

A Work Project presented as part of the requirements for the Award of a Master's degree in Impact Entrepreneurship and Innovation and Business Analytics from the Nova School of Business and Economics.

Overcoming Financing Challenges for Impact: The Development of a Match-Making Platform for Impact Organizations and Public Funding Opportunities – The Role Of Technology

Moritz Constantin Gütersberger

Work project carried out under the supervision of:

António Miguel

Leid Zejnilovic

24/01/2023

Table of contents

1	<i>Introduction</i>	2
2	<i>Literature Review</i>	3
2.1	Definition of impact investing	3
2.2	Definition impact start-ups	5
2.3	Impact Measurement	6
2.4	Current sources of funding for Impact Startups	7
2.4.1	Definition of different forms of finance	9
2.4.2	Financial Instruments connected to financing institutions	11
2.5	Deep-dive Grants Funding and the potential of EU and National Funds	12
2.5.1	Historic background and purpose of government grants	12
2.5.2	Structure of government grants: direct, indirect & shared management	15
2.6	Budget assessment	17
3	<i>Problem statement</i>	18
3.1	The causes: lack of promotion, dispersed information, and lengthy application process for grants	19
4	<i>Benchmark of solutions</i>	21
4.1	The role of consultancies	21
4.2	Current tools and their limitations	22
5	<i>Our solution</i>	24
5.1	Market analysis and target market	25

5.2	The Company's Assets	26
5.3	Customer acquisition and monetization	27
5.4	Future overlook and trend analysis	28
6	<i>The Role of Technology</i>	29
6.1	The Role of Data on the Platform	29
6.2	Data Scraping Process.....	30
6.3	Data Engineering with Keywords	31
6.4	Keyword extraction with RAKE_NLTK.....	32
6.5	Determine the similarity between keywords	33
6.6	Analysis of Results	34
6.7	AI Development: The Roadmap.....	36
6.8	Interpretation and business case	39
7	<i>Conclusion</i>	40
8	<i>Appendix</i>	42
8.1	References:	42
8.2	Figures	46
8.3	Tables:.....	52
8.4	Abbreviations:.....	53
8.5	Python Implementation.....	54
8.5.1	Web Scraping.....	54

8.5.2 Matching Model..... 63

Abstracts

Group: This paper examines the challenges faced by impact start-ups and NGOs in obtaining sufficient financial means to bring their vision to life and proposes a solution to the challenge of efficient funding distribution in the impact investment ecosystem. The paper presents the development of the JamBank platform, which connects impact start-ups and NGOs with the best funding options from public institutions, such as grants and loans, through a sophisticated matching approach. The paper discusses the obstacles faced by organizations in the funding process, provides an overview of the funding landscape, and proposes a goal-oriented economic approach to address these challenges. The paper also offers insight into the technical implementation of the JamBank platform and its potential to provide a valuable solution to the challenge of efficient funding distribution for impact start-ups and NGOs.

Individual: This paper aims to provide practical insights into data collection and the development of a data-driven matching platform, showcasing the potential of automated digital platforms in the venture development environment. Through documented implementation and a long-term strategy, technology serves as a core strategic element in reducing information asymmetries and facilitating efficient resource allocation. The paper emphasizes the combination of real-world data and an analysis of its potential.

Keywords: Impact Investments, Entrepreneurship, Innovation, EU Funds, Keyword-Matching, Webscraping

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

In recent years, the emergence of impact start-ups and NGOs has solidified their place as significant players in the financial ecosystem. Impact investing aims to create both positive social and environmental impact and financial returns, while impact start-ups strive to achieve both financial and social or environmental returns. This project focuses on addressing one of the most significant obstacles in the impact investment ecosystem: the efficient distribution of funds. Although impact ventures and initiatives often have the best of intentions, they frequently struggle to obtain sufficient financial means to bring their vision to life. This is often due to the dispersed and intransparent nature of funding sources, requiring organizations to invest significant resources in finding the appropriate source of funding. With JamBank, our team has developed a platform that connects impact start-ups and NGOs with the best funding options from public institutions, such as grants and loans. The platform employs a sophisticated matching approach to determine the similarity between an organization and available projects, allowing both sides to benefit. Despite some limitations, the platform provides a valuable contribution to the allocation of funds for meaningful ventures in a more effective manner. Through this project, we aim to highlight the current obstacles that organizations face in the funding process, provide an overview of the funding landscape, propose a goal-oriented economic approach, and offer insight into the technical implementation of the platform. This paper asserts that the JamBank platform offers a valuable solution to the challenge of efficient funding distribution for impact start-ups and NGOs, by connecting these organizations with the best funding options from public institutions and utilizing a sophisticated matching approach to align their goals and needs.

2 Literature Review

In order to develop an entrepreneurial idea with an academic perspective, literature plays a crucial role in outlining the context of what was achieved in the sector of impact investment involving grants making and funds from governmental organizations focus on the European Union settings. Within this research, a wide perspective of the impact investing world will be given, illustrating the current infrastructure together with its actors that play a crucial role in its development. After that, a deep dive into the most common forms of funding and the respective providers of capital will be done, emphasizing the work performed by the European Commission in providing financial support in solving the world's most pressing challenges. This academic review will be provided with the spirit of finding the best tools and resources to be leveraged by changemakers in the transition to more sustainable ways of living for our society and for our planet.

2.1 Definition of impact investing

The field of finance is a large and diverse ecosystem whose role in the broader economy and the environment can be crucial for achieving critical objectives in our society. The efficiency of capital allocation presupposes the long-run sustainability of natural and human ecosystems within which economic activities occur (Bose, Satyajit, Dong; Guo, Simpson Anne, 2019). Mainstream investments ultimately hasten disease or even cause death in our communities and nature as a whole for the sake of profit. Industries like tobacco, gambling, and weapons violate human rights and our natural resources, consequently threatening the evolution of a more equal and sustainable world (Monitor Institute, 2022). For this reason, investors find themselves on the hunt for different profit-making opportunities that result in being more aligned with their values. What if, through our investment decisions, we could not only stop harming our planet and its inhabitants but also encourage the sustainable development of our communities?

Here is where Impact Investing comes in. The idea of investing for the common good is “moving from a periphery of activist investors to the core of mainstream financial institutions,” according to the Monitor Institute in its *Investing for Social and Environmental Impact*. The GIIN (Global Impact Investing Network) describes impact investing as: “**Investments made with the intention to generate positive, measurable social and environmental impact alongside a financial return**” (Monitor Institute, 2022). This definition implies that the investment meets at least the three following criteria (Vecchi, Veronica, Balbo, Luciano, Brusoni, Manuela, Caselli, Stefano. 2016):

- **Intentionality** of creating extra-financial impact before the investment is made;
- **Use Investment and Impact Data in Investment Design** that will be useful in contributing to social and environmental benefits;
- **Manage Impact Performance** which means that the results are often communicated in feedback loops that improve the investment chain and perfect the approach to generate even more impact;
- **Contribute to the Growth of the Industry** by the contribution of investors that share useful data about terms, conventions, and indicators relevant for determining strategies, goals and performance.

Measuring outcomes cover a major role in defining whether an investment is considered impactful or not. The metric should be targeted, direct, and potentially measurable through repercussions in society (UNICEF, 2014). Moreover, impact investing brings attention to the importance of managing processes to ensure the achievement of the targeted outcomes and ensure that the initial problem is appropriately addressed or not. For this reason, impact management is of utmost importance for investors and institutions involved. This means that

the invested parties are cautiously monitored on the progression toward stated objectives and occasionally advised on intended and non-intended outputs (Big Society Capital, 2020).

The more the new line of thought of measuring outcomes instead of output gains significance, the more investors realize the potential of impact investment markets in addressing the world's most critical issues in sectors, namely sustainable agriculture, renewable energy, conservation, microfinance, and affordable and accessible basic services including housing, healthcare, and education (Impact Investing Hub, 2022).

2.2 Definition impact start-ups

Although beneficiaries of impact investments are diverse in size and legal structure, for the goal of the entrepreneurial project developed through this research, it will be relevant to analyze a special kind of organization: Impact Startups. According to Big Society Capital, the leading financial institution devoted to social impact investment in the UK, an Impact Startup is defined as **“a startup whose purpose is to contribute to solutions that create positive change and over time enables and evidences strong impact performance through proportionate impact practice”**. This is an indication of the founders' strong desire to have “purpose” as the main value of the company's mission statement. The intended impact is translated into the effort of transforming the social or environmental problems of a large group of people or geographic area into sustainable, well-performing solutions. Examples of these solutions are products and services that reduce CO2 emissions or kilos of plastic avoided either during the manufacturing period or the sale itself (Big Society Capital, 2022).

A commonly accepted framework for identifying what qualifies as Social and Environmental Impact is the United Nation's Sustainable Development Goals (SDGs) (UNDP, 2022). If a startup works to solve one of the 17 goals above mentioned, it will be obvious that it is tied to a significant social or environmental issue, which is a defining characteristic of impact-driven firms (Impact Prosper, 2022).

Impact Prosper claims that it is not only possible but also the most responsible approach of doing business while doing good. Impact and business are not mutually exclusive concepts; instead, they support one another. In this sense, making an impact can be viewed as a commercial opportunity.

There are three common models adopted by founders of impact ventures:

For-Profit Business: Business with purpose, founded with the purpose of generating both, social/environmental value (measurable impact) and economic value (revenue);

Cooperative (COOP): Increasingly adopted by entrepreneurs, focused on creating a self-sufficient group of individuals who are connected by shared economic needs and social/environmental values through a co-owned and co-managed enterprise;

Non-Profit (ONG): In the absence of government action, it is frequently up to ONGs to fill this existing gap, contributing significantly to the creation of societies that are safer, healthier, and more equal.

2.3 Impact Measurement

Finding a framework to make impact tangible is as much important as measuring profit. Accessing this data allows informed decision-making for the management, other than an accountable reporting source for stakeholders and shareholders. Whether it is about the number of lives saved, CO2 reduced, or the total of animals rescued from polluted areas, the best-known impact measurement frameworks are:

UN Sustainable Development Goals: Internationally recognized standards and most commonly referenced by business organizations and investment institutions;

IRIS by Global Investment Network (GIIN): Includes performance indicators for social, environmental, and financial change based on answering simple questions: what, who, how much, contribution and risk.

Also gaining more and more popularity:

Social Value International (SVI): Helps define a business' Social ROI (Return on Investment) by supporting the reasoning on how the change affects stakeholders and how much;

The Impact Beacon: In a simple and more straightforward approach, it allows entrepreneurs to use facts and figures to identify, evaluate and articulate their impact.

By increasing the use of such impact measurement frameworks, organizations can have a compass to understand towards which direction their actions are going and how they can be improved. Based on this, organizations create their long-term strategy and reiterate the process until they reach the intended outcome. Providing concrete numbers increases confidence in the interaction with stakeholders and potential investors, who expect concrete results by the actions performed over time (Shultz, A., 2022). Furthermore, still very little data is available on successful impact business models which might nurture skepticism and prevent crucial actors from investing resources in the area.

2.4 Current sources of funding for Impact Startups

In order to effectively promote entrepreneurial solutions for the impact investment market, it is essential to comprehensively assess the existing infrastructure, comprehend the behaviors of the associated entities, and evaluate the mechanisms used to provide monetary resources to the recipients. Thus, impact investments can be seen as a marketplace where demand (impact-driven organizations), supply (impact investors in a broader sense), and intermediaries intersect to generate value together (Social Impact Investment Taskforce, 2014). The main components in the impact investment ecosystem are:

Impact-seeking purchasers: consumers, corporations, foundations, and the government provide the sources of revenues that underpin investment in impact-driven organizations;

Impact-driven organizations: either social sector organizations, impact-driven businesses, or any kind of organization that has a long-term social and environmental impact vision, sets outcome objectives, and measures their achievements;

These two sides interact through tools, channels and legal mechanisms that simplify the process in which the capital is channeled. Each of them will be briefly analyzed in terms of relevance and role on the overall chain, so then the focus can be drifted on some specific fund providers and forms of finance, which are at the core of the entrepreneurial solution proposed in the following pages. Figure 1. provides a broad overview of the current impact investments infrastructure market and how actors interactions are intertwined.

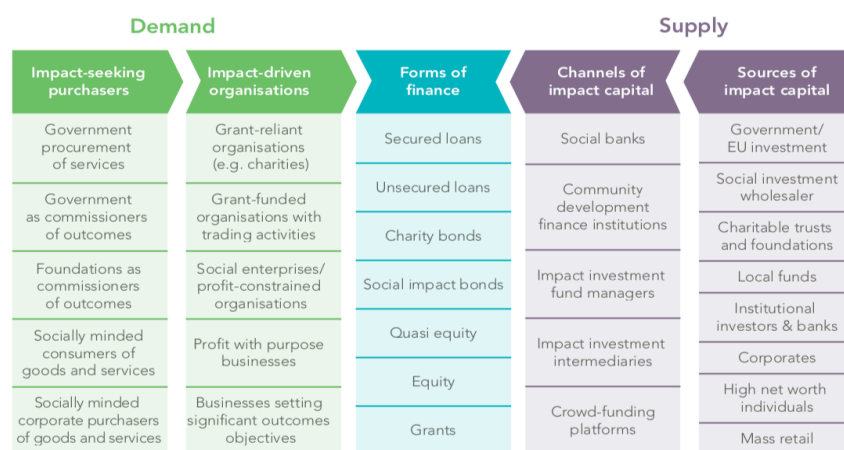


Figure 1. Social Impact Investment Taskforce, 2014. *Impact Investments: The Invisible Heart of Markets*. UK: G8

It is important to mention that social investment and grant-making lead to two different ways of funding the organization. The first aims at investing in financially sustainable enterprises whereas the second rather focuses on impactful results. Most investors that choose to invest on impactful projects are clearly more interested in the former, given the higher chances of return. Impact startups and social enterprises find themselves in the middle, trying to catalyze both funding opportunities and often setting up a separate entity to circumvent legal barriers and benefit from both (Social Impact Investment Taskforce, 2014).

2.4.1 Definition of different forms of finance

There are several ways of financing impact and entering the Impact Ecosystem as a monetary resource provider. The main ones that are recognized by EVPA (European Venture Philanthropy Association) are:

Social Impact Funds: Pioneers in impact investing in Europe, born out of the necessity to address pressing societal issues by deploying patient capital over a ten- to twelve-year period in high-risk social-impact projects with hardly any success track record. As a market facilitator, the EU Commission launched several programs that support social enterprises accessing these funds. Some examples are the EU Program for Employment and Social Innovation (EaSI), InvestEU, and the European Social Fund Plus (ESF+) which offer subsidies to cover transactional costs and support the management and maintenance on the side of the fund managers (European Commission, 2022).

Social Investment Crowdfunding Platforms: Created to bridge the gap between donors and financial beneficiaries by deploying equity to support early-stage and high-risk social enterprises, typically providing small tickets between EUR 50,000 and EUR 400,000 (EVPA, 2022);

Foundations: Many traditional foundations are adopting the engaged and strategic approach of venture philanthropy, including impact measurements, non-financial support (expertise and connections in the market), and long-term grant-giving, other than experimenting with other financial instruments like debt, equity, or hybrid (EVPA, 2022);

Corporate Social Investors (CSIs): Vehicles normally related to a company, and which aim to create social impact linked to companies. e.g. corporate foundations, corporate social businesses, corporate social investment funds, and corporate social accelerators. They play an important role in generating direct social impact in society and, on the other side, push the corporation to change its business practices to become more sustainable (indirect social impact);

Banks: Ethical banks can show that it is possible to achieve financial results while creating a positive impact, by supporting impact organizations. After the 2009 crisis, banks started to ponder the negative effects of their actions on society, which started demanding a more sustainable way of banking. Banks developed ESG funds with the issue of green and sustainable bonds, microfinance initiatives for the “underbanked”, and favourable loans with “pay-for-success” approaches to measure the impact of the investee projects. Moreover, they play a crucial role in the development of SIBs (Social Impact Bonds) in proactively building and tapping into their large network of stakeholders, other than counselling and advising their customers on the best impact investment opportunities according to their risk appetite.

Public Sector: National, regional and local governments have the mandate to provide services to all segments of the population, while giving special attention to the most disadvantaged. Additionally, these governments play a pivotal role in supporting initiatives to create services in collaboration with the private sector, as well as developing solutions to address pressing societal issues. In a recent OECD report, the government's role was identified in four distinct areas: steering (strategic creation of the market), ruling, financing (allocating capital towards impactful initiatives), and informing (sharing relevant data). In addition, the European Union has a prominent role in achieving sustainable development, job creation, and economic growth in the European Community. This role is facilitated by the availability of grants that address particular societal problems and can fund Impact and Social enterprises (OECD, 2022).

Other actors: Other players in the game are Institutional Investors (90/10 of solidarity funds are allocating 5% to 10% to unlisted social enterprises while the rest 90-95% of funds follow SRI principles), Family Offices (improving profile assessment for their clients) and Asset Management Companies.

Among the ones analyzed, the focus of this research will go towards to Public Sector and its unexploited potential. With this in mind, we will analyze the most common financial instruments to then narrow down even more to the ones relevant to the solution proposed.

2.4.2 Financial Instruments connected to financing institutions

The previously mentioned institutions have been actively involved in the development of innovative financial instruments that would intrinsically require the cooperation of different stakeholders. The last year witnessed collaborative partnerships among various types of organizations which would include governments and public institutions, private investors, impact startups and regulators. Only proactive engagement and expertise of different partners ensure the successful implementation of the following financial instruments:

Secured and Unsecured Loans: Loans with assets provided as collateral or not (CFI, 2022);

Charity Bonds: tradable loan from a group of social investors to a charity or social enterprise with a fixed period of time and rate (Good Finance, 2022);

Social Impact Bonds: Unique public-private partnerships that fund effective social services through performance-based contracts. Impact investors provide the capital to scale the work of high-quality service providers. The government repays those investors if and when the project achieves outcomes that generate public value (Social Finance, 2022).

Quasi-Equity and Equity: Quasi-Equity investments can be structured as debt but convertible into equity or preferred equity. They have a higher risk than senior debt but a lower risk than common equity (Interreg Europe, 2022).

Grants: Within the programs led by Governments and Public Sector like the European Union, there are some sections targeted to Employment, Social Innovation and Inclusive Entrepreneurship that, even if they are not explicitly part of a “Social Impact Investment

program”, address the same problems that the previously mentioned financial tools do (European Commission, 2022). This will be the one the research will focus on the most.

2.5 Deep-dive Grants Funding and the potential of EU and National Funds

A form of funding for impact startups often underestimated is grants and funding from governments and public institutions. More and more financial resources are poured by public institutions into social enterprises and NGOs and their potential is often undervalued or not sufficiently recognized (European Commission, 2022). In particular, the European Union has been actively involved in sponsoring initiatives of varying sizes and scopes to address certain problems within and beyond EU borders. Infrastructures, research, education, and the most essential economic sectors, including agriculture, are just a few examples of how it has an impact on the daily lives of Europeans. The results of the regional and cohesion policy efforts are under everyone’s eyes. However, their structure’s limitations affect a broader diffusion among the public. In the following pages, an overview of the background, goals and potential of EU funds will be provided. Consequently, a more throughout analysis of their application process will highlight the limitations and strengths of the current model, laying the foundation of a solution to the identified weaknesses.

2.5.1 Historic background and purpose of government grants

The European Union's Cohesion Policy was born out of the Treaty of Rome in 1957, wherein Shuman highlighted the need to reduce disparities among regions and foster a unified market. Subsequent treaties and agreements saw the introduction of resources, decentralization, and cooperation, providing a flexible policy to meet the needs of its members and support the most disadvantaged regions. In the period from 1985 to 1995, several community initiatives were launched, such as Interreg for cross-border collaboration, Leader for local and rural development and Resider, for industrial area renewal. From 1983 to 1993, GDP increased by

3%, with 600,000 jobs created through Structural Funds in Greece, Ireland, Portugal, and Spain, and nearly one million individuals trained through the European Social Fund. In 1994, the Amsterdam Treaty established the Cohesion Policy, with a focus on urban areas and doubling the funds available to foster regional development. The 2000 Growth Strategy allocated €200 billion to support social inclusion and advance the Lisbon Strategy, which sought to make the European Union the world's most competitive and dynamic knowledge-based economy by 2010. As more countries joined the EU, the disparity between its 250 regions increased, and so from 2007-2013, the Cohesion Policy was amended to place greater emphasis on job creation and innovation and to allocate additional funds to the most deprived regions. In 2021-2027, the same goals will be pursued, with even more resources to tackle current challenges.

As seen, the European Union has the privilege of being a safety net for its citizens by pooling together resources from all areas and countries in Europe to fight common challenges and create EU Added Value. It also presents a big funding opportunity for startups and impact organizations that aim to bring added value to society and the planet. But first, let's have a look at its history. Money comes in different forms, each one suited to different purposes and goals. Its allocation is carried out by strict rules to ensure tight control, in a way that the money is spent transparently and accountably (European Commission, 2022). The EU provides both repayable and non-repayable finance to fund social enterprises and environmental solutions. Moreover, grants and microcredit solutions are provided to specific vulnerable groups who find it hard to access conventional credit markets, for startups and existing micro-enterprises (European Commission, 2022).

Through the provision of both repayable and non-repayable finance, the EU has enabled the development of social enterprises and environmental solutions, as well as providing grants and microcredit solutions to vulnerable groups who are unable to access conventional credit markets (European Commission, 2022). In this way, the EU has become a safety net for its citizens,

pooling together resources from all areas and countries within Europe.

Moreover, the EU provides a significant funding opportunity for startups and impact organizations that are seeking to bring an added value to society and the planet. These organizations may be eligible for a variety of grants and loans, including those provided by the European Fund for Strategic Investments and the European Investment Bank (European Commission, 2022). Through this funding, the EU seeks to promote economic growth, create jobs, and reduce regional disparities across Europe. Consequently, the EU has become a major source of financial support for organizations that are dedicated to making a positive contribution to society.

The European Commission's objectives, such as fostering a more inclusive and sustainable Europe, are in line with the goals of impact investing, which endeavor to improve society, the environment, and the ecosystem. The Commission's calls and programs are crafted to meet the specific needs of the European context and form part of the strategic objectives that the European Commission hopes to fulfill with its 7-year EU budget. The primary areas of focus for the EU budget for the period 2021-2027 are:

Single Market, Innovation, and Digital: the EU has positioned itself as a driver of economic, technological, and social development through initiatives such as Horizon Europe, Invest EU, and the Digital Europe Program. Through these programs, it provides not only financial resources to support the implementation of projects that promote progress, but also alternative incentives such as tax reductions and facilitation of access to customers and beneficiaries outside the countries of operation.

Community Development and Reducing Inequalities: through programs like Erasmus+, the European Social Fund+, ReactEU, and the most common European Regional Development Fund (ERDF), the European Commission has the ambition of correcting imbalances between regions in terms of economic and social disparities, emerge stronger from a crisis like the one

brought by Covid-19. This is done also by tackling education, training, and by building more resilient health systems;

Natural Resources and Environment: Funds on agriculture, fishing and food sufficiency have always been significantly weighted on the overall EU budget over the years. These were recently implemented by funds focused on the environment (protecting and restoring natural flora and fauna of areas in danger) and in the transition towards a carbon-neutral economy.

Migration and Security: The Eu puts a lot of effort into Humanitarian Aid for communities who look for asylum. This goes together with ensuring security from crime, terrorism and radicalization.

The European Commission used an adaptive approach to policymaking in recent years in response to changing circumstances. This modest degree of flexibility made it possible to allocate resources to challenging situations during the health crisis Covid-19 pandemic and in geopolitical tensions like the Ukrainian conflict.

2.5.2 Structure of government grants: direct, indirect & shared management

The European Commission has implemented a policy whereby both internal and external institutions are engaged in the distribution of funds, in accordance with the mode of program implementation. Strict regulations are adhered to during the allocation of the funding regardless of the program, to ensure that transparency and responsible usage are closely monitored. There are three main implementation types that the EU follows to distribute its budget:

Direct management: EU funding is managed directly by the European Commission, which is responsible for all aspects of program implementation, including launching calls for proposals, evaluating submitted proposals, signing grants agreements, monitoring project implementation, assessing results, and making payments. Calls for proposals are periodically published in the European Commission's Funding and Tenders Portal. Direct management of the EU budget for 2012-2027 accounts for approximately 20% of the budget. The Next Generation EU, a

temporary recovery instrument, will also be implemented in direct management mode. Applications must be submitted through the Funding and Tenders portal of the European Commission.

Shared Management: This is a mechanism employed by the European Commission and national authorities, such as ministries and public institutions, to jointly run programs. It is estimated that approximately 70% of all EU programs are managed in this manner. Member states' administrations are responsible for overseeing selection and award procedures, monitoring, and managing operations, as well as payments. To inform citizens of calls for proposals, each Member State has established their own Funding and Tenders Portal. Recently, some of the Member States have included all country-specific programs on these websites.

Indirect management: The European Union (EU) employs indirect management of its budget through the delegation of budget execution tasks to partner organizations, both inside and outside the EU. Most of the EU budget is allocated to humanitarian aid and international development. Examples of partner organizations include international organizations such as the United Nations (UN), the International Monetary Fund (IMF), the World Bank, the European Investment Bank (EIB) and the European Investment Fund (EIF). Additionally, the EU works with local and regional authorities, national governments, civil society, and private sector representatives.

Therefore, the European Union provides the necessary financial support for numerous programs and initiatives, with member states and other organizations delegated the task of overseeing their day-to-day management. A part of the EU budget is allocated and managed by member states and delegated institutions, while the European Commission is responsible for ensuring strict supervision and proper expenditure of the funds.

Additionally, as previously mentioned, The European Union utilizes ad hoc organizations, such as the European Investment Bank and the European Investment Fund, the European Investment Council (EIC) and the European Investment Council, to distribute funding among its members. By issuing calls for proposals for certain initiatives and organizations, these organizations promote infrastructure development through co-sponsoring impact investing funds and creating innovative forms of finance, such as microloans, quasi-equity, and social impact bonds. Consequently, these bodies become closely intertwined with the private investing ecosystem and the beneficiaries of their funding.

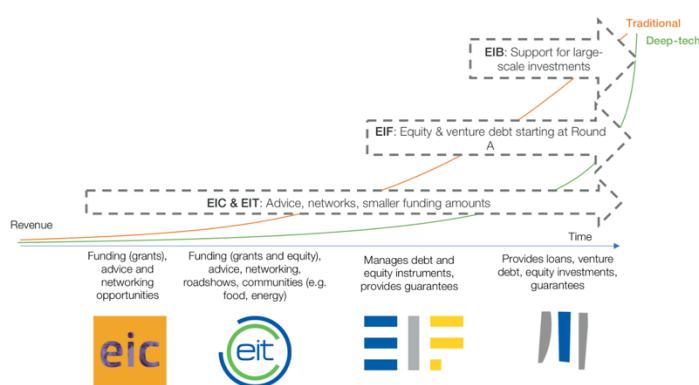


Figure 2. World Economic Forum. 2020. Bridging the gap in European Scale-up Funding: The Green Imperative in an Unprecedented Time.

(Funding from European Union in start-up lifecycle)

2.6 Budget assessment

The European Union budget is an essential tool for ensuring the continued democratic, peaceful, prosperous, and competitive position of Europe in the global arena. Its contribution to the single market allows for the collective addressing of issues such as immigration, terrorism, and climate change, which would otherwise be too difficult for individual countries to manage. In addition, the EU budget creates jobs, supports projects that improve health, education, transport, and energy infrastructure, builds EU supercomputer capacity and energy links, strengthens the security of borders, and helps to combat climate change and ease society's digital transition.

A prime example of the efficacy of the European Union budget is its response to the COVID-19 crisis. With the resources provided by the EU budget and Next Generation EU, every EU

country is able to address the economic and social consequences of the crisis and begin to recover. The 27 member states of the EU, representing a population of 450 million, agree on the EU budget and how it is to be financed each year, which typically ranges from €160 to €180 billion. The main sources for the 2021-2027 budget are customs duties, contributions based on the Value Added Tax collected by Member States, and direct contributions from EU countries, also known as Gross National Income-based contributions. Furthermore, as of January 1st, 2021, a new source for the EU budget is a contribution based on non-recycled plastic packaging waste. According to the latest research, members have a return on investment of one to six on their contribution to the European Union. Together, the long-term budget and the Next Generation EU, worth €2 trillion, can effectively address the economic and social consequences of the COVID-19 crisis. Figure 3. provides a graphical depiction of the most recent allocation of the European Union's budget.

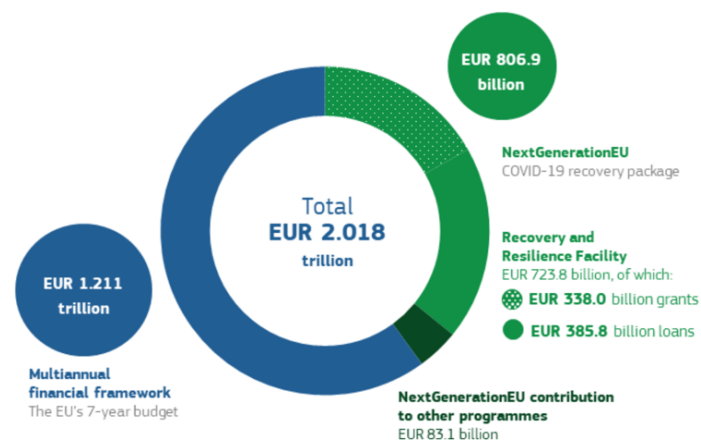


Figure 3. European Commission. 2021. Fact Check on the EU Budget

3 Problem statement

Now that an overview was given on the Impact Investing ecosystem, what impact startups are and their potential to solve current world challenges, the limitations of the current funding options, and what is causing inefficiencies. In particular, we will analyze grants and government

subsidies, their distribution, and management. This research will constitute a basis for the creation of an entrepreneurial idea that can solve these challenges.

3.1 The causes: lack of promotion, dispersed information, and lengthy application process for grants

The European funding environment is undergoing constant changes, with the European Commission playing a significant role in providing financial resources for innovation via public policy activities and direct funding to promising startups through the action of subsidized bodies such as the EIT, EIC and EIB (World Economic Forum, 2020). Despite this, information about such funding opportunities remains scarce and fragmented across the websites of each of these institutions responsible for its distribution. This lack of a unified platform for beneficiaries to access the best available funding opportunities has impeded their ability to locate the most suitable option for their area of expertise and size (World Economic Forum, 2020). Furthermore, the application process is often lengthy and full of technicalities, with prospective recipients needing to navigate the Funding and Tenders portal of the European Commission, (Deloitte, 2014) together with regional and other European websites to better understand the best option for their situation. This means that very specific criteria, language and value of each of the activities conducted have a big weight on the acceptance of the proposal. Strict rules and procedures are also set out to ensure that every euro is spent in a transparent and accountable manner. Additionally, the rigorous pre-, during and post-implementation reporting requirements may be off-putting to potential applicants, necessitating the introduction of a more modernized budgetary approach (European Commission, 2022). Consequently, start-ups and impact organizations need a clear public funding pathway from early-stage innovation funding to growth-stage investments.

Lack of Awareness & Complex Scouting Process

The potential of public funding is vastly underutilized due to a lack of awareness within the startup community. According to market research, startups are aware of this funding tool, however, they perceive the application process as lengthy and complicated, leading them to opt for alternatives, although expensive, such as crowdfunding and private capital from venture capitalists.

Organizations seeking funding from programs such as Erasmus+, Horizon Europe, and COSME must rely on the assistance of specialized consultants to guide them through the process, which entails scouting and the application itself. Additionally, they must sign in to the websites of the European Commission, its related institutions and the National institutions that provide funding to acquire the most suitable solution. Each program has its own language, such as Horizon Europe's scientific terminology, which can make the process even more challenging. The scouting and application process for a single call can take anywhere from 10 days to a month and is often tied to a specific timeline, which can range up to a year. The length of time required to receive feedback and capital aid can have a detrimental effect on the success of a newly launched startup (Deloitte, 2014).

Dispersed information among the websites

The dispersed information among websites presents a challenge for startups seeking funding opportunities. It can be difficult for them to keep up with the various calls across national and regional entities, as well as philanthropy entities associated with the government, with the lengthy scouting process often requiring specialist knowledge. The European Commission has developed a “funding for tenders” website that lists all funding opportunities from its programs, though there is scope to make it more user-friendly. However, this platform only encompasses the funds directly from the European Union, leaving out those from the European Investment

Bank and other bodies. Furthermore, there is a lack of knowledge of the indirect funds available from each country and those from EIT, EIC, and EIB, often compounded by bureaucracy processes that prevent startups from completing the steps necessary to obtain the financial resources needed (World Economic Forum, 2020). For example, the Portuguese government has worked to make financial incentives more accessible through the Portugal 2020 program and the current Portugal 2030 (Portugal 2030, 2021) through institutions such as IAPMEI, Startup Portugal, and the co-financing of startup accelerators like Startup Lisboa. However, startups from abroad may still face difficulties due to legal and language barriers in navigating the bureaucracy of these websites.

4 Benchmark of solutions

To assess the efficacy of a potential new solution to tackle the problem mentioned above, the current platforms must be examined in greater detail to understand where to orient the team's efforts. Additionally, the roles of the active players involved must be evaluated. These active players include start-ups seeking funding opportunities, users of the platform, consultancy companies offering services to simplify the application process for grants, and matching platforms that are attempting to solve the problem.

4.1 The role of consultancies

Consulting firms and independent contractors are essential in connecting organizations with potential financing opportunities, particularly those available through the European Union. The process of acquiring EU funding is highly competitive and requires not only an excellent project but also the ability to present the project to the funders and a comprehensive understanding of EU policy. To assist clients in taking advantage of these opportunities, we offer tailored project application and implementation services, ranging from identification of public grants, recruitment of consortia, bid writing, selection of experts, negotiations with granting

authorities, budgeting, quality control, and project implementation. Their business model typically consists of a success fee of 5-10% of the total funds obtained. This means that a €200,000 fund would result in a 10% fee of €20,000 for the consultant. Additionally, consultants can choose to join the funded organization as project managers and coordinate the financed projects for their duration, or simply provide advice for the application process.

4.2 Current tools and their limitations

There are several types of platforms already present that specialize in matching fund seekers and investors. Since this particular problem includes two parties, present solutions address this with multiple-sided platforms (MSPs). Erics Brousseau and Thierry Penard suggest three dimensions to compare the business model of digital platforms Economies of Scale, Network Externalities and Access & Quality. Network Externalities influence the impact of transaction costs and therefore aim to reduce the friction between both sides of the market. In this setting, transaction costs are represented by the time each user must invest to manually search through all sources to identify eligible funding sources. Secondly, Economies of Scale refer to the assembly process of functionalities or information available. For the funding platform, this is directly related to the efficiency and accuracy of the matching process, as well as the functions that are available. More data means more accuracy in matching the queries of the user, to which economies of scale apply. The third dimension is knowledge management, a vital aspect of the aggregation platform, which focuses on the effectiveness with which the data supplied by customers of digital goods is used to enhance services and develop. To conclude, effective knowledge management is the exchange of information at the core of the value creation of this platform.

There are multiple major digital aggregation platforms that are engaged in collecting information on European cooperation projects and attempting to match start-ups that seek funds in form of grants and investors. What these platforms all truly lack is the ability to connect those

who are considering government grants as a funding option to the right grant and the right partner to be successfully approved for this grant. In this section, we focus on analyzing competitors in order to identify gaps in value creation and showcase how our solution is able to fill these gaps leaving both users and customers with a better solution.

Up2Europe: This platform is designed around the idea to create a central port of call for finding a partner to pursue cooperation on a European project. In its endeavor to create a European ecosystem, the platform offers a variety of services to boost cooperation. The website hosts separate digital spaces for both project owners and project seekers to connect them with a partner. It is important to mention that there is no matchmaking in these spaces, it is rather designed as a feed with filtering options to scroll through projects (Up2Europe, 2022).

The matching process is conducted manually by the funding seeker by scrolling the project feed. Lastly, there is an accelerator platform that is designed for public authorities to gain access to project seekers within a distinct territory. This service specifically addresses local authorities to increase allocated funding to their region. It is noteworthy that the platform collaborates with authorities, aside from the main website. The result is individual websites that have a local focus tailored to the specific region, i.e., a site that focuses on the development of projects in the South of France. Businesswise, the platform monetizes the consultant side by offering a premium membership on the platform to increase visibility as a potential partner. Moreover, the platform generates revenue through designing and hosting sub-platforms, tailored specifically to local authorities (Mar Région Sud, 2022).

Dealflow.eu: The platform positions itself as a portal that connects EU-funded innovations with Investors and Corporates. It offers three core functions matchmaking and fundraising, venture building and tech due diligence, as well as networking and partnerships. It is worth mentioning that the 2020 Horizon Research and Innovation Program of the European Union provides

funding for Dealflow.eu. Powered by dealroom.com and Innovation Radar, the platform possesses a tendency towards traditional markets with a focus on profits rather than impact. Logically, the available funding sources are equity investments and not governmental grants. As presented earlier in this paper, this signifies a detrimental effect on the impact of an entity itself because the organization is now obliged to account for returns to investors. In conclusion, the platform has a distinct focus on the private venture capital market but does not serve as an ideal tool for impact organizations to gain access to funding.

The web analytics report reveals that there is a large gap in monthly users at multisided aggregation platforms (Monthly visitors in September: up2Europe >5k, dealflow.eu >33k) compared to the primary websites publishing information on grants and calls (Monthly visitors in September: Eureka Network 168k, EIF 96k, EEA 301K and European Commission 24M). This ambiguity supports our assumption that impact organizations currently rely mainly on primary sources to search for government grants. As monthly visitors do occur significantly on the primary website, we can conclude that a need for information in this area does exist, but it is not yet channeled toward a centralized source of information (Dealflow.eu, 2022).

5 Our solution

Given the analysis of the impact investment world and the potential and limitations of European Funds, the ultimate goal of this research project was to develop a full-fledged solution to address the problem of dispersed information on funds and grants among European public institutions' websites and turn it into a business opportunity. Through a combination of expertise acquired in postgraduate studies in Impact Entrepreneurship and Innovation on one side and Business Analytics on the other, the team was able to devise a match-making platform, JamBank, which connects impact organizations and public funding opportunities and incentives. This platform offers users a comprehensive range of funding opportunities directly from the European Commission, national governments, and EU-funded institutions such as the European

Investment Bank, European Investment Fund and the European Institute of Innovation and Technology by only uploading a small description of the prospective organization. It furthermore provides a variety of consultants to support the application process, to reduce bureaucratic barriers. In the future, it is anticipated that the platform will be further automated by technology and data gathering, allowing for streamlined access to the funds already available. This will ultimately free up resources for CEOs and finance departments of impact organizations, enabling them to focus on other areas of the business.

5.1 Market analysis and target market

This represents a tremendous opportunity for the European impact startup community to tap into the EUR 2 trillion of the budget made available by the European Commission for European citizens to thrive. We can consider the Total Addressable Market (TAM) as the total EU Budget available for the period 2021-2027 which relates to how much the platform could potentially serve. The Serviceable Addressable Market (SAM) is the portion of the TAM available to the platform, which, in this case, is EUR 600 billion, which corresponds to the funds dedicated to climate actions and social, according to DealFlow. In this case, based on market research which suggests that only 20% of this funds market can be realistically grasped, the total Serviceable Obtainable Market (SOM) represents EUR 120 billion for the next 7 years. This is data that cannot be avoided if considering the potential that can have in European impact startups. Figure 4. Provides a summary of JamBank's market analysis.

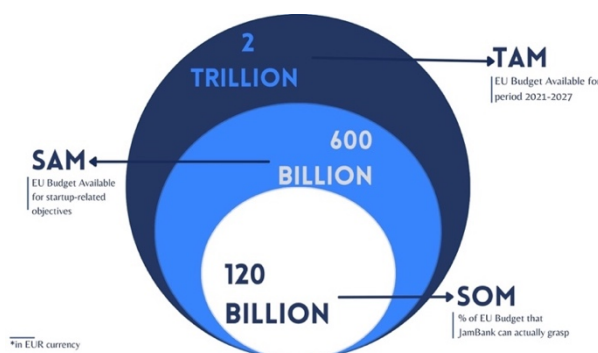


Figure 4. JamBank's market analysis

5.2 The Company's Assets

For the entrepreneurial concept to realize its potential success, not only must the requisite competencies and financial resources be present, but also a competitive and intrinsic advantage to which competitors cannot easily gain access. The combination of the following assets will show how JamBank can disrupt the startup and NGO funding market.

Key Resources: JamBank's key resources include its team members, who have expertise in both technical development and industry experience, as well as its platform and database of funding options. The team's technical and industry expertise will be crucial to developing and operating the platform, while the platform and database will be the main products offered to customers.

Key Partners: JamBank's key partners are likely to include public institutions that provide grants and loans, as well as consulting companies that assist startups with grant applications. These partners will provide the funding opportunities and consulting services that are offered on the platform and will help to expand the reach and impact of JamBank's services.

Team: Building and operating this platform requires two major areas of expertise, technical capabilities, and industry experience. Inhouse-development of the platform allows designing the solution according to specific needs that may arise during its conception process. On the other hand, team members on the business side proved long-term experience in the industry and add their valuable network to the venture. Having worked on both sides, fund-seeking and application assistance provides them with a skill set that is well-tailored to tackle the challenges of this venture. This combination ensures that JamBank provides a comprehensive and secure solution to the market. The team also works closely with industry partners to ensure that the platform and database remain up to date with the latest funding opportunities and grant applications.

The platform's activities will involve the development and operation of the platform, the maintenance and updating of the database of funding options, and the connection of startups with consulting companies. This necessitates the establishment and maintenance of the technical infrastructure of the platform, the aggregation and structuring of information on funding opportunities, and the facilitation of connections between startups and consulting companies.

5.3 Customer acquisition and monetization

JamBank plans to generate revenue through two main channels: a premium version of the platform for organizations and advertisements from consulting companies.

Premium Version for Organizations: Impact Startups and NGOs seeking to maximize their success in grant applications can take advantage of the premium version of the software, which offers a range of analytical tools and proposal drafting features that increase the chances of success in grant applications. This may include analytics, proposal drafting tools, and other specialized features. A monthly fee is required to access these features.

Advertisement from Consulting Companies: The platform includes an ads component that will enable consulting companies to generate leads and augment their visibility through an auction-based system. This system will allow consulting companies to bid for ad spaces, with the most substantial bid being awarded the most prominent placement.

In addition, European funds will be utilized in the initial months to finance the platform's activities as an example of the viability of public grants for the funding of a startup.

Key Costs: JamBank's key costs include platform development and maintenance, marketing and sales efforts, and any fees associated with partnering with public institutions and consulting companies. The costs of developing and maintaining the platform will be a significant expense,

as will the marketing and sales efforts needed to attract and retain customers. Any fees associated with partnering with public institutions and consulting companies will also need to be considered in the overall business model.

5.4 Future overlook and trend analysis

The primary research area of this platform will be the Lisbon region and Portugal as a whole, and it will feature startups and consulting firms in the area. The platform will provide information about grants from the European Union, tax credits and financing options from the Portuguese government.

In the medium-term, the platform will be expanded to encompass the entire application process, featuring roadmap and budget drafting, a comprehensive calendar, and a listing of activities that will be conducted once the funding is procured. Additionally, the platform will be extended horizontally to incorporate European markets, providing funding prospects from national governments, social impact funds, and impact venture capital funds. JamBank will expand its reach to social impact funds and venture capital funds dedicated to impact investing. In this way, the goal of create a more equitable access to impact and public funds will be finally reached.

The platform has a long-term goal of achieving global reach and providing an automated process for grant applications, from A to Z. Additionally, offices will be established globally to offer support to startups and NGOs. The ultimate goal is to democratize impact and public funds and make them accessible to organizations from all around the world.

6 The Role of Technology

The technical functionality of the project will be emphasized in this section to showcase its operability and provide an analysis of the results. As technology is a key driver for this project's competitive advantage and differentiation, a functioning minimal viable product was designed with a preliminary database to provide insight into the structure of the database and the web scraping process. Additionally, a matching algorithm was implemented to demonstrate the model's ability to allocate projects to entities based on an automated similarity index calculation.

6.1 The Role of Data on the Platform

The core technical function of the platform is to reduce the manual work involved in scrolling through potential sources of funding by applying web scraping and harmonizing data in a coherent database. This database can then be filtered and tailored to the specific type of grant that the organization is searching for. In this project's scope, a functioning prototype of the database is provided to showcase core functionalities. This minimal viable product connects to 5 major providers of publicly accessible grants in entire Europe and specific countries. On the project side, we were able to obtain data on more than 900 projects from 5 different data sources that list current projects for funding. The data sources include national, as well as international institutions that offer funds for completing projects with significant social impact. The largest source connected is the database from the European Commission to which the algorithm can connect automatically, complemented by minor national and private institutions that offer to fund in specific areas. This data was obtained using URL requests and server-side APIs to retrieve data automatically from the website. On the entity side, 425 entities were captured from 2 different

Project tags counted	
count	902.000000
mean	162.046563
std	118.171596
min	0.000000
25%	93.000000
50%	149.000000
75%	205.000000
max	1185.000000
Entity tags counted	
count	425.000000
mean	19.472941
std	13.902112
min	1.000000
25%	10.000000
50%	16.000000
75%	25.000000
max	107.000000

data sources. Most entries come from NGO explorer, a multinational database that collects information on NGOs, complemented by Startups with a focus on impact captured by dealflow.eu a major database for information regarding the start-up ecosystem. The data was gathered by manual downloads directly from the website. Both sources ensure data consistency and data quality through continuous supervision and updates in the databases. (Table 1: Implemented Data Sources).

6.2 Data Scraping Process

In this research, data were obtained from five different sources on December 05, 2022 (Table 1: Implemented Data Sources). There are two ways to retrieve data from web pages: using a web application programming interface (API) or a web scraping library in a data processing language, such as BeautifulSoup in Python. A web API is a web development technology that facilitates the transfer of data, often on the client side. By using Python to make requests with a specified data type, a JSON object containing structured and labeled data can be obtained. In this case, the European Commission implemented an API to facilitate the transmission of available projects. Furthermore, the technique of transferring Hypertext Markup Language (HTML) data into data processing languages such as Python is known as web scraping. Python has several libraries for running web scraping, and for this project, the BeautifulSoup library is used. Through the linked elements in HTML, it is possible to access and retrieve information stored on a website. This practice allows for efficient and structured data retrieval from web pages. Web scraping is particularly effective when the information is systematically stored, such as in tables and lists. The structure and quality of data are important because it directly impacts the accuracy and reliability of insights and conclusions derived from that data. When data is well-structured and high quality, it is easier to analyze and interpret, leading to more accurate and meaningful results. The data processed in the database is either manually mined through web scraping or pre-structured API downloads. This process ensures that data is well-

structured and of high quality. This approach informs and supports decision-making more robustly and reliably, helping the company to make better decisions and achieve goals effectively.

6.3 Data Engineering with Keywords

The platform's core function is to reduce the resources required for an organization to locate a suitable grant for which it can apply. Consequently, a crucial element in matching both funding seekers with active funding projects is to devise a method for calculating the similarity between the two sides. In practice, this involves calculating the similarity between one organization and n-projects. By determining these metrics, it becomes possible to rank projects numerically according to their similarity index.

The data obtained from both organizations and active projects suggests that common ground between the two sides will be based on the text in descriptions. These descriptions are in natural language, i.e. written in full sentences. Therefore, the field of data processing and machine learning that is relevant to this scenario is natural language processing (NLP). NLP is a subfield of computer science and artificial intelligence concerned with the interaction between computers and human languages. It focuses on developing algorithms and models that can understand, interpret, and generate human language, allowing computers to process and analyze large amounts of text or speech data.

There are two ways of calculating the similarity in a text that differ in the input data used by the models to classify similarities in natural language. These methods are either using the full text or using subsets of the text, i.e., keywords, as input to the model to determine similarity. One advantage of using keywords to compare the similarity of texts is that it can be computationally more efficient than using the full text. Keywords are typically a small subset of the words in a text, and so comparing the similarity of texts based on keywords can require

less computation than comparing the full text. This can be particularly useful for large or complex texts, where the computational cost of comparing the full text may be prohibitively high. Additionally, keywords are typically the most important or representative words in a text, so comparing the similarity of texts based on keywords can capture the most important information in those texts (Justeson, John S., and Slava M. Katz, 1995).

6.4 Keyword extraction with RAKE_NLTK

The Rapid Automatic Keyword Extraction (RAKE) algorithm, implemented in the `rake_nltk` library, is a technique that utilizes natural language processing (NLP) and statistical methods to automatically identify and extract the most important keywords or phrases from a given text.

By taking into account the frequency and relevance of the keywords to the overall content, the `rake_nltk` algorithm can accurately identify the most important keywords in the text. Upon identification, these keywords are extracted and returned as a list. Overall, RAKE with `rake_nltk` is a powerful and efficient tool for identifying and extracting the most essential keywords from a text (Rose, Stuart, Dave Engel, Nick Cramer, and Wendy Cowley, 2010). To improve the semantic relevance of the keywords, the model was refined to only include singular and plural nouns. The

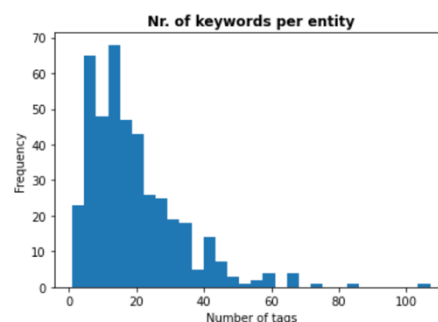


Figure 11: Distribution of tags on entities

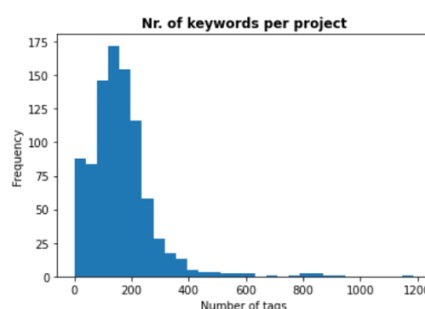


Figure 12: Distribution of tags on projects

NLTK library provides an efficient method for achieving this by tokenizing each word and adding a categorical tag. This allows for part-of-speech tagging, which assigns a tag to each tokenized keyword indicating whether it is a noun, verb, or adjective. An example of this can be seen in the Penn Treebank Project, which aims to develop annotated corpora for use in natural language processing research. Reducing the keywords to nouns has several advantages,

including improved relevance and specificity, as well as reduced noise (Marcinkiewicz, Mary Ann, 1995) The refinement of the model to only include singular and plural nouns has several advantages. By reducing the keywords to nouns, the model can improve relevance, increase specificity, and reduce noise. Additionally, this allows the model to increase in accuracy and relevance while requiring less computational power. This demonstrates the effectiveness of limiting the keywords to nouns in terms of both performance and efficiency. Even though entities have significantly fewer keywords compared to projects, we can logically infer that this is a result of the fact that project descriptions typically contain more text than the descriptions of companies and NGOs, which tend to be more concise. Furthermore, this is beneficial for the matching process, as the more precise descriptions of entities allow for a more accurate allocation of projects.

6.5 Determine the similarity between keywords

When calculating text similarity based on keywords, several factors must be taken into consideration. Firstly, there is the selection of keywords: the first step in calculating text similarity based on keywords is to identify which words will be used for comparison. This requires choosing a subset of words from the texts that are most relevant and important for the task at hand. In this case, the RAKE_NLTK algorithm was chosen for its ability to provide relevant and specific input data for calculating similarity. Secondly, there is a method for calculating similarity: once the keywords have been selected, the next step is to choose a method for determining the similarity between the texts based on these keywords. In this case, the platform's matching model is based on the GenSim algorithm, a free open-source Python library for representing documents as semantic vectors. Gensim is designed to process raw, unstructured digital texts by applying unsupervised machine-learning algorithms (Gensim 2022). The model has many strengths, including its ability to accurately represent texts using word vectors, which capture the meaning and context of the text. Additionally, the model is

highly flexible, thanks to its ability to be extensively trained, and is computationally efficient (Rehurek Radim, and Petr Sojka, 2010). Another crucial factor to consider when calculating text similarity based on keywords is the context in which the keywords appear. The meaning and significance of a keyword can vary depending on the context in which it appears, so it is essential to take this into account when determining text similarity (Wang, Rui, Wei Liu, and Chris McDonald, 2014).

Ultimately, the algorithm will return a similarity score between 0 and 1 that directly measures the similarity between two keywords. To gain a comprehensive understanding of the similarity of projects, it is necessary to compare each keyword of an entity to all keywords of each project. For example, if an entity has 5 keywords and a project has 6 keywords, the algorithm will require 30 calculations to determine the similarity of each pair. In the next step, the scores are aggregated so that each pairing of an entity with a project receives a unique score. This is accomplished by calculating the mean of relevant keywords, where relevant keywords are filtered by a minimum similarity threshold.

6.6 Analysis of Results

A sample of 30 entities and 250 projects was collected and analyzed to assess the effectiveness of the matching process. The algorithm matches each entity with all projects, resulting in 7500 unique pairings. For each of these pairings, the average similarity index was calculated. As previously discussed, entities have far fewer keywords per entity, with a mean of 19.4, compared to the 162 average keywords per project. Furthermore, all keywords of all entities and projects were paired, resulting in a total of 4,681,261 keyword pairs. The similarity of each of these pairs was determined, with a mean similarity index of 0.2199. This indicates that the model returns an overall similarity of keywords of 22%. This is reasonable,

Similarity	
count	7401.000000
mean	0.219898
std	0.010681
min	0.150670
25%	0.214700
50%	0.219450
75%	0.224600
max	0.404110

Figure 13: Statistics on keyword distribution

given that the input data of entities and projects come from similar fields related to impact entrepreneurship, social and ecological welfare, and funding projects with a focus on achieving positive social impact. Additionally, the 25% and 75% quartiles show a dense distribution around the

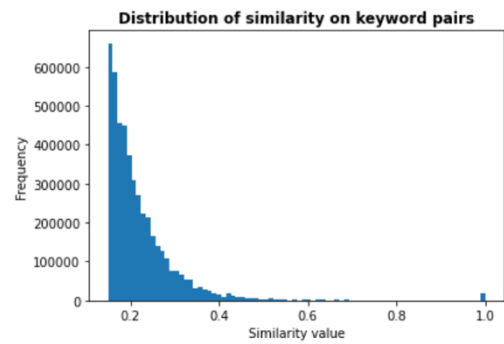


Figure 14: Distribution of similarity on keyword pairs

mean and median, indicating that the average similarity of projects is significantly above the mean, and even the 75% quartile of pairings possess characteristics that reflect this similarity. Further exploration will involve determining a threshold for classifying a match. Additionally, by conducting sanity checks on the column of ‘Highest Matched Keywords’ in the resulting data frame, it can be demonstrated that identical keywords are matched correctly, with pairs of (“health”:” health”) and (“technology”: “technology”) obtaining a score of 1, indicating 100% similarity.

It is important to consider the limitations and potential biases of this project. First, the input data is a vulnerable factor that can cause errors in the matching process. This may be due to inconsistencies in the descriptions of projects or entities, as well as the engineered keywords from the text. Additionally, it should be noted that the number of keywords is not equally distributed. Outliers with significantly more or fewer keywords than the average can distort the results, as significantly fewer keywords can be matched more easily and receive a higher score. Conversely, long descriptions result in more keywords, which usually lowers the overall similarity index in this calculation logic. Secondly, the GenSim model, a large-scale semantic NLP model, requires a significant number of computational resources to process large amounts of text data. This can make it difficult to apply the algorithm to very large datasets or to use it in real-time applications where speed is important. Additionally, the GenSim algorithm is based on unsupervised machine learning, which means that it may not be able to incorporate domain-

specific knowledge or prior assumptions about the data. This can make it less effective at generating high-quality results in certain contexts.

6.7 AI Development: The Roadmap

In this chapter, the value that technology can bring to the business and its stakeholders will be highlighted. The development and implementation of the AI project will be outlined in a clear and actionable roadmap, which will help to clarify and articulate the value and benefits of the project to stakeholders and partners. By including an AI Development Roadmap, the platform can ensure that it reaches its full potential and is implemented in a timely, efficient, and effective manner. This will help to ensure that the expected outcomes, impacts, and benefits are delivered.

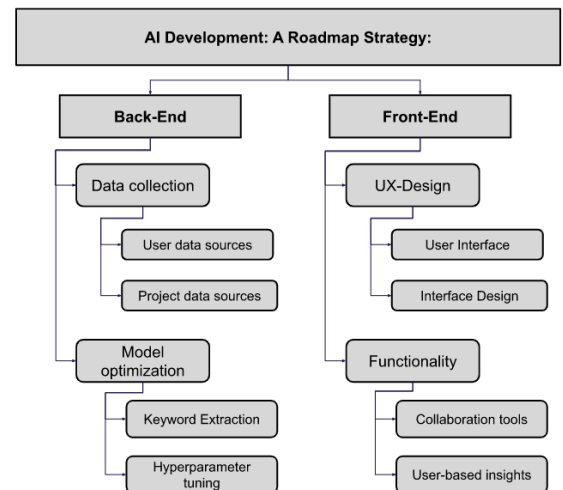


Figure 15: AI Development Roadmap

Advances in data collection: A key strategic asset of the platform is the availability of data. As with any data-driven application, the comprehensiveness of the database is crucial to the success of the venture. Therefore, it is essential to have a strategic plan for expanding the database and determining which features should be considered in the matching process. To provide value to all stakeholders, it is necessary to develop a dual data collection strategy for both users and customers.

On the user side are organizations that leverage the platform to shorten the scouting process for funding. Although the platform's primary function is to connect signed-up users with existing projects in the database, it is possible to collect data and information on existing and potential customers through publicly available information online. Tools such as 'NGO Explorer'

provides insight into NGOs that are active in a certain country or region. Based on the description it is possible to engineer keywords to match organizations to particular projects. This is particularly powerful because this enables us not only to match existing users to matching projects but also to identify potential matches. On the business side, this is a significant leap forwards because users can be directly acquired with a specific value proposition (NGO Explorer 2022). Similarly, "dealroom.eu" is a comprehensive source for obtaining data on start-ups and companies. The demo dataset for Portugal also includes a column for tags, which can potentially be matched with the engineered keywords for each project (Dealflow 2022).

On the stakeholder side, the platform focuses on the scouting process rather than the actual application process, which requires more individual expertise. However, it can capture and leverage the information relevant to this task. With aligned interests, consultancies aim to provide value to their customers through effective and successful application services. The platform can eventually infer whether an application was successful through two sources of information. Firstly, a survey on customer satisfaction can be sent to the user once it is decided that the organization will apply for a grant. Secondly, if the organization subsequently attempts to find another partner for a grant application. Through a basic algorithm, the database can approximate the success rate of consultancies, as well as the amount of time the application process required. The visibility and ranks of recommended partners will also depend on their success rate to enhance user satisfaction. By following this approach, the platform will not only be more accurate in the matching process, but its continuous growth in stakeholders will also make it more sustainable.

Learning Tasks: This section presents an outlook and roadmap for the platform's potential growth, as well as the learning tasks that need to be completed to implement and refine its functionalities. Overall, the development of a powerful and efficient AI involves a combination

of research, development, and evaluation activities, as well as regular iteration and optimization to ensure that it meets its performance and accuracy goals. In this project, we present a minimal viable product that already contains the core functionalities. With this foundation in place, we will continue to explore opportunities for evaluation and optimization to enhance the performance of the model.

Reliable evaluation metrics are crucial in obtaining a realistic perspective on the model's performance, thus key performance indicators (KPIs) must mirror the objectives of the platform. For this reason, KPIs were defined to cover the core strategic aspects of the platform that is user interaction, success in the matching process, and operational efficiency (Table 2: Strategic KPIs).

Performance optimization is essential for sustaining competitiveness through improved accuracy in the matching process, as well as for reducing resource usage and increasing computational efficiency in order to save costs. Additionally, developing a differentiation method can help the platform to prevail in the market. There are various ways to boost the model's performance, such as experimenting with and comparing different NLP models for project matching, such as the Cosine and Jaccard similarity algorithms. Another approach to increase accuracy would be to improve the quality of the input data by refining the keyword-selector model and making manual adjustments for keywords that do not add value because they are not specific to a particular topic.

Another crucial aspect is the user experience and user interface. Advancing the platform will always require improvements in the accessibility and usability of the digital platform, as users can only retrieve the value that they are able to generate through it. Therefore, it is essential to make the platform easy to use with a user-centric interface. Within this scope, further functionalities can be implemented to enhance the user experience, such as an estimation of the

duration of an application and the likelihood of admission, as well as collaboration tools such as discussion forums, project management tools, and shared calendars. Additionally, analytics and insights based on user data can provide trends and patterns in the funding landscape. Through a thorough, strategic approach that applies a user-centric technology development approach, the platform has the potential to deliver its value proposition to the market and assist entities in finding relevant funding opportunities. On the stakeholder side, the algorithm is also capable of determining the similarity between consultancies and particular projects.

Scalability: Scalability is important because organizations and funding projects can vary widely in size and complexity, and a scalable platform can accommodate this diversity and support a wide range of users and applications. Scalability can also be vital from a business perspective, as it allows the platform to grow and expand as more organizations and funding providers join, and as the platform's user base and data volume increase. This can help the platform to generate more value and revenue, and to support the growth and success of its users. Ultimately, scalability is a strength of a digital platform that calculates the similarity between organizations and funding opportunities through keywords, as it enables the platform to handle a wide range of data and users without sacrificing performance or accuracy and to support the growth and success of its users.

6.8 Interpretation and business case

To conclude, technology plays a central role in the competitive advantage that the platform can achieve through extensive data collection, a user-centric value proposition, and high scalability. It is an essential asset of the project that complements its business side. Through a comprehensive data strategy, the company can collect, store, and analyze data in a way that is efficient, effective, and scalable. This enables it to use its data to train and evaluate machine learning models, as well as to make data-driven decisions based on accurate and reliable

insights. The platform extracts value from its data by identifying trends, patterns, and insights that can inform and support decision-making, and help the company to improve its operations, develop new products and services, and gain a competitive advantage in the market. Overall, through the combination of a powerful technology strategy and a target-oriented business strategy, the platform will be successful in its mission to facilitate the scouting process for entities to acquire funding.

7 Conclusion

In conclusion, impact start-ups and NGOs are emerging as significant players in the financial ecosystem. Impact investing seeks to generate both positive social and environmental impact and financial returns, while impact start-ups aim to achieve both financial and social or environmental returns. By aligning private financial flows with the common good, impact investing and impact start-ups have the potential to promote sustainable development and drive positive change in our communities and the natural world. However, the funding environment for impact start-ups in Europe is challenging due to a lack of promotion and dispersed information about government funding opportunities, as well as a lengthy and complex application process for grants. The strict reporting rules and lack of a common platform for accessing funding opportunities can make it difficult for start-ups to find the best options available to them. These problems can lead to inefficiencies in the distribution and management of grants and government subsidies for impact start-ups. In response to these challenges, this project has created a platform that connects these organizations with the best funding options from public institutions, such as grants and loans. The platform uses a sophisticated matching approach to determine the similarity between an organization and available projects, allowing both sides to benefit. With a well-aligned technology and business strategy, the service can provide value to its users through technology and revenue generation, as well as financial

sustainability through its business strategy. Moreover, it should be noted that the current state of the project has some limitations. In particular, errors in the matching process can be caused by inconsistencies in engineered keywords, and the number of keywords is not equally distributed for entities and projects, which can distort the accuracy of the model. Additionally, the algorithm is unsupervised and may not be able to incorporate domain-specific knowledge or prior assumptions about the data, potentially reducing its effectiveness in the matching. This project presented a comprehensive and viable solution with a fully functional algorithm to minimize the resources needed for entities seeking the appropriate source of funding. Operating in the rapidly expanding realm of sustainable investments and promoting the notion of combining business and purpose, the platform offers a valuable contribution to the allocation of funds for meaningful ventures in a more effective manner, laying the groundwork for a better world.

8 Appendix

8.1 References:

Literature:

- Brousseau, E. and Penard, T. The economics of digital business models: A framework for analyzing the economics of platforms. *Review of Network Economics*, 6(2). (2007).
- Christen, Peter. "The data matching process." In *Data matching*, pp. 23-35. Springer, Berlin, Heidelberg, (2012).
- Cusumano, Michael, David Yoffie, and Annabelle Gawer. *The future of platforms*. MIT Sloan Management Review, (2020).
- Eisenmann, Thomas, Geoffrey Parker, and Marshall W. Van Alstyne. "Strategies for two-sided markets." *Harvard business review* 84, no. 10 (2006): 92.
- Evans, David S., and Richard Schmalensee. *Matchmakers: The new economics of multisided platforms*. Harvard Business Review Press, (2016).
- Hagiu, Andrei. "Pricing and commitment by two-sided platforms." *The RAND Journal of Economics* 37, no. 3 (2006): 720-737.
- Justeson, John S., and Slava M. Katz. "Technical terminology: some linguistic properties and an algorithm for identification in text." *Natural language engineering* 1, no. 1 (1995): 9-27.
- Lamy, Y. S., C. Leijonhufvud, and N. O'Donohoe. "The Next 10 Years of Impact

Investment." Sandford social innovation review Mar 16 (2021).

- Marcinkiewicz, Mary Ann. "Building a large, annotated corpus of English: The Penn Treebank." *Using Large Corpora* 273 (1994).
- Rehurek, Radim, and Petr Sojka. "Software framework for topic modelling with large corpora." In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*. (2010).
- Rose, Stuart, Dave Engel, Nick Cramer, and Wendy Cowley. "Automatic keyword extraction from individual documents." *Text mining: applications and theory* (2010): 1-20.
- Wang, Rui, Wei Liu, and Chris McDonald. "Corpus-independent generic key phrase extraction using word embedding vectors." In *Software engineering research conference*, vol. 39, pp. 1-8. (2014).

Reports:

- Deloitte. "A quick guide to EU Funding" (2014).
- European Commission. "Beginners' Guide to EU Funding". (2012).
- EVPA. "The 2020 Investing For Impact Survey: A Deep-Dive Into Impact Funds". EVPA Publishing, (2020)
- Freireich, Jessica, Fulton, Katherine. Monitor Institute. "Investing for Social & Environmental Impact" (2009).
- Gaggiotti, G., and Gianoncelli, A. EVPA, "Accelerating Impact - Main takeaways from the first harmonised European impact investment market sizing exercise". (2022).
- Hagel, John. "The power of platforms." Part of the Business Trends series, Deloitte

Insights (2015).

- Hand Dean, Ringel Ben, Danel Alexander. Global Impact Investing Network. “GIINsight: Sizing the Impact Investing Market 2022”. (2022)
- International Finance Corporation. “Investing for Impact: The Global Impact Investing Market 2020”. World Bank, (2021).
- Krausova, Tereza, Walsh, John. “2021-2027: Cohesion policy EU budget allocations”. (2021).
- Mackevičiūtė, Raimonda, Žilvinas Martinaitis, Fiorenza Lipparini, Barbara Constance Scheck, and Izabela STYCZYŃSKA. "Social Impact Investment-Best Practices and Recommendations for the Next Generation." (2020).
- OECD. Social Impact Investment 2019 The Impact Imperative for Sustainable Development. OECD Publishing, (2019).
- Patricia Roger. UNICEF Office Research. “Theory of Change, Methodological Briefs: Impact Evaluation 2” (2014).
- Similarweb. “November 2022 Website Analysis: Up2Europe”. (2022).
- Similarweb. “November 2022 Website Analysis: Dealflow”. (2022).

Web Resources:

- Agency Up2Europe. “Up2Europe.com - Partner Platform.” Accessed December 14, 2022. <https://up2europe.com/services/matchmaking-platforms/partners>. (2022)
- BIG Society Capital. 2022. "What is an Impact Startup?" Accessed November 15, 2022. <https://bigsocietycapital.com/latest/what-is-an-impact-startup/> (2022).
- Dealflow.EU. “Dealflow - Europe’s Best Investment Opportunities.” Accessed December 12, 2022. <https://dealflow.eu/> (2022).

- European Commission. 2022. “Social Economy and inclusive entrepreneurship”. Accessed November 17, (2022).
<https://ec.europa.eu/social/main.jsp?catId=1561&langId=en> (2022)
- European Investment Bank. 2022. “EFSI and its legacy”. Accessed December 01, 2022. <https://www.eib.org/en/products/mandates-partnerships/efsi/index.htm> (2022)
- European Investment Fund. 2022. “The EIF and InvestEU”. Accessed December 10, 2022. <https://www.eif.org/InvestEU/index.htm> (2022).
- Good Finance. 2022. “Charity Bonds”. Accessed December 14, 2022.
<https://www.goodfinance.org.uk/charity-bonds> (2022).
- Impact Investing Hub 2022. “Understand Impact Investing”. Accessed December 02, 2022. <https://www.impactinvestinghub.com.au/> (2022)
- Impact Startup Nordic. 2022. "What is Impact?" Accessed November 16, 2022.
<https://www.impactstartupnordic.com/impact> (2022).
- Mar Région Sud. Up2Europe. “Mar Région Sud: Accélérateur d’Idées Pour La Cooperation Européen”. Accessed November 18, 2022.
<https://maregionsud.up2europe.eu/> (2020)
- NGO Explorer. “Ngoexplorer.org”. Accessed December 12, 2022. Accessed December 12, 2022 <https://ngoexplorer.org/about> (2022).
- Radim Řehůřek. “Gensim: Topic Modelling for Humans.” Accessed December 12, 2022. <https://radimrehurek.com/gensim/> (2022).
- Rees, Matt. 2017. “Quasi-equity: A new financial structure for a new challenge”. European Investment Bank. Accessed October 20, 22.
<https://www.eib.org/en/stories/quasi-equity-a-new-financial-structure-for-a-new-challenge> (2017)

- Morgan Stanley and Co LLC. "Sustainable Signals: Asset Owners Embrace Sustainability." Accessed December 10, 2022.
<https://www.morganstanley.com/assets/pdfs/sustainablesignals-asset-owners-2018-survey.pdf> (2018).
- Trager, Lily. 2018. "Why Investors Are Getting in on This Growing Trend | Morgan Stanley." Morgan Stanley. Accessed November 19, 2022.
<https://www.morganstanley.com/access/impact-investing-why-it-matters> (2018).

8.2 Figures

Figure 1: Social Impact Investment Taskforce. 2014. Impact Investments: The Invisible Heart of Markets. UK: G8

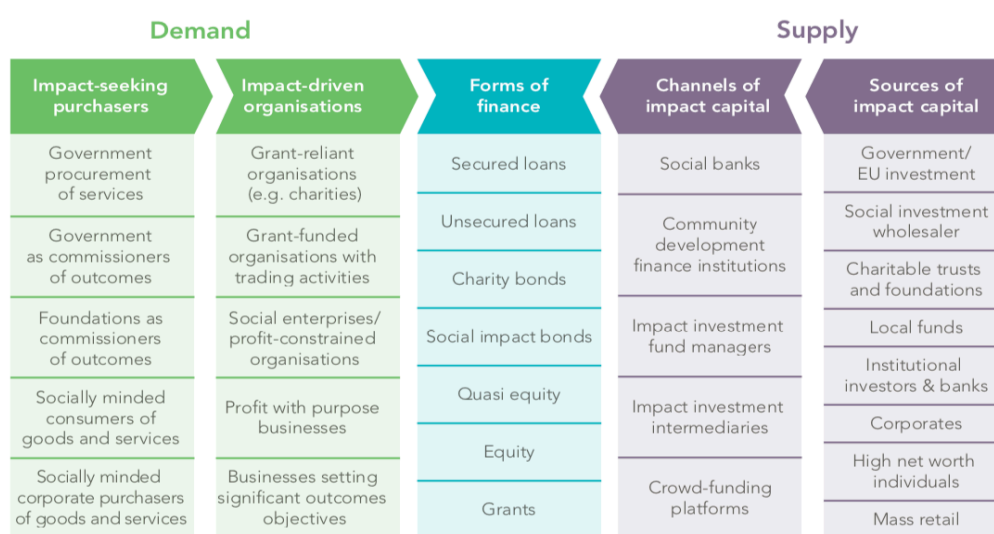


Figure 2: World Economic Forum. 2020. Bridging the gap in European Scale-up Funding: The Green Imperative in an Unprecedented Time. (Funding from European Union in start-up lifecycle)

Figure 3: European Commission. 2021. Fact Check on the EU Budget. Accessed December 3rd, 2022.

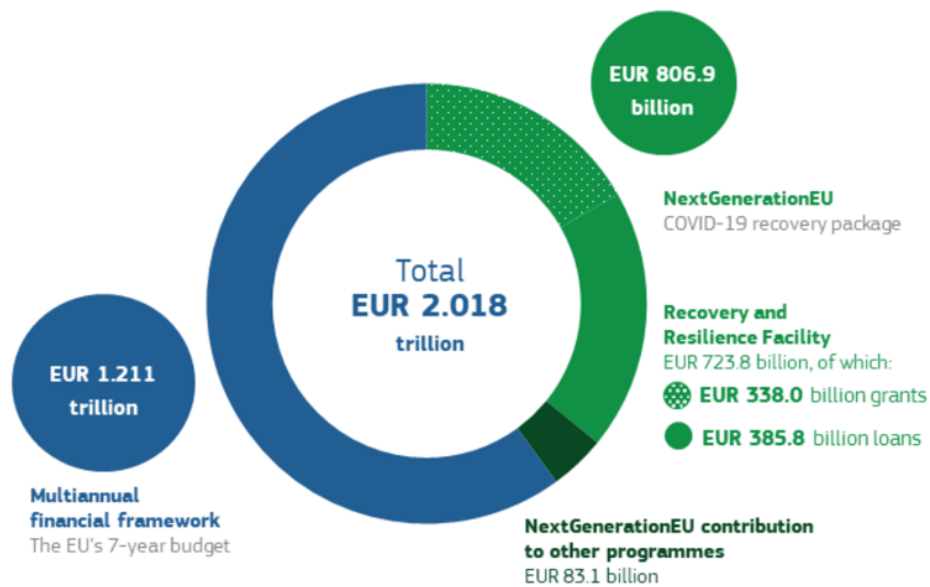


Figure 4: Value Proposition Canvas– Impact Startups customer segment

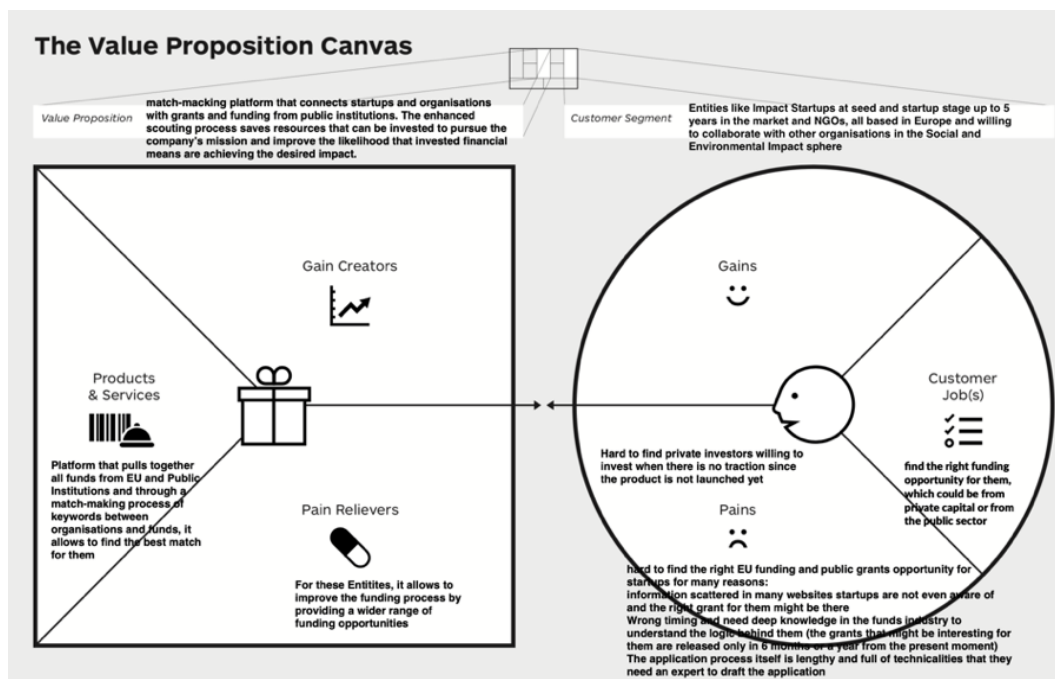


Figure 5: Value Proposition Canvas – Consulting company's customer segment

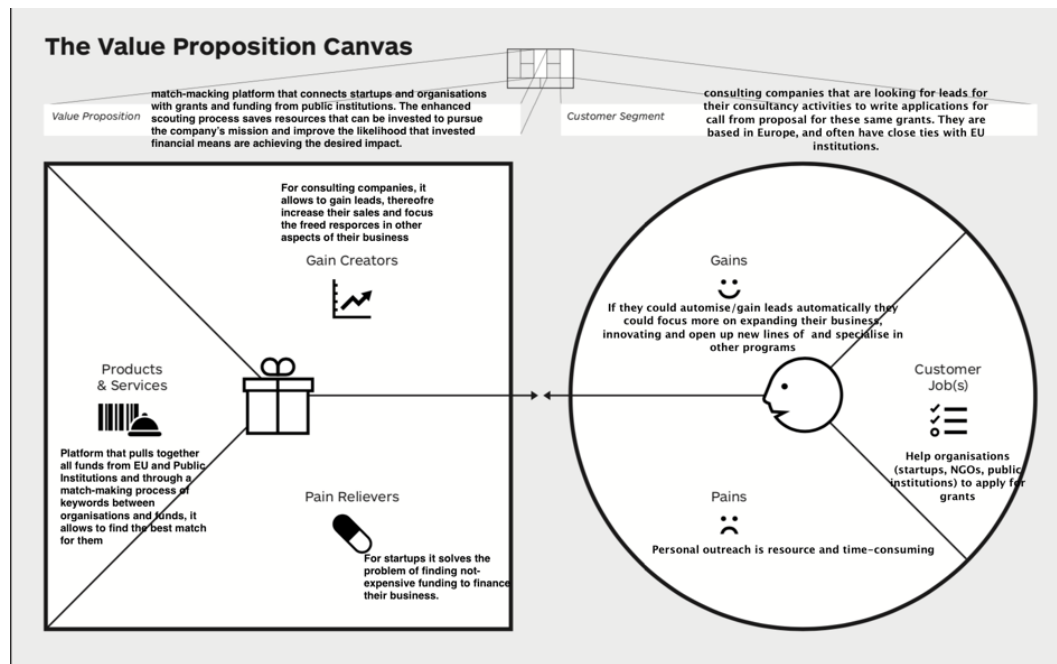


Figure 6: Statistics on captured projects

Project tags counted	
count	902.000000
mean	162.046563
std	118.171596
min	0.000000
25%	93.000000
50%	149.000000
75%	205.000000
max	1185.000000

Figure 7: Statistics on captured entities

Entity tags counted	
count	425.000000
mean	19.472941
std	13.902112
min	1.000000
25%	10.000000
50%	16.000000
75%	25.000000
max	107.000000

Figure 8: Distribution of tags on entities

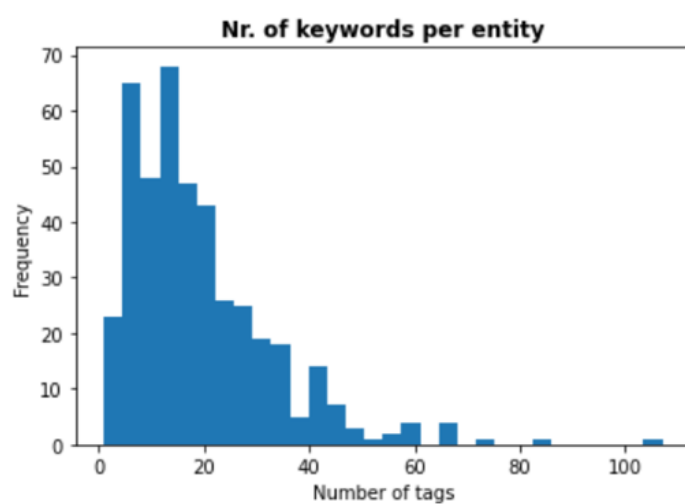


Figure 9: Distribution of tags on projects

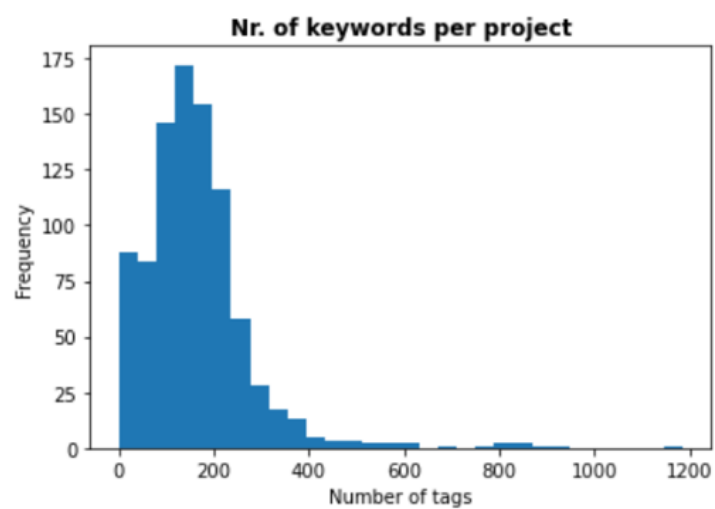


Figure 10: Statistics on keyword distribution

Similarity	
count	7401.000000
mean	0.219898
std	0.010681
min	0.150670
25%	0.214700
50%	0.219450
75%	0.224600
max	0.404110

Figure 11: Distribution of similarity on keyword pairs

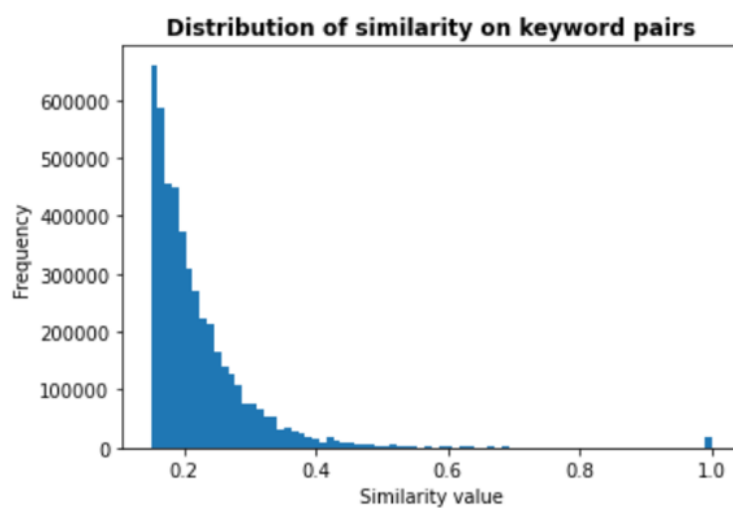
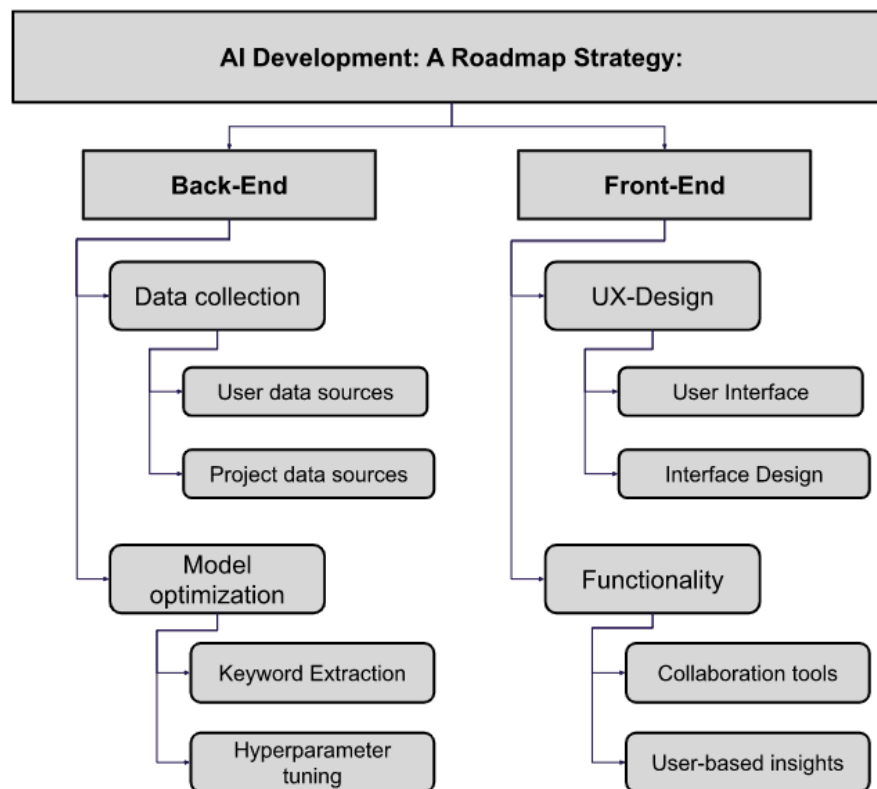


Figure 12: AI Development Roadmap



8.3 Tables:

Table 1: Implemented Data Sources

Funding entity	EEA	EIF	European Commission	Eureka Network	StartUp Portugal	Dealflow.eu	NGOexplorer.org
Description	Funded by Norway, Liechtenstein and Iceland the program provides funding for 15 countries within Europe across various programs	The EIF designs and provides equity and debt products for projects that benefit societal welfare	The Funding & Tenders Portal is the entry point for participants and experts in funding programs and tenders managed by the European Commission and other EU bodies	Eureka is an agreement between 18 countries and the European Commission to foster competitiveness to boost R&D cooperation. It includes 47 countries in and beyond Europe	StartUp Portugal is a non-profit organization aiming to foster entrepreneurship and innovation in Portugal with a particular focus on Portugal and entrepreneurs	A European portal that connects thousands of EU-funded innovations with Investors & Corporates	An explorer for data on development charities based in England and Wales
Data retrieved	Project data	Project data	Project data	Project data	Project data	Entity data	Entity data
Data retrieval technique	Web scraping with BeautifulSoup	Web scraping with BeautifulSoup	Access through web API	Web scraping with BeautifulSoup	Web scraping with BeautifulSoup	Download queries as CSV-file	Download queries as CSV-file

Table 2: Strategic KPIs

KPI	Number of active users	Number of search queries	Number of matches	Funding success rate	User satisfaction	Computation time
Definition	Number of users registered on the platform that conducted a query in the last 100 days	Total number of search queries performed by users	Total number of matches that were achieved through queries	The amount of successful application by the total amount of matches	Percentage of organizations that have successfully secured funding through the platform, as well as the average amount of funding that has been awarded	Time required for an organization to perform a query through all projects

8.4 Abbreviations:

Abbreviation	Definition
API	Application Programming Interface
CLV	Customer Lifetime Value
CSI	Corporate Social Investor
EIB	European Investment Bank
EIC	European Investment Council
EIF	European Investment Fund
EIT	European Institute of Innovation & Technology
ERDF	European Regional Development Fund
ESF+	European Social Fund Plus
EU	European Union
GIN	Global Investment Network
HTML	Hypertext Markup Language
IMF	International Monetary Fund
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
NGO	Non-Governmental Organization
NLTK	Natural Language Toolkit
RAKE	Rapid Automatic Keyword Extraction
ROI	Return on Investment
SAM	Serviceable Addressable Market
SDG	Sustainable Development Goal
SIB	Social Impact Bond
SME	Small-Medium Enterprises
SOM	Serviceable Obtainable Market
SVI	Social Value International
TAM	Total Addressable Market
UN	United Nations
WOM	Word-of-Mouth

8.5 Python Implementation

8.5.1 Web Scrapping

Master_Scraping_05122022

15.12.22, 10:57

Overcoming Financing Challenges for Impact: The Development of a Match-Making Platform for Impact Organisations and Public Funding Opportunities

Obtaining data through web scrapping

For this project, the following code connects and systematically downloads data from websites. This notebook contains several functions to gather and clean the data from different sources.

Eventually all gathered data tables will be joined in one harmonized database. The database will then be saved in a CSV-format to enable further processing.

Imports

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 1 of 17

Master_Scraping_05122022

15.12.22, 10:57

```
In [1]: #Library imports
import bs4
import numpy as np
import pandas as pd
import requests
from urllib.request import urlopen as uReq
from bs4 import BeautifulSoup
from datetime import date
import csv
import warnings

#Display options
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
warnings.filterwarnings("ignore")

In [2]: pip install pyopenssl ndg-httpclient pyasn1 #Avoid connection timeouts with EU website

Requirement already satisfied: pyopenssl in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (22.0.0)
Requirement already satisfied: ndg-httpclient in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (0.5.1)
Requirement already satisfied: pyasn1 in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (0.4.8)
Requirement already satisfied: cryptography>=35.0 in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (from pyopenssl) (36.0.0)
Requirement already satisfied: cffi>=1.12 in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (from cryptography>=35.0->pyopenssl) (1.15.0)
Requirement already satisfied: pycparser in /Users/moritzguttersberger/opt/anaconda3/envs/Energy/lib/python3.9/site-packages (from cffi>=1.12->cryptography>=35.0->pyopenssl) (2.21)
Note: you may need to restart the kernel to use updated packages.
```

Webscraping functions

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 2 of 17

EEA Website:

Funded by Norway, Liechtenstein and Iceland the program provides funding for 15 countries within Europe across various programs

```
In [3]: #PageScroller for EEA
def page_scroller_eea(og_link):
    pages = []
    for i in np.arange(5):
        site=og_link+str(i)
        pages.append(site)
    return pages

#Scraper for EEA website
def EEA_scraper(urls):

    cols = ['Title', 'Description', 'URL', 'Country', 'Budget', 'Deadline']
    db = pd.DataFrame(data = None, columns = cols)

    #Scroll thorough pages
    for p in urls:

        #Open URL
        page = requests.get(p)

        #Open Page with BS
        soup = BeautifulSoup(page.content, "html.parser")
        test = soup.find_all(class_='node node--type-calls node--view-mode-teaser ds-lcol clearfix')

        #Create lists to save vals
        title, description, url, country, budget, deadline = [], [], [], [], [], []

        for grant in test:
            #Titles
            title_element = grant.find("a").find('h2').text.strip().split(': ')

            description_element = title_element[1]
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 3 of 17

```
title_element = title_element[0]
title_element = title_element[3:]

title.append(title_element)
description.append(description_element)

#URLs
for a in grant.find_all('a', href=True):
    url_full = 'https://eeagrants.org/' + a['href']
    url.append(url_full)

#Country
country_element = grant.find('div', class_='meta').text.strip()
country_element = country_element[:-4]
country.append(country_element)

#Content with deadline and budget
budget_element = grant.find('div', class_='content').text.strip().replace('\n', '').split('|')

deadline_element = budget_element[0].replace(' ', '').split(':')[1].replace(' ', '')
deadline.append(deadline_element)

budget_element = budget_element[1].replace(' Budget: ', '')
budget.append(budget_element)

#Build DB
db_add = pd.DataFrame({'Title':title, 'Description':description, 'URL':url,
                       'Country':country, 'Budget':budget, 'Deadline':deadline, 'Source':'EEA',
                       'Retrieval Date':date.today(), 'Application Process':None, 'Program':None,
                       'Status':'Open'})

db = pd.concat([db, db_add], axis = 0, ignore_index = True)

return db
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 4 of 17

EIF Website:

The EIF designs and provides equity and debt products for projects that benefit societal welfare.

```
In [4]: def EIF_scraper(urls):

    cols = ['Title']
    db = pd.DataFrame(data = None, columns = cols)

    #Scroll thorough pages
    for p in urls:

        #Open URL
        page = requests.get(p)

        #Open Page with BS
        soup = BeautifulSoup(page.content, "html.parser")

        title, description = [], []

        #Title
        for a in soup.find_all('h1'):
            title.append(a.text.strip())

        #Description_full
        for a in soup.find_all('div', class_ = 'searchable'):
            description.append(a.text.strip())

        #Build DB
        db_add = pd.DataFrame({'Title': title, 'Description':description})
        db = pd.concat([db, db_add], axis = 0, ignore_index = True)

    return db

def EIF_details(db_full):
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 5 of 17

```
text = db_full['Description']
col_names = ['Title', 'Description', 'Application Process', 'Deadline', 'Description Accurate']
app, deadline, description_acc = [], [], []

for a in text:
    #Get application process
    app_1 = a.split('\nApplication process')[1].split('\nDeadline')[0].replace('\n','')
    #Get Deadlines for each applications
    app_2 = a.split('\nDeadline for applications: ')[1].split('\n')[0]
    #Get accurate description
    app_3 = a.split('Eligible financial intermediaries\n')[0].replace('\n','')
    #Append features to list
    app.append(app_1)
    deadline.append(app_2)
    description_acc.append(app_3)

#Complete Database
db_add = pd.DataFrame({'Application Process': app, 'Deadline':deadline, 'Description Accurate':description_acc})
db_full = pd.concat([db_full, db_add], axis = 1, ignore_index = True)
db_full.columns = col_names

#Remove full HTML text from df
db_full.pop('Description')
#Rename Column
db_full.rename(columns={"Description Accurate": "Description"}, inplace = True)

#Add Source
db_full['Source'] = 'EIF'
db_full['Retrieval Date'] = date.today()
db_full['Country'] = 'All EU'
db_full['Budget'] = None
db_full['Application Process'] = None
db_full['Program'] = None
db_full['Status'] = 'Open'

return db_full
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 6 of 17

Eureka Website:

Eureka is an agreement between 18 countries and the European Commission to foster competitiveness to boost R&D cooperation. It includes 47 countries in and beyond Europe.

```
In [5]: def Eureka_page_scroller(og_link): #Page scroller
        pages = []
        for i in np.arange(1,2):
            site=og_link+str(i)+'&country=all'
            pages.append(site)
        return pages

        def Eureka_scraper(urls): #Landing Page Scraper

            cols = ['Program', 'Title', 'Description', 'URL', 'Deadline']
            db = pd.DataFrame(data = None, columns = cols)

            #Scroll thorough pages
            for p in urls:

                #Open URL
                page = requests.get(p)

                #Open Page with BS
                soup = BeautifulSoup(page.content, "html.parser")

                url, title, deadline, description, program = [], [], [], [], []

                #URLs
                for a in soup.find_all('a', href=True, class_ = 'call-list-macro__call-link'):
                    url_full = 'https://www.eurekanetwork.org' + a['href']
                    url.append(url_full)

                #Titles
                for a in soup.find_all('h3', class_ = 'call-list-macro__call-title'):
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 7 of 17

```
        a=a.text.strip()
        title.append(a)

        #Deadline
        for a in soup.find_all('p', class_ = 'call-list-macro__call-end-date'):
            a=a.text.strip().split('\n')[1]
            deadline.append(a)

        #Description
        for a in soup.find_all('div', class_ = 'call-list-macro__call-description'):
            a = a.text.strip()
            if a not in description: #Ensure doubling is excludud
                description.append(a)

        #Program
        for a in soup.find_all('ul', class_ = 'tags-cloud'):
            a = a.text.strip()
            program.append(a)

        #Build DB
        db_add = pd.DataFrame({'Program':program, 'Title': title, 'Description':description, 'URL':url,
                              'Deadline':deadline})
        #db_add = pd.DataFrame({'Description':description})

        db = pd.concat([db, db_add], axis = 0, ignore_index = True)

        return db_add

    def Eureka_scraper_details(urls): #Details pages Scraper

        pages = urls
        db = pd.DataFrame(data = None)

        #Scroll thorough pages
        for p in pages:

            #Open URL
            page = requests.get(p)
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 8 of 17

```

#Open Page with BS
soup = BeautifulSoup(page.content, "html.parser")

description, countries = [], []

#Get Description
for a in soup.find('div', class_ = 'wysiwyg'):
    description.append(a.text)

description = description[9:] #Remove /n from list

del description[1::2] # delete every second element, starting from the second
description = ' '.join(description)

#Get Countries
c_list = []
for a in soup.find_all('li', class_ = 'call_countries-item'):
    c_list.append(a.text.strip())
countries.append(c_list)

#Build DB
db_add = pd.DataFrame({'Full Description':description, 'Country':countries, 'Source':'Eureka',
                      'Retrieval Date': date.today(), 'Status': 'Open'})
db = pd.concat([db, db_add], axis = 0, ignore_index = True)

return db

```

European Comission Website:

The Funding & Tenders Portal is the entry point for participants and experts in funding programs and tenders managed by the European Commission and other EU bodies.

```

In [6]: def eu_commission_projects(og_url):
        #Connect with API of EU website to retrieve all projects that are labeled as open grants

```

```

#Get API Response
projects = requests.get(og_url).json()

att = ['type', 'ccm2Id', 'identifier', 'title', 'status']

length = np.arange(len(projects['fundingData']['GrantTenderObj']))
tp, ccm2Id, identifier, title, status = [], [], [], [], []

for i in length:
    tp.append(projects['fundingData']['GrantTenderObj'][i]['type'])
    ccm2Id.append(projects['fundingData']['GrantTenderObj'][i]['ccm2Id'])
    identifier.append(projects['fundingData']['GrantTenderObj'][i]['identifier'])
    title.append(projects['fundingData']['GrantTenderObj'][i]['title'])
    status.append(projects['fundingData']['GrantTenderObj'][i]['status'])

project_db = pd.DataFrame(data = {'Type: Grant=1 / Call=0': tp, 'ccm2Id':ccm2Id, 'TopicIdentifier':identifier,

#Include status in DB
url_db = project_db['TopicIdentifier'].str.lower()

#Create URLS
u=[]

for i in url_db:
    url = 'https://ec.europa.eu/info/funding-tenders/opportunities/data/topicDetails/' + str(i) + '.json?lang=
    u.append(url)

z = pd.DataFrame({'URL':u})

url_db = pd.concat([project_db, z], axis = 1)

#Exclude all closed projects
url_db = url_db[url_db['Status'] != 'Closed']

#Split between Grants and Tenders
open_grants = url_db[url_db['Type: Grant=1 / Call=0']==1]
open_tenders = url_db[url_db['Type: Grant=1 / Call=0']==0]

```

```

return open_grants

def extract_grants(project_list):
    master_columns = ["identifier", "title", 'callTitle', 'allowPartnerSearch', 'frameworkProgramme_abbreviation',
                      'frameworkProgramme_description', 'actions_status_id', 'actions_status_description',
                      'action_types_description', 'planned_OpeningDate', 'submissionProcedure_id',
                      'submissionProcedure_abbreviation', 'deadlineDates', 'URL', 'ApprovalDate', 'Description Text']
    master_df = pd.DataFrame(columns = master_columns)

    for i in project_list:

        api_url = i

        #Get API Response
        response_full = requests.get(api_url).json()
        response_topic = requests.get(api_url).json()["TopicDetails"]

        #Define which columns are fetched from the API to extract information from
        columns = ["identifier", "title", 'callTitle', 'allowPartnerSearch', 'frameworkProgramme', 'actions']

        x = list()
        for i in columns:
            x.append(response_topic[i])

        #Column names df
        c_df = ["identifier", "Title", 'callTitle', 'allowPartnerSearch', 'frameworkProgramme_abbreviation',
                'frameworkProgramme_description', 'actions_status_id', 'actions_status_description',
                'action_types_description', 'planned_OpeningDate', 'submissionProcedure_id',
                'submissionProcedure_abbreviation', 'Deadline', 'URL', 'ApprovalDate', 'Description']

        #Values df
        v_df = x[:4]

        frameWork = list()
        frameWork.append(x[4]['abbreviation'])
        frameWork.append(x[4]['description'])

```

```

action = list()

try:
    action.append(x[5][0]['status']['id'])
    action.append(x[5][0]['status']['description'])
    action.append(x[5][0]['types'][0]['typeOfMGA'][0]['description'])
    action.append(x[5][0]['plannedOpeningDate'])
    action.append(x[5][0]['submissionProcedure']['id'])
    action.append(x[5][0]['submissionProcedure']['abbreviation'])
    action.append(x[5][0]['deadlineDates'][0])

except:
    action.append('')

url=[api_url]

try:
    latest_changes=[response_topic['latestInfos'][1]['approvalDate']]
except:
    latest_changes=['']

#Join lists
val_df = v_df + frameWork + action + url + latest_changes

#Get project description
try:
    project_description = response_full['TopicDetails']['description']
except:
    project_description = ''

#Clean HTML tags
cleantext = BeautifulSoup(project_description, "lxml").text
cleantext = cleantext.strip('\n')

#Split text by paragraph into: Objective, Scope and Expected impact
clean_des_para = cleantext.split("\n")

```

```

#Append description to text
val_df.append(clean_text)

#Create DF
db = pd.DataFrame(data= val_df).transpose()

master_df = pd.concat([master_df, db], axis=0)

#Format and rename columns
eul = master_df.iloc[:, :16]
eul.set_axis(c_df, axis='columns', inplace=True)

#Add Source
eul['Source'] = 'European Commission'
eul['Retrieval Date'] = date.today()

eul.rename(columns = {'actions_status_description': 'Status',
                      'frameworkProgramme_description': 'Program'}, inplace = True)

return eul

```

StartUp Portugal:

StartUp Portugal is a non-profit organization aiming to foster entrepreneurship and innovation in Portugal with a particular focus on Portugal and entrepreneurs

```

In [7]: def startup_portugal_scraper(url):

        cols = ['Title']
        db = pd.DataFrame(data = None, columns = cols)

        #Open URL
        page = requests.get(url)

        #Open Page with BS
        soup = BeautifulSoup(page.content, "html.parser")

```

```

title, tag, href, country = [], [], [], []

#Title
for a in soup.find_all(class_ = 'program-name'):
    title.append(a.text.strip())

#Tag
for a in soup.find_all(class_ = 'd-flex justify-content-between justify-content-sm-end py-2 py-sm-0'):
    tag.append(a.text.strip())

#Href details
for a in soup.find_all('a', class_ = 'btn-link'):
    a = str(a)
    a = a.split('href="')[1].split('"')[0]
    href.append(a)

#Country
for a in range(len(href)):
    country.append('Portugal')

#Build DB
db_add = pd.DataFrame({'Title': title, 'Tag': tag, 'URL': href, 'Country': country})

#Add Source
db_add['Source'] = 'StartUp Portugal'
db_add['Deadline'] = None
db_add['Retrieval Date'] = date.today()
db_add['Status'] = 'Open'

return db_add

```


Executed Code

The following section saves the links the URL-requests are run on and execute the functions to save the table in a pandas DataFrame format.

Website links

```
In [8]: #EEA pages
eea_url = "https://eeagrants.org/currently-available-funding?combine=&page="

#EIF pages
eif_pages = ['https://www.eif.org/InvestEU/guarantee_products_calls',
             'https://www.eif.org/InvestEU/equity_products_calls',
             'https://www.eif.org/InvestEU/capacity-building-investments-calls']

#Eureka pages
eureka_url = "https://www.eurekanetwork.org/open-calls/?page="

#European commission pages and column names
eu_com_url = "https://ec.europa.eu/info/funding-tenders/opportunities/data/referenceData/grantsTenders.json"
c_df = ['identifier', 'title', 'callTitle', 'allowPartnerSearch', 'frameworkProgramme_abbreviation',
        'frameworkProgramme_description', 'actions_status_id', 'actions_status_description',
        'action_types_description', 'planned_OpeningDate', 'submissionProcedure_id',
        'submissionProcedure_abbreviation', 'deadlineDates', 'URL', 'ApprovalDate', 'Description Text']

#Startup Portugal
Startup_PT_url = 'https://startupportugal.com/programs/#filter=.startups'
```

Building Single Databases:

Run each function with the source URLs and create a database for each source.

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 15 of 17

Master_Scraping_05122022

15.12.22, 10:57

```
In [10]: #EEA
eea_pages = page_scroller_eea(eea_url) #Generate URLs to scroll through the pages
eea_db = EEA_scraper(eea_pages) #Extract

#EIF
eif_db = EIF_scraper(eif_pages) #Landing page scraper
eif_db = EIF_details(eif_db) #Details page scraper
eif_db['URL'] = eif_pages

#Eureka
eureka_pages = Eureka_page_scroller(eureka_url) #Create list of pages that should be scraped
eureka_db = Eureka_scraper(eureka_pages) #Run EEA Landing Page Scraper
project_scroller = eureka_db['URL'] #Withdraw Project Scroller from EEA Scraper
details = Eureka_scraper_details(project_scroller) #Run Details Scraper
eureka_db_full = pd.concat([eureka_db, details], axis = 1) #Concat results of both scrapers to one table

#Startup Portugal
startup_db = startup_portugal_scraper(Startup_PT_url)

#European Commission
eu_grants_list = eu_commission_projects(eu_com_url) #Access EU website to filter all open grants and regarding URL
eu_grants_urls = eu_grants_list['URL']
#eu_grants = extract_grants(eu_grants_urls) #Go through all URLs from the grants list and extract relevant informat

#Cleaning European Commission DF
eu_grants_clean = eu_grants.drop(['identifier', 'frameworkProgramme_abbreviation', 'allowPartnerSearch',
                                  'actions_status_id', 'action_types_description', 'planned_OpeningDate',
                                  'submissionProcedure_id', 'submissionProcedure_abbreviation',
                                  'ApprovalDate'], inplace = True, axis = 1)
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/Master_Scraping_05122022.ipynb?download=false

Page 16 of 17

```
In [10]: #EEA
eea_pages = page_scroller_eea(eea_url) #Generate URLs to scroll through the pages
eea_db = EEA_scraper(eea_pages) #Extract

#EIF
eif_db = EIF_scraper(eif_pages) #Landing page scraper
eif_db = EIF_details(eif_db) #Details page scraper
eif_db['URL'] = eif_pages

#Eureka
eureka_pages = Eureka_page_scroller(eureka_url) #Create list of pages that should be scraped
eureka_db = Eureka_scraper(eureka_pages) #Run EEA Landing Page Scraper
project_scroller = eureka_db['URL'] #Withdraw Project Scroller from EEA Scraper
details = Eureka_scraper_details(project_scroller) #Run Details Scraper
eureka_db_full = pd.concat([eureka_db, details], axis = 1) #Concat results of both scrapers to one table

#Startup Portugal
startup_db = startup_portugal_scraper(Startup_PT_url)

#European Commission
eu_grants_list = eu_commission_projects(eu_com_url) #Access EU website to filter all open grants and regarding URL
eu_grants_urls = eu_grants_list['URL']
#eu_grants = extract_grants(eu_grants_urls) #Go through all URLs from the grants list and extract relevant informat

#Cleaning European Commission DF
eu_grants_clean = eu_grants.drop(['identifier', 'frameworkProgramme_abbreviation', 'allowPartnerSearch',
                                'actions_status_id', 'action_types_description', 'planned_OpeningDate',
                                'submissionProcedure_id', 'submissionProcedure_abbreviation',
                                'ApprovalDate'], inplace = True, axis = 1)
```

Building one coherent database:

Merging the various single databases to one coherent and harmonized database that can then be further processed. Eventually the database is saved as a CSV-file.

```
In [ ]: #Merge Databases
platform_db = pd.concat([eea_db, eif_db, eureka_db_full, startup_db], axis = 0,
                        join = 'outer', ignore_index=True)

#Remove irrelevant columns
platform_db.drop(['Full Description', 'Application Process', 'Tag'], inplace = True, axis = 1)

#Add EU Grants
platform_db_full = pd.concat([platform_db, eu_grants], axis = 0,
                            join = 'outer', ignore_index=True)

#Clean NaNs
platform_clean = platform_db_full.replace({np.nan: None})

#Save as CSV
platform_clean.to_csv('Platform_db.csv')
```

8.5.2 Matching Model

NLP_testing

15.12.22, 11:18

Overcoming Financing Challenges for Impact: The Development of a Match-Making Platform for Impact Organisations and Public Funding Opportunities

Imports

For the general imports we import the numpy and pandas library to effectively handle large amounts of data. The CSV library to read data from CSV-files, time library to track runtime of the code and matplotlib to create visualizations of results.

For the scraping imports we use the requests library, as well as urllib to access websites and APIs to retrieve data. BeautifulSoup is important to effectively navigate through HTML-data and systematically obtain relevant information.

For the NLP models we import the rake_nltk, as a nlp-toolkit to process data in human language and engineer keywords. Additionally, the GenSim model is downloaded with a trained dataset.

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 1 of 544

NLP_testing

15.12.22, 11:18

```
In [1]: #General imports
import numpy as np
import pandas as pd
from datetime import date
import csv
import time
import matplotlib.pyplot as plt

#Scraping imports
import requests
from urllib.request import urlopen as uReq
from bs4 import BeautifulSoup
import bs4

#NLP imports
import rake_nltk
from rake_nltk import Rake
import nltk
import gensim.downloader as api
wv = api.load('word2vec-google-news-300')

#Display options
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

Read and format input data

The obtained data is stored in several CSV-files. This cell imports the CSV-files and reads them into Pandas DataFrame format. Additionally, harmonization in column names is performed and the input data is labelled. Thus, types of entities, start-ups and NGOs can be distinguished

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 2 of 544

```
In [2]: #Import data from CSV files
ngo_db = pd.read_csv('ngoexplorer-portugal.csv')
startup_db = pd.read_csv('Impact_start_ups_Europe.csv')
platform_clean = pd.read_csv('Platform_db.csv')

#Rename columns to prepare for working on the description
ngo_db = ngo_db.rename(columns={"objectives": "Description"})
ngo_db['Type'] = 'NGO'

startup_db = startup_db.rename(columns={"TAGS": "Description", "NAME": 'name'})
startup_db['Type'] = 'Start-Up'
```

Engineer keywords and keep nouns only

Feature engineering is performed on the description of the different data sources. Tags are created applying RAKE-model on the Description of each database. Keywords are refined by including only tags that consist of max. 2 words and min. consist of 1 word. Repeated tags are removed.

The created tags are then saved in a new column for each database.

```
In [3]: #Create tags out of the description
def create_tag(row):
    rake_nltk_var = Rake(min_length=1, max_length=2, include_repeated_phrases=False) #Create keywords with max
    text = str(row['Description'])
    text = str.lower(text)
    rake_nltk_var.extract_keywords_from_text(text)
    keyword_extracted = rake_nltk_var.get_ranked_phrases()
    return keyword_extracted

#Add column for tags for each db
platform_clean['Tags'] = platform_clean.apply(lambda row: create_tag(row) , axis=1)
ngo_db['Tags'] = ngo_db.apply(lambda row: create_tag(row) , axis=1)
startup_db['Tags'] = startup_db.apply(lambda row: create_tag(row) , axis=1)
```

Refine keywords with nouns only

To reduce noise in the tags captured and increase the relevance of tags, all word types besides nouns are removed from the tags. Each word is tokenized and labelled with a category of the word type. All non-nouns are filtered in the next step.

The results are appended on the existing databases for each input source in a new column as 'Tags_clean'

```
In [4]: def filter_nouns(row):
x = []

for i in row['Tags']:
    #Tokenize entire tag
    tokens = nltk.word_tokenize(i)

    #Tag each token with word type
    tagged = nltk.pos_tag(tokens)

    #Filter and keep only nouns
    nouns = [word for word,pos in tagged if (pos == 'NN' or pos == 'NNP' or pos == 'NNS' or pos == 'NNPS')]

    x.append(nouns)

return x

#Create another row with cleaned and tags
startup_db['Tags_clean'] = startup_db.apply(lambda row: filter_nouns(row) , axis=1)
ngo_db['Tags_clean'] = ngo_db.apply(lambda row: filter_nouns(row) , axis=1)
platform_clean['Tags_clean'] = platform_clean.apply(lambda row: filter_nouns(row) , axis=1)
```

Calculating similarity

To calculate the similarity between an entity and one project we need to break down the steps since for this process different functions are applied that work recursively.

First of all, pairings of keywords are created with the **create_pairs** function. These pairings have the keywords of the entity on one side and the keywords of the project on the other side. Assuming an entity has n-keywords and a project has m-keywords then the function create_pairs will create n pairs for each keyword of the entity and n x m keywords for the entire entity-project match.

The **flatten** function formats the input data that nested lists are removed, as well as empty list elements.

The **calculate_similarity** function receives the pairs of keywords that were returned by the create_pairs function. For each pairing the function applies the GenSim model to determine similarity between the two words and converts the lists to a coherent dataframe with the columns: Keyword_A, Keyword_B and the Similarity. The values are then sorted according to their similarity to give insight on which keywords are most similar in the calculation.

Eventually, the **get_tuples** function will add a column to the dataframe to highlight highest matched keywords for each entity-project match.

Note: For the course of this project, data was sampled to reduce runtime.

```
In [5]: #Sampling
sample_project = platform_clean.iloc[180]
sample_ngo = ngo_db.iloc[15]
sample_start_up = startup_db

test_ngo = sample_ngo['Tags_clean']
test_project = sample_project['Tags_clean'][:10]
test_start_up = sample_start_up
```

Perform pairing of keywords and calculate similarity for each pair

```
In [6]: #Flatten format of nested lists in input data
def flatten(l):
    #Flatten
    x = [item for sublist in l for item in sublist]
    #Remove empty elements
    while("" in x):
        x.remove("")

    return x

#Create pairs to feed into GenSim
def create_pairs(entity, project):
    #Receives two lists of keywords from entity and projects and creates a flattened pairing of both keywords
    #so every keyword of an entity is matched with a keyword of the project
    #Returns a list of tuples (x,y)

    pair = []

    #Flatten both lists
    entity = flatten(entity)
    project = flatten(project)
    i = 0

    for i in entity:
        x = [i] * len(project)
        y = project

        #Create tuples
        pairing = list(zip(x,y))

        pair.append(pairing)

    return flatten(pair)

#Calculate similarities for each pair of keywords
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 7 of 544

```
def calculate_similarity(pairs):
    #Receives a list of tuples (x,y) based on which pairings the similarity is calculated and returns a
    #dataframe with the two keywords as well as the calculated similarity index

    data = []

    for w1, w2 in pairs:
        try:
            sim = (w1, w2, wv.similarity(w1, w2))

            #Create threshold
            if wv.similarity(w1, w2) >= 0.15:
                data.append(sim)

        except KeyError:
            pass

    df = pd.DataFrame(data, columns=['Keyword_A', 'Keyword_B', 'Similarity'])
    df = df.sort_values(by=['Similarity'], ascending=False)
    return df
```

Append tuples of keywords with highest similarity scores for sanity checks

This function adds a column to the dataframe to which it adds keyword pairs and the similarity score saved in a tuple. The tuples in one row are sorted in descending order for the similarity score.

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 8 of 544

```
In [7]: def get_tuples(df):
#Receives a df with three columns and the float of similarity that scans through the list of matched
#keywords and returns a list of tuples if the similarity between both keywords is larger than 0.2

tuples = []
for index, row in df.iterrows():
    col1 = row['Keyword_A']
    col2 = row['Keyword_B']
    col3 = round(row['Similarity'],3)

    #Threshold
    if col3 > 0.05:
        tuples.append((col1, col2, col3))

    #Sort the list of tuples in descending order
    tuples.sort(key=lambda t: t[2], reverse=True)

return tuples
```

Automize comparison process with organizations and entities

This function connects receives a list of organizations and projects and connects all the other functions to automatize the process for a large number of projects and organizations. It returns a dataframe with the organization in one column, the project and the similarity index in another column.

```
In [8]: #Calculate the similarity for each organization and all projects
def similar_projects(orgs, projects):

    data = []

    #Convert df to lists to ease the inputs for other functions
    orgs_name = orgs['name'] #Organizations
    orgs_tags = orgs['Tags_clean']
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 9 of 544

```
pro_name = projects['Title'] #Projects
pro_tags = projects['Tags_clean']

#Iterate through all organizations and projects
n = 0
m = 1

#Organizations
for i in range(len(orgs_name)):
    x = orgs.iloc[i]['Tags_clean']

    n +=1
    start = time.time()

    if n >=1:
        print(str(m) + ' organizations matched out of ' + str(len(orgs_name)))
        end = time.time()
        time_req = str(round((end - start),2))

    m=m+1

#Projects
for j in range(len(pro_tags)):

    y = projects.iloc[j]['Tags_clean']

    #Create pairs for all project for one organization
    pairs = create_pairs(x, y)

    #Calculate similarity
    sim = calculate_similarity(pairs)
    avg_sim = np.mean(sim['Similarity']) #Mean for all compared keywords

    #Retrieve highest matched Keywords
    matched_keywords = get_tuples(sim)

    #Create tuple to add to df
```

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 10 of 544

```

add = (orgs.iloc[i]['name'], projects.iloc[j]['Title'], round(avg_sim,5), matched_keywords)
data.append(add)

#Create DF to get all projects with their similarities
df = pd.DataFrame(data, columns=['Entity', 'Project', 'Similarity', 'Highest Matched Keywords'])
df = df.sort_values(by=['Similarity'], ascending = False )

return df

```

Sampling and testing

For feasibility reasons a subsample of the entire dataframe was created to showcase the matching mechanism.

```

In [30]:
#NGO
test_ngo2 = ngo_db.iloc[:,['name', 'Tags_clean', 'Type']]
test_ngo2

#Projects
test_project2 = platform_clean.iloc[:,['Title', 'Tags_clean']]
#test_project2

#Start_Ups
test_start_up = sample_start_up[:,['name', 'Tags_clean', 'Type']]
test_start_up

#All entities
entities_db = pd.concat([test_start_up, test_ngo2], axis = 0)

#Create a sample of 30 random picked organizations to assess overall similarity of keywords
test_entities = entities_db.reset_index(drop=True).sample(n = 30)
test_entities = count_tags(test_entities)

#Create a sample of 50 random picked organizations to assess overall similarity of keywords
test_project2 = test_project2.sample(n=250)
test_project2 = count_tags(test_project2)

```

```

In [31]:
#Track runtime
start = time.time()

#Testing
testing_sample_entities = similar_projects(test_entities, test_project2)

end = time.time()
time_req = str(round((end - start)/60,2))

print('\nTotal time required for the code to run was: ' + time_req + 'minutes')

#Testing the sample
testing_sample_entities = testing_sample_entities.reset_index()

```



```

1 organizations matched out of 30
2 organizations matched out of 30
3 organizations matched out of 30
4 organizations matched out of 30
5 organizations matched out of 30
6 organizations matched out of 30
7 organizations matched out of 30
8 organizations matched out of 30
9 organizations matched out of 30
10 organizations matched out of 30
11 organizations matched out of 30
12 organizations matched out of 30
13 organizations matched out of 30
14 organizations matched out of 30
15 organizations matched out of 30
16 organizations matched out of 30
17 organizations matched out of 30
18 organizations matched out of 30
19 organizations matched out of 30
20 organizations matched out of 30
21 organizations matched out of 30
22 organizations matched out of 30
23 organizations matched out of 30
24 organizations matched out of 30
25 organizations matched out of 30
26 organizations matched out of 30
27 organizations matched out of 30
28 organizations matched out of 30
29 organizations matched out of 30
30 organizations matched out of 30

```

Total time required for the code to run was: 12.05minutes

Data visualization

To achieve further insights into the results we're applying data visualization to get a detailed view of the distribution of the results.

In [32]: `#Count how many keywords are allocated to each entity`

http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 13 of 544

```

#Count how many keywords are allocated to each entity
def count_tags(df):
    #The function counts the number of keywords and saves them in a new column
    df['Project tags counted'] = 0

    for index, row in df.iterrows():
        tags = row['Tags_clean']

        count = 0
        for tag_list in tags:
            count += len(tag_list)

        #Set value of the "tag_count" column
        df.at[index, 'Project tags counted'] = count

    return df

#Visualize tags counted in a histogram
def create_histogram(df):
    # Import the matplotlib library
    import matplotlib.pyplot as plt

    # Get the values of the "Tags_counted" column
    tags = df['Project tags counted']

    # Create the histogram
    plt.hist(tags, bins=30)

    # Set the x-axis and y-axis labels
    plt.xlabel('Number of tags')
    plt.ylabel('Frequency')
    plt.title('Nr. of keywords per project', fontweight = 'bold')

    # Show the histogram
    plt.show()

test_entities = pd.concat([test_start_up, test_ngo2], axis = 0)

```

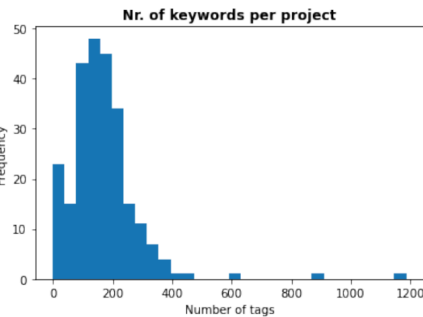
http://localhost:8888/nbconvert/html/Desktop/Business_Analytics/Work%20Project/NLP_testing.ipynb?download=false

Page 14 of 544

```
#Create count tags column
viz_ent = count_tags(test_entities)
viz_project = count_tags(test_project2)

#Create histogram
#create_histogram(viz_ent)
create_histogram(viz_project)

#Get key metrics descriptive statistics
viz_ent.describe()
viz_project.describe()
```



Out[32]:

Project tags counted	
count	250.000000
mean	166.744000
std	120.886844
min	0.000000
25%	104.250000
50%	154.500000
75%	208.000000
max	1185.000000

Histogram

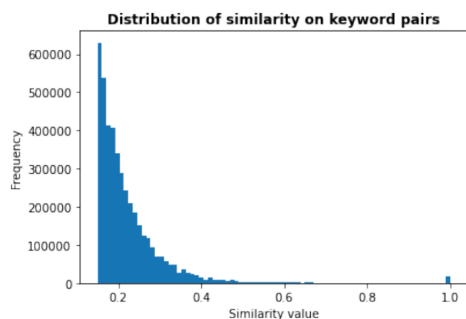
```
In [33]: def plot_histogram(series):
# Initialize a list for the float values
values = []

# Loop through the lists in the series
for lst in series:
    # Loop through the tuples in the list
    for tup in lst:
        # Check if the third element of the tuple is a float
        if isinstance(tup[2], float):
            # Add the third element of the tuple to the list of float values
            values.append(tup[2])

# Create a histogram using the list of float values
plt.hist(values, bins=80)
# Label the x-axis as "Float value"
plt.xlabel('Similarity value')
# Label the y-axis as "Frequency"
plt.ylabel('Frequency')
plt.title('Distribution of similarity on keyword pairs', fontweight = 'bold')
# Show the plot
plt.show()

#Create a series
floats = testing_sample_entities['Highest Matched Keywords']

#Plot histogram
plot_histogram(floats)
```



Analysis of Results

Calculate overall similarity of keywords

```
In [40]: #Overall average of similarity in keywords and the number of keyword pairs considered
def calc_avg_and_count(data):
    keywords = data['Highest Matched Keywords']

    #Initialize variables for the sum and count of float elements
    sum_floats = 0
    count_tuples = 0

    #Loop through the lists of tuples in the column
    for list_of_tuples in keywords:
        for tup in list_of_tuples:
            if isinstance(tup[2], float):
                sum_floats += tup[2] #Calculate sum of floats
                count_tuples += 1 #Count number of entries

    #Calculate the average of the float elements
    avg_floats = sum_floats / count_tuples

    # Print the average and count of float elements
    print('Average of float elements:', round(avg_floats,4))
    print('Count of tuples:', count_tuples)

calc_avg_and_count(testing_sample_entities)

testing_sample_entities = testing_sample_entities.reset_index(drop = True)
testing_sample_entities.describe()
```

```
Average of float elements: 0.2209
Count of tuples: 4297573

Out[40]:
```

	index	Similarity
count	7500.000000	7137.000000
mean	3749.500000	0.219695
std	2165.207842	0.014274
min	0.000000	0.153660
25%	1874.750000	0.212470
50%	3749.500000	0.218730
75%	5624.250000	0.225150
max	7499.000000	0.360000

```
In [41]: test_print = testing_sample_entities[testing_sample_entities['Similarity'] != np.nan]
test_print
```

```
Out[41]:
```

	index	Entity	Project	Similarity	Highest Matched Keywords
0	3389	SILOAM CHRISTIAN MINISTRIES LIMITED	Support for Hiring (IEFP)	0.36000	[(support, support, 1.0), (support, support, 1.0), ...]
1	1196	DFN PROJECT SEARCH	GOVERNANCE	0.35018	[(people, people, 1.0), (people, people, 1.0), ...]
2	2085	machineMD	Canada France call for innovative projects	0.34678	[(technologies, projects, 0.347)]
3	2047	machineMD	Tech Visa	0.34073	[(technologies, tech, 0.413), (diagnostics, te...)]
4	7389	ORCA RESCUES FOUNDATION	Support for Hiring (IEFP)	0.33035	[(support, support, 1.0), (help, support, 0.46...)]

5	3889	KERYGMA 180	Support for Hiring (IEFP)	0.31799	[(support, support, 1.0), (services, support, ...
6	4685	PROJECT MOZAMBIQUE	INNOVATION	0.30579	[(provision, proposals, 0.333), (provision, in...
7	2389	THE VIRTUOUS WOMAN	Support for Hiring (IEFP)	0.29983	[(support, support, 1.0), (benefit, support, 0...
8	2033	machineMD	Call Tourism (Portugal Ventures)	0.29976	[(technologies, ventures, 0.3)]
9	2168	machineMD	Pervasive photonics - multi-technology integra...	0.29867	[(technologies, systems, 0.612), (technologies...
10	2169	machineMD	CELTIC-NEXT Spring 2023 call for projects	0.29842	[(technologies, technology, 0.833), (diagnosti...
11	3585	FOOD OF WAR PLATFORM	Canada France call for innovative projects	0.29810	[(uk, canada, 0.682), (uk, france, 0.515), (et...
12	1639	PAO VIVO	Support for Hiring (IEFP)	0.29726	[(benefit, support, 0.424), (faith, support, 0...
13	4946	ALSTROM SYNDROME EUROPE	GOVERNANCE	0.29626	[(people, people, 1.0), (people, people, 1.0),...
14	2080	machineMD	Future Space Ecosystem: Management and Coordin...	0.29429	[(technologies, technologies, 1.0), (technolog...
15	2030	machineMD	Industrial leadership in AI, Data and Robotics...	0.29416	[(technologies, technologies, 1.0), (technolog...
16	2086	machineMD	Intelligent data acquisition and analysis of m...	0.29251	[(technologies, technologies, 1.0), (technolog...
17	2027	machineMD	Interoperable solutions for positive energy di...	0.29134	[(technologies, technologies, 1.0), (technolog...
18	2196	machineMD	GOVERNANCE	0.29095	[(technologies, initiatives, 0.429), (diagnost...
19	4435	THE CONFERENCE OF EUROPEAN RABBIS	INNOVATION	0.29076	[(education, initiatives, 0.363), (advancement...