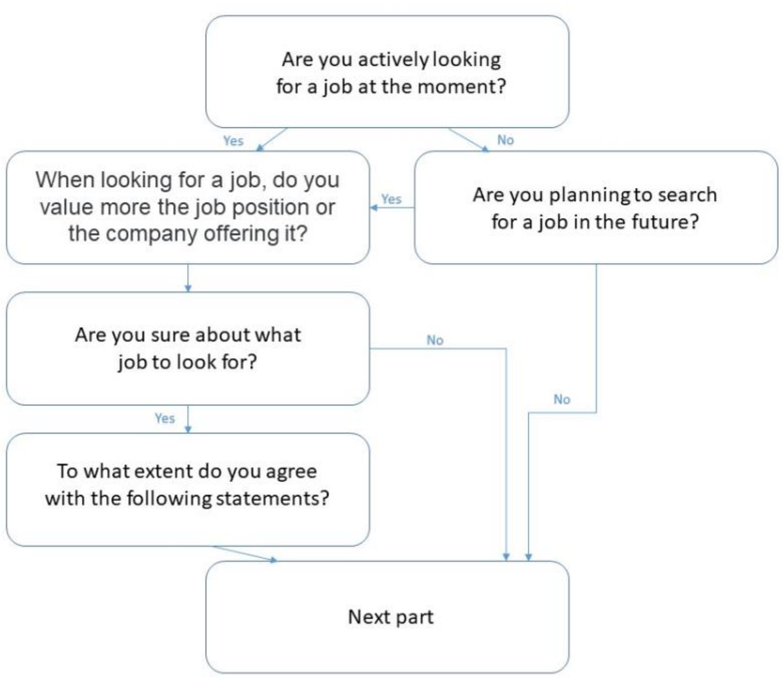
This experiment was conducted through a survey, which was carefully designed to execute our experiment. The survey consisted of three parts: the first part focused on demographics, the second part included questions aimed at testing three of the five links: "Mismatch," "People looking for jobs," and "Position over company." The third part was more complex and intended to assess whether hard skills were more effective in landing a job compared to soft skills or viceversa, it also involved a random experiment, but we will discuss that later.

In the first demographic section, we collected information on age, gender, whether the subjects lived in Italy or abroad, and a crucial piece of information necessary for the third part of the survey: whether the respondents had a background in STEM or not. We collected this information in advance to use it as a filter later in the survey to distinguish between STEM and non-STEM respondents.

Data Collection

The second part of the survey was dedicated to testing three main attributes, following the structure depicted in Figure 1. We chose this structure to screen out individuals who were not seeking employment or had no plans to do so in the near future, aiming to enhance the data quality.

Figure 1:



With the first question in this session, "Are you actively looking for a job at the moment?" we are filtering out subjects who do not possess the necessary characteristics for our survey.
If the response to this initial question was "no," we would then direct our subjects to the question "Are you planning to seek employment in the near future?" If the answer was "yes," we believed that this specific group of individuals might still provide valuable insights into the job market, so we chose to redirect them back to the primary line of questioning. Conversely, if the response remained "no," we would have the subjects skip the remainder of this section, as their responses would not carry significance and would only introduce confusion into the dataset.
Now, we move on to the primary line of questioning, starting with the second question: "When seeking a job, do you prioritize the job position or the company offering it?" This question is designed to assess the "Position over Company" attribute. Our expectation for this question was a roughly equal split, with about 50-50 responses.
The subsequent question served as another filtering mechanism: "Are you certain about the type of job you are seeking?" This question enabled us to identify the appropriate individuals to respond to our final set of questions, which aimed to evaluate the "Mismatch" attribute through a series of questions employing a Likert scale.
Our subjects were presented with the following statements and asked to provide responses on a Likert scale.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I am aware of the specific requirements, skills and experience needed to apply for the job I am looking for	○	○	○	○	○
I am aware of how I rank among my peers looking for the same job	○	○	○	○	○

This marks the conclusion of the second part of the survey.

Experiment

In the third and final segment, we assessed the "Hard vs. Soft Skills" attribute through a specific experiment. Initially, we categorized respondents into two groups: STEM and non-STEM, using a filtering question in the first part. These two groups were presented with two distinct descriptions of an arbitrary individual named Mario Rossi, which included his field of study and skill set. Subsequently, the groups were tasked with estimating the likelihood of Mario Rossi securing a specific job position (with different job positions specified for each group). Now, the critical and actual test phase commenced. In the subsequent question, respondents were informed that Mario Rossi had either improved or acquired one hard skill or one soft skill (from a random selection of three possible soft skills and three possible hard skills). They were then asked to reassess the probability of Mario obtaining the same jobs. Our primary interest lies not in the absolute values of the collected probabilities but in the deviation from the baseline when Mario Rossi acquired either a soft skill or a hard skill. We anticipate that the increase in probability for hard skills will be greater than that for soft skills, underscoring the greater importance of hard skills. To facilitate a clearer understanding of this section, we will now provide Mario's descriptions and the respective job positions for both STEM and non-STEM interviewees.

For STEM interviewees:
“Mario Rossi has a bachelor’s degree in economics and management, currently pursuing a master’s degree in business Analytics. Moreover, he has a basic knowledge of R, Python, and SQL, average knowledge of Microsoft Office, and a proven track record of good abilities to work collaboratively in a team. He applied for the position of consultant, data scientist, and data analyst.”
Possible hard skills:

1. An advanced level of proficiency with Python
2. An advanced level of proficiency with R and SQL
3. A professional level of expertise in Excel.

Possible soft skills:

1. Proven records of good leadership skills
2. Proven records of great communication skills
3. Proven records of great problem-solving skills

For non-STEM interviewees:
“Mario Rossi has a bachelor’s degree in economics and management, currently pursuing a master’s degree in business Analytics. He has basic financial modeling skills, average knowledge of Microsoft Office, proven track record of good abilities to work collaboratively in a team. He applied for the position of consultant, M&A analyst, and data analyst.”
one hard or soft skill and then evaluate the probability of getting a job.
Possible hard skills:

1. Python fundamentals
2. Improved financial modeling skills
3. A professional level of expertise in Excel.

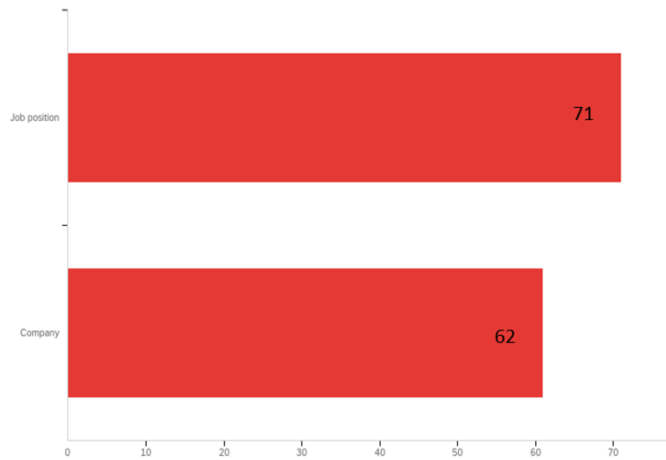
Possible soft skills:

1. Proven records of good leadership skills
2. Proven records of great communication skills
3. Proven records of great problem-solving skills

In order to pick which hard and soft skills would have actually been more relevant we scraped LinkedIn job postings and analyzed occurrences of skills in each posting. Insights of this analysis were also leveraged as a preliminary study for the popularity of hard and soft skills.
(While scraping is usually considered a gray area, we did some research and according to the European Data Act, proposed by the Commission in February 2022, web scraping of public commercial data that is not subject to copyright, or privacy laws is legal in the EU)
We are now ready to present the results of our analysis.

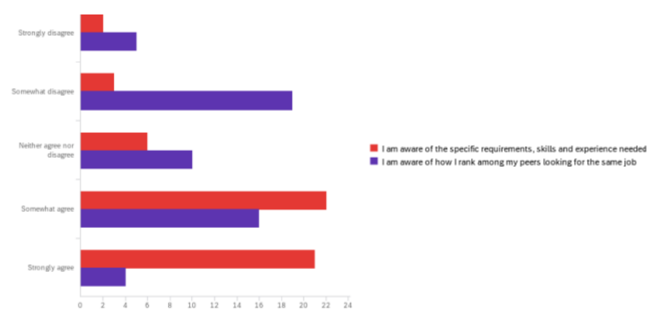
Data Analysis

Do people sort by position or by company?



A slight majority of the survey respondents (53.79%) consider the job position to be more important than the company itself when looking for a job. The relatively close percentages indicate that both factors—job position and company—are crucial considerations for job seekers. We should also consider that this question was posed to people who were already aware of what job to look for, which could also imply that some of them were also aware of what firms were offering the most interesting opportunities for their desired job. In other words, they are still concerned about the job position, and they are not willing to be employed in a different job even at their dream company.

Skills Mismatch



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance
1	I am aware of the specific requirements, skills and experience needed to apply for the job I am looking for	1.00	5.00	4.06	1.03	1.05
2	I am aware of how I rank among my peers looking for the same job	1.00	5.00	2.91	1.14	1.31

There is a high level of self-reported awareness among the respondents about the specific requirements, skills, and experience needed for the jobs they are seeking. Respondents show less certainty about how they rank among their peers in the job-seeking process. This is shown by the more evenly distributed responses across the Likert scale, a lower mean, and greater standard deviation and variance.

The marked difference between the two means suggests that while people may feel confident about their own qualifications, they are less certain about how their qualifications compare to those of other job seekers. This might imply a need for better mechanisms or tools that help job seekers assess their competitiveness in the job market.

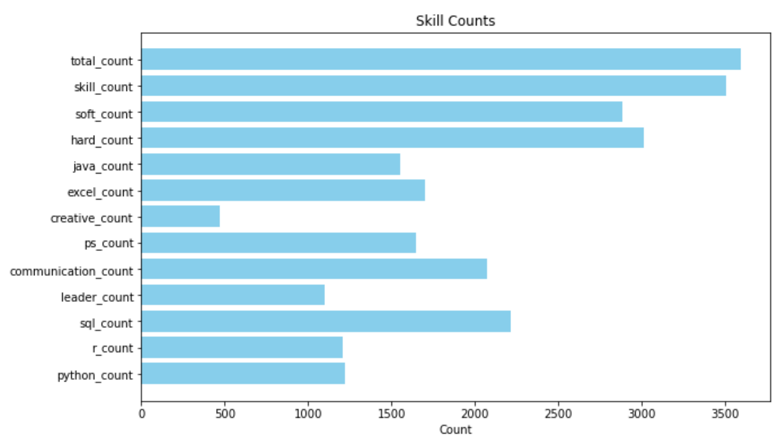
Importance of skills

In this section, we explore how important different skills are for getting a job. We started by looking at job descriptions from LinkedIn to see which skills employers are looking for.

Afterward, we looked at how having certain skills affects people's chances of getting a job, using a Difference-in-Difference regression to compare the importance of 'hard' skills, like programming, to 'soft' skills, like communication.

Secondary data analysis

As a preliminary analysis for considering to what extent skills matter, we scraped LinkedIn job descriptions and analyzed occurrences of each skill. We performed the scraper on Python using the library *Selenium,* a very popular tool for online scraping. We queried on LinkedIn for job titles and scraped all results automatically. We integrate the dataset, that was composed of 750 job description, with another dataset from 2022 of job description found on Kaggle, that was composed of 2845 observations from the US. Diversifying sources of data, as long as they were comparable, enable us to generalize the results across time and be somewhat independent of trend. We shortlisted some skills that, based on our own evaluation, were the most common appearing in job descriptions. We plot the results and obtained this:



{'total_count': 3595, 'skill_count': 3510, 'hard_count': 3012, 'soft_count': 2884, 'python_count': 1223, 'r_count': 1205, 'sql_count': 2217, 'leader_count': 1100, 'communication_count': 2075, 'ps_count': 1650, 'creative_count': 471, 'excel_count': 1699, 'java_count': 1556}

As we can observe, The 'total_count' bar is the longest, indicating the cumulative mentions of all observations. 'python_count' and 'r_count' are the counts for specific programming languages, with Python seemingly mentioned more frequently than R. Soft skills such as 'communication_count', 'leader_count', 'ps_count', and 'creative_count' are also tracked, with communication skills being notably prevalent. The counts for 'soft_count' and 'hard_count' indicate the overall demand for these types of skills, with hard skills being more commonly sought after.

Based on this analysis, we can have an initial understanding on the importance of skills in the job market. First of all, nearly all of the job description analyzed mentioned some sort of skills in the description, and most of them at least 1 hard and 1 soft. This is testament of their importance. Seemingly, there is a slight advantage for hard skills in this context, even though soft skills are very frequent also.

Primary data analysis

We used the data about demographics collected in the first section as controls variables. We want to test whether to focus on soft skills or hard skills in our proposal. We will use a DifferenceInDifference framework, which will be explained more carefully later on. We extracted the dataset referring to the last section from Qualtrics to Stata, which is the tool we will be using to perform regression analysis. Several columns that are not needed for analysis, such as *StartDate, EndDate, ResponseType,* etc., are dropped. This step simplifies the dataset by removing irrelevant information. We divide the dataset according to the education of the respondent, in order to separate the analysis. For each response, we consider the average of the probability of landing 3 different (but comparable) jobs. This way, our analysis is more robust to different job. The final dataset looks like this:

proba	after	treatm~t	gender	location	age_or~d	id
.6666667	1	1	Female	Italy	18-21	11
.77	1	0	Male	Italy	22-25	23
.4533333	0	0	Female	Italy	18-21	16
.2166667	0	0	Male	Italy	22-25	44
.6366667	1	1	Female	Italy	22-25	8
.2733333	1	0	Male	Italy	22-25	44
.7333333	1	1	Female	Abroad	22-25	15
.2733333	1	0	Male	Italy	22-25	51
.58	1	1	Male	Italy	22-25	48
.54	1	0	Male	Italy	18-21	22

- *proba*: Average of probability of landing a job proposed by each respondent.
- *after*: A dummy variable indicating if the recorded probability refers to the ex ante (0) or ex post(1) probability.
- *treatment*: Treatment variable, another dummy indicating whether the individual is part of the treatment (1) or control (0) group. The treatment group refers to observing the candidate developing an hard skill, whereas the control group refers to observing the candidate developing a soft skills.
- *gender*: Categorical variable indicating the gender of the individual, a control variable.
- *location*: Categorical variable indicating the geographic location of the individual, either Italy and Abroad.
- *age_ordered*: an ordinal variable indicating the age range category of the individual, such as 18-21, 22-25, etc.
- *id*: An identifier variable, unique to each individual or observation in the dataset.

The fact that we have 2 observations for each ID makes our dataset a panel, therefore we would need to include fixed effects in our analysis.

For example, if we look at the 2 observations with id = 44 in the excerpt of the dataset found above, we can see that the average the respondent provided initially was 0.21, and was updated to 0.27 after he was told that the candidate developed a soft skill (given that treatment is 0).

We may want to observe if the randomization was carried out correctly by Qualtrics. We do so by running Balance Tables assuming after equal to 0. This way we can observe eventual differences in the assignment of the two groups, not only in terms of demographics but also in terms of proposed probability given the same candidate.

```
preserve
keep if after == 0
iebaltab age_1 age_2 age_3 age_4 gender_1 gender_2 location_1 location_2 proba,
grpvar(treatment) savexlsx(bt.xlsx)  replace
restore
```

The result is:

		(1)		(2)		(1)-(2)
		0		1		Pairwise t-test
Variable	N	Mean/(SE)	N	Mean/(SE)	N	Mean difference
age_1	160	0.263	122	0.377	282	-0.115**
		(0.035)		(0.044)		
age_2	160	0.650	122	0.574	282	0.076
		(0.038)		(0.045)		
age_3	160	0.037	122	0.000	282	0.037**
		(0.015)		(0.000)		
age_4	160	0.050	122	0.049	282	0.001
		(0.017)		(0.020)		
gender_1	160	0.350	122	0.426	282	-0.076
		(0.038)		(0.045)		
gender_2	160	0.650	122	0.574	282	0.076
		(0.038)		(0.045)		
location_1	160	0.425	122	0.230	282	0.195***
		(0.039)		(0.038)		
location_2	160	0.575	122	0.770	282	-0.195***
		(0.039)		(0.038)		
proba	160	0.451	122	0.511	282	-0.060**
		(0.016)		(0.017)		

Where we observe that, actually, there are some problems in how Qualtrics carried out the randomization. This may be due to the limited number of observations collected. Anyhow, we remind ourselves to perform regressions including robust standard errors. The fact that we will be using a DifferenceInDifference framework ensures that the imbalance between the two groups' observed probability before the treatment not a real problem. In fact, the DiffinDiff regression is able to capture the heterogeneity across the two groups at after = 0 and isolate the true difference among the two groups caused by the treatment, which is what we want to determine. Thus, our equation would look like:

$$y_{it} = \beta_0 + \delta_0 \cdot dAfter_t + \beta_1 \cdot dTreated_i + \delta_1 \cdot dAfter_t \cdot Treated_i + \epsilon_{it}$$

Where the error term includes also the individual fixed effect. We are looking for δ_1, that gives us the effect of our experiment.

We did separate regression analysis for the two groups, which can be found in the appendix. Here, we show the final output of the regression of the merged dataset of Stem and Non-Stem, where we have included a column indicating if the individual studies a degree in a Stem-related subject or not.

```
reg proba treatment i.after intera gender_enc location_enc educ_enc i.id, r
```

The regression output is:

proba	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
treatment	.1844962	.0280511	6.58	0.000	.1290342	.2399582
1.after	.0527083	.008487	6.21	0.000	.0359281	.0694886
intera	.0476742	.0138469	3.44	0.001	.0202964	.0750519
gender_enc	.3283333	.0811645	4.05	0.000	.1678567	.48881
location_enc	-.0066667	.0582111	-0.11	0.909	-.1217604	.1084271
educ_enc	-.3716666	.0786005	-4.73	0.000	-.5270738	-.2162595
id						
2	-.08	.021367	-3.74	0.000	-.1222464	-.0377536
3	-.7866666	.1303358	-6.04	0.000	-1.044364	-.5289696
4	-.4716666	.0886076	-5.32	0.000	-.6468596	-.2964737
5	.145	.0823855	1.76	0.081	-.0178908	.3078909

Note: the intercept is equal to 0.55

We can observe that all coefficients are significant. As anticipated from the balance tables, the coefficient of *treatment,* signalling a non-perfect randomization, is positive and significant. However, also the coefficient of the interaction, *intera,* is positive and significant. The coefficient of 0.047 tells us that, on average, people who were shown a candidate equipped with hard skills believed that the candidate was, on average, 5 percentage points more likely to land a job. The after coefficient is also positive and significant, which implies that both hard and soft skills do matter in evaluating a candidate.

Thus, our estimated equation is:

$$y_{it} = 0.5 + 0.05 \cdot After_t + 0.18 \cdot Treated_i + 0.047 \cdot After_t \cdot Treated_i + \gamma \cdot X_i + \epsilon_{it}$$

Where γX_i include any possible controls that may or may affect our variable. We now may want to observe if results are any different in Stem vs. non-Stem. Thus, we include an interaction coefficient with our education variable. The output regression is shown below:

Linear regression

Number of obs	=	282
F(143, 138)	=	4228.29
Prob > F	=	0.0000
R-squared	=	0.9599
Root MSE	=	.05689

proba	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
treatment	.1831319	.0304897	6.01	0.000	.1228446	.2434193
1.after	.0527083	.0085177	6.19	0.000	.0358663	.0695504
1.intera	.0450336	.0135213	3.33	0.001	.0182978	.0717694
educ_enc						
Stem	-.3743512	.0802999	-4.66	0.000	-.5331285	-.215574
intera#educ_enc						
1#Stem	.0053692	.0221481	0.24	0.809	-.0384243	.0491626
gender_enc	.3283333	.082761	3.97	0.000	.1646898	.4919769
location_enc	-.0066667	.0584217	-0.11	0.909	-.122184	.1088507
id						
2	-.08	.0215032	-3.72	0.000	-.1225184	-.0374816
3	-.7866666	.1314783	-5.98	0.000	-1.046639	-.5266941
4	-.4716666	.0894359	-5.27	0.000	-.6485086	-.2948247
5	.1423154	.0840347	1.69	0.093	-.0238466	.3084775

We see a coefficient that is not significant with a very high probability. This means that the effect of the treatment across the two groups is the same.

As for whether skill matter, we can take a look at the *after* coefficient, which is positive and significant at <1%. This means, as expected, that the extra skill had an impact on improving the odds of the candidate to obtain a specific job. This happened even though some of the skills include were not relevant in one of more workplaces.

Insights

The insights from our analysis provide a promising outlook for our business, especially since a slight majority values job positions over the company. This aligns well with our product's focus on job roles, suggesting a potentially higher demand.

The presence of a skills mismatch is critical for our product's relevance. If job seekers were completely aware of and matched to job requirements, there would be little need for our services. However, the data suggests that while job seekers are aware of the skills required, they are uncertain about their competitiveness. This gap underscores the necessity of a tool that aids in assessing and enhancing job market standing.

Regarding the importance of skills, our analysis indicates that skills are certainly a priority for job seekers. Job seekers who focus on skill enhancement are likely to value tools that provide insights into skill gaps and recommendations for improvement.

Lastly, results from the experimental part of our survey suggest that hard skills are more effective than soft skills in boosting candidates chance in landing a job.

Update of probabilities

Position over company

According to our survey, we gathered information that a quite similar number of people value position and company as same important. Our previous estimations are strongly underestimated. It seems rational to set a new probability of people valuing position over the company as around 50% which allows us to be more optimistic about the success of our business.

Mismatch

Mismatch refers to the disconnect between a user's profile and the requirements of their desired roles. The mismatch is detrimental as it lowers a user's chances of securing suitable positions, leading to frustration and reduced user satisfaction. Mitigating mismatch through accurate assessment, targeted guidance, and personalized recommendations is crucial for enhancing user success rates. Our research shows that low self-confidence in comparison to other applicants may hugely increase the demand for our product.

The mismatch can also have a much higher influence on our business's success than we previously supposed. As we previously mentioned it can lead to big dissatisfaction on a professional level. Therefore, we decided to increase these values in our theory.

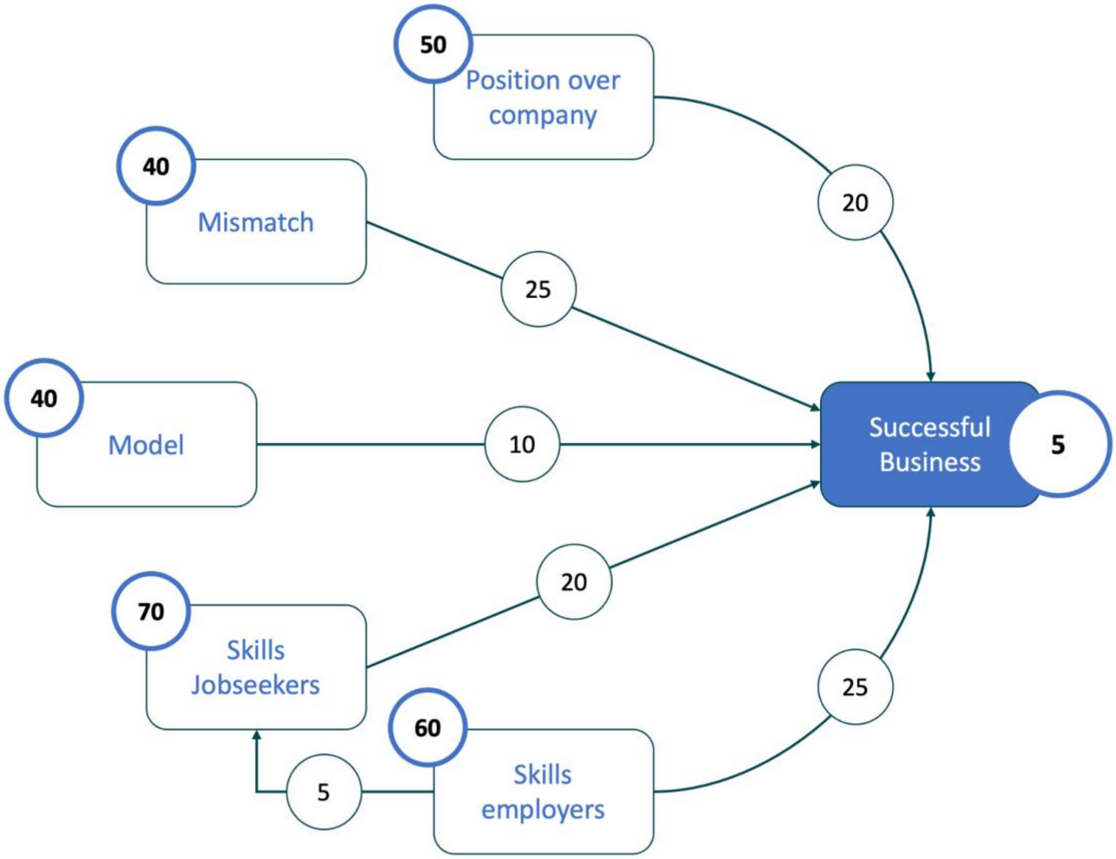
Skills matter for jobseekers

We have observed, as we expected, that skills are very important for jobseekers, therefore we increased the probability to 70%.

Skills matter for employers

Given our literature review, we have observed how weight employers put on in-demand skills, therefore we have increased our attribute to 60%.

After all the experiment we conducted we updated our theory in the following way:



This puts our probability of success assuming our theory is true at around 56%, and since we have done some testing we can be a bit more confident and assume that the probability that our theory is true is 20% higher than before, so around 70%. We also updated out probability in case our theory is not true, as we received a lot of good feedback and interest in the product from our colleagues. Then our overall probability of success is going to be 51%.

$$V = \omega V_{\Theta} + (1 - \omega) V_{\hat{\Theta}}$$
$$\omega = 0.56$$
$$V_{\Theta} = 0.7$$
$$V_{\hat{\Theta}} = 0.4$$
$$V = 0.51$$

Conclusions

Let's take a moment to reflect on our journey and where we've landed. When we started, our mission was clear: to ease the often daunting transition from academic life to the professional world for students and recent graduates. Our initial strategy was built around an AI-powered platform – think of it as a personal career guide, mapping out career paths, assessing skills, and connecting users with the right job opportunities. Initially, we were looking at a 26% chance of success with this idea. It was a solid start, but we knew there was room to grow.

We also explored an alternative path, leveraging ChatGPT as a virtual interview coach. This approach stemmed from a different perspective on the recruitment process, where the emphasis shifted from skill sets to the candidate's proficiency in interview situations. It suggested that perhaps the art of a well-conducted interview could be as crucial as the skills themselves. The potential success here was slightly better, pegged at around 29%.

However, a pivotal transformation occurred following our analysis of data and feedback. Through meticulous survey results and data examination, we significantly refined our original concept. The outcome of these efforts is our revised theory, which now boasts a 51% probability of success.

Consequently, we are advancing with this updated strategy, a fusion of innovative thinking, empirical insights, and comprehension of the challenges and goals of our target demographic. We are dedicated to making a significant and impactful contribution in facilitating the transition between academic achievements and professional success, advancing methodically and thoughtfully.

References

McGuinness, S., Pouliakas, K., & Redmond, P. (2018). Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys*, *32*(4), 985-1015.

OECD. (2013) OECD Skills Outlook 2013: First Results from the Survey of Adult Skills. Paris: OECD Publishing.

Desjardins, R. and K. Rubenson (2011), "An Analysis of Skill Mismatch Using Direct Measures of Skills", OECD Education Working Papers, No. 63, OECD Publishing, Paris.

Verhaest, D., Omey, E. (2012). Overeducation, Undereducation and Earnings: Further Evidence on the Importance of Ability and Measurement Error Bias. J Labor Res 33, 76–90.

Cedefop. (2015) Skill Shortages and Gaps in European Enterprises. Luxembourg: Cedefop Reference Series102, Publications Office of the European Union.

World Economic Forum, J. (2020). The future of jobs report 2020. *Retrieved from Geneva*.

Andrews, J. and Higson, H. (2008) Graduate Employability, ‘Soft Skills’ Versus ‘Hard’ Business Knowledge: A European Study, Higher Education in Europe, 33:4, 411-422.

Lyu, W., & Liu, J. (2021). Soft skills, hard skills: What matters most? Evidence from job postings. Applied Energy, 300, 117307.

Brüning, N. and P. Mangeol (2020), "What skills do employers seek in graduates?: Using online job posting data to support policy and practice in higher education", OECD Education Working Papers, No. 231, OECD Publishing, Paris.

Deming, D., & Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, *36*(S1), S337-S369.

Appendix

proba	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.treatment	-.318471	.0186686	-17.06	0.000	-.3556299	-.2813121
1.after	.0574667	.0114987	5.00	0.000	.034579	.0803543
treatment#after						
1 1	.0402753	.0155847	2.58	0.012	.0092547	.0712958
gender	.1633333	.0242002	6.75	0.000	.1151641	.2115025
location	-.6083333	.030781	-19.76	0.000	-.6696014	-.5470652
age_ordered	-.0683333	.0210904	-3.24	0.002	-.1103127	-.0263539
id						
2	-.08	.0215408	-3.71	0.000	-.1228759	-.037124
3	.32	.0236729	13.52	0.000	.2728803	.3671197
4	-.0483333	.0423005	-1.14	0.257	-.1325304	.0358637
5	.13	.0390215	3.33	0.001	.0523297	.2076703

Regression for Non-Stem respondents

proba	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.treatment	.1641667	.0308527	5.32	0.000	.1024082	.2259252
1.after	.0447778	.0121786	3.68	0.001	.0203996	.0691559
treatment#after						
1 1	.0583333	.0230938	2.53	0.014	.0121061	.1045606
gender	-.3133333	.0860842	-3.64	0.001	-.4856496	-.141017
location	-.0083333	.0627005	-0.13	0.895	-.1338419	.1171753
age_ordered	.1	.0188786	5.30	0.000	.0622105	.1377895
id						
2	.0683333	.023299	2.93	0.005	.0216954	.1149713
3	-.0783333	.0683931	-1.15	0.257	-.215237	.0585703
4	-.1166667	.0336666	-3.47	0.001	-.1840576	-.0492757
5	.2583334	.0855675	3.02	0.004	.0870514	.4296153

Regression for Stem respondents

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	When looking for a job, do you value more the job position or the company offering it?	1.00	2.00	1.46	0.50	0.25	132

#	Field	Choice Count	
1	Job position	53.79%	71
2	Company	46.21%	61
			132

Showing rows 1 - 3 of 3

Detailed results of the survey question about job position or company

#	Field	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	Total
1	I am aware of the specific requirements, skills and experience needed to apply for the job I am looking for	3.70% 2	5.56% 3	11.11% 6	40.74% 22	38.89% 21	54
2	I am aware of how I rank among my peers looking for the same job	9.26% 5	35.19% 19	18.52% 10	29.63% 16	7.41% 4	54

Showing rows 1 - 2 of 2

Detailed results of the survey question about job market awareness