



UNIVERSITÀ DEGLI STUDI DI TRENTO

DEPARTMENT OF MATHEMATICS

MASTER'S DEGREE IN DATA SCIENCE

**Measuring corporate contributions to
UN SDGs using alternative data
and NLP techniques**

Supervisor:

Giuseppe Alessandro Veltri

Signature

Giuseppe A. Veltri

Candidate:

Antonia Donvito

DATE OF DISCUSSION

23 October 2020

*«Accept your lack of knowledge
and use it as your asset»*

Natalie Portman

Contents

Introduction	4
1. Evaluating Corporate Sustainability	6
1.1. Sustainable Development and UN SDGs	6
1.2. Corporate Sustainability Reporting	11
1.3. Corporate Social Responsibility Assessment	15
2. FinScience Sustainability Score	19
2.1. Score definition and methodology	19
2.2. Internal Data	24
2.2.1. Corporate Website	25
2.2.2. CSR/Sustainability Report	31
2.2.3. Memberships & Shareholder Proposals	34
2.2.4. Balance Sheet	35
2.3. External Data	36
2.3.1. FinScience news	37
2.3.2. Rankings and Indexes	38
2.3.3. Controversial Activities	39
2.3.4. Reviews	40
3. Data analysis and findings on S&P 500 companies	43
3.1. Internal Data	43
3.2. External Data	56
3.3. Scores	76
Conclusion	87
References	90
Annexes	94

Introduction

While sustainability has become extraordinarily popular in the public domain over the last decade, with the young generations urging both politicians and corporations to make a change, the debate about corporate responsibility dates back to more than a century and a half ago. In order to achieve sustainability, the interaction between three actors is indeed necessary: governments, civil society and private businesses on local, national and international levels, with specific sustainability strategies enabled by national regulative guidelines. The introduction of the Sustainable Development Goals by the United Nations in 2015 has established a common framework for both private and public sectors to take action and measure progress towards sustainability.

This thesis will be focusing on corporate sustainability performance and, more specifically, on the construction of a sustainability scoring system that evaluates business contributions to the 17 SDGs, by combining traditional and alternative data. This constituted the core of the project that I was involved in during the last nine months as a Data Science intern at FinScience S.r.l., a data-driven fintech based in Milan, Italy. Hence, I will retrace the steps that were taken from the definition of a novel methodology to the final results, providing an in-depth discussion of the data collection and data analysis phases. I will do so in three distinct chapters.

The first chapter will be centred on a literature review of corporate sustainability and its evolution over time. The introduction of the UN SDGs will be discussed to understand how this framework poses itself with respect to the previous global agenda and to the private sector in particular. Then, the existing tools to measure corporate sustainability performance will be discussed, as well as the existing rankings and indices to evaluate and compare corporate sustainability performances.

The second chapter will move from the current sustainability assessment systems

to the FinScience Score. I will provide details on the methodology behind the score, with reference to its main distinctive features: the SDGs framework, the combination of traditional and alternative data, the employment of Natural Language Processing techniques, the value added by a materiality analysis. The score is also characterized by a dualistic approach to the corporate sustainability performance, that considers both the information disclosed by the company (internal score) and the external perception of corporate performance by third parties (external score). Then, the data collection process will be extensively described, with reference to the Python scripts that I wrote to automatize the extraction of relevant information from more than 300 hundred sources.

In the last part of this work, I will discuss the results of the data collection and of the data analysis performed on the 500 companies that constitute the Standard and Poor 500 index, which were selected as sample data for testing. Insights on both internal and external data will be extracted and analysed with the aim of finding trends and patterns in corporate sustainability disclosure. Then, I will show the results of the final scores that emerge from such data, looking at the performance by SDG, by data source type (how well the company performs based on self-disclosed data versus external reactions), by industry and by additional exploratory criteria. To conclude, I will review the limitations of this study and suggest what could be improved to fine-tune corporate sustainability assessment.

1. Evaluating Corporate Sustainability

What is sustainability? What are the UN Sustainable Development Goals? What does corporate sustainability mean and how can it be measured? The first chapter of this thesis will answer these preliminary questions to define the background in which the need for a corporate sustainability score has emerged. The first paragraph proposes a literature review on the definitions of sustainability, with a specific focus on the business perspective and on the recent introduction of the UN Sustainable Development Goals. The second paragraph explores the existing framework to report on corporate sustainability and to operationalize business contributions to SDGs through measurable indicators. The last part focuses on the most famous existing indices, ranking and score systems that have been proposed in the last years as tools for both investors and corporations to evaluate and compare sustainability performances.

1.1. Sustainable Development and UN SDGs

Although the term of sustainability has roots in the past, the concept of sustainable development only emerged after the damages caused by the industrial revolutions and the two world wars on the environment, when it became evident that the idea of progress as continuous improvement was not maintainable and the expectations of constant growth were an illusion (Du Pisan, 2007; Waas, Hugé, Verbruggen, & Wright, 2011). Global environmental awareness emerged into numerous discussions involving the United Nations, but it was the publication in 1987 of the report “Our Common Future” by Brundtland, at that time chair of the UN General Assembly, that paved the way for the modern thinking on sustainability, with the famous definition: *“Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs”* (WCED, 1987, pag.16). In other words, the economic growth should be pursued in a way that enables all humans to satisfy their basic

needs by guaranteeing the future generations the possibility to also do so. This general framework was expanded and operationalized in 1994 by Elkington, with the aim of measuring corporate sustainability performance in America. He defined an accounting framework that incorporates three dimensions, namely economy, environment and society, also known as the Triple Bottom Line (Elkington, 1992). The model evaluates the performance of corporations by expanding the traditional profit-centred accounting frameworks to the social and environmental impact made by the company (Goh, Chong, Jack, & Mohd Faris, 2020). As the TBL framework simultaneously focuses on the business impact on Profit, Planet and People, it is also known as the 3Ps model. The last twenty years of the 20th century were also dominated by Socially Responsible Investing (SRI), an expression that refers to any type of investing strategies that aims at having a positive social impact, with the selection of assets that meet a set of ethical criteria (Hill, Ainscough, Shank, & Manullang, 2007).

Some years later, in 2004, the acronym ESG, that refers to Environmental, Social and Governance factors, was coined in the report “Who Cares Wins” by the UN Global Compact and, since then, the ESG investment strategy has become the main guideline for sustainability investing and SRI in the last two decades (United Nations, The Global Compact, 2004). Environmental factors include environmental risks such as gas emission, waste management and energy production; Social factors regards human rights, such as workforce diversity, customer satisfaction and labour standards; Governance factors are about, for instance, bribery and corruption, board composition, lobbying (CFA Institute, 2015). The fact that ESG issues have become extremely popular is demonstrated by the ESG investing having grown to more than \$30 trillion in 2018, and expected to reach \$50 trillion this year (Global Sustainable Investment Alliance, 2018). The predominate strategy in Europe is negative screening, the exclusion of industry sector or companies

involved in controversial activities, such as army or tobacco production, while the United States prefer ESG integration, which is the inclusion of ESG issues in investment analysis¹.

In the meantime, the debate around sustainable development policy at the international level became more and more open to the private sector, as the crucial role of small and big companies in achieving sustainable goals was acknowledged (Scheyvens, Banks, & Hughes, 2016). In this sense, an important milestone was placed by the United Nations Millennium Summit in 2000, when the Eight Millennium Development Goals, to be achieved by 2015, were set (i.e. to eradicate extreme poverty and hunger; to achieve universal primary education; to promote gender equality and empower women; to reduce child mortality; to improve maternal health; to combat HIV/AIDS, malaria, and other diseases; to ensure environmental sustainability; to develop a global partnership for development)². While the MDGs have succeeded in drawing the global attention to the needs of the poorest countries and to poverty alleviation (Lomazzi, Borisch, & Laaser, 2014), this framework, by mainly addressing human development and well-being, did not adequately cover environmental issues, such as climate change, and social equality, nor it was clearly structured (for a review, see Fehling, Nelson, & Venkatapuram, 2013).

The MDGs were overcome by the definition of a novel set of goals, the 17 Sustainable Development Goals, announced by the United Nations in Rio de Janeiro in 2012 and adopted from 2015 as part of the 2030 Agenda for Sustainable Development³. The 17 Sustainable Development Goals, underpinned by 169 targets, are: Goal 1: *No Poverty*; Goal 2: *Zero Hunger*; Goal 3: *Good Health and Well-Being*; Goal 4: *Quality Education*; Goal 5: *Gender Equality*; Goal 6: *Clean Water and Sanitation*; Goal 7:

¹ <https://sdg.iisd.org/news/gsia-report-finds-increase-in-sustainable-investing/>

² https://www.who.int/topics/millennium_development_goals/about/en/

³ For a timeline see

https://www.unece.org/are/en/home/sustainable-development/international-cooperation/2030agenda/un- _milestones-in-sustainable-development.html

Affordable and Clean Energy; Goal 8: Decent Work and Economic Growth; Goal 9: Industry, Innovation and Infrastructure; Goal 10: Reduced Inequalities; Goal 11: Sustainable Cities and Communities; Goal 12: Responsible Consumption and Production; Goal 13: Climate Action; Goal 14: Life Below Water; Goal 15: Life on Land; Goal 16: Peace, Justice and Strong Institutions; Goal 17: Partnership for the Goals (Figure 1)⁴.



Figure 1: the 17 United Nations Sustainable Development Goals

The 2030 Agenda in which the goals are included was defined as “*a plan of action for people, planet and prosperity. It also seeks to strengthen universal peace in larger freedom. We recognize that eradicating poverty in all its forms and dimensions, including extreme poverty, is the greatest global challenge and an indispensable requirement for sustainable development.*” (UN General Assembly, 2015, pag. 1). Hence, compared to the MDGs, the SDGs come with a more comprehensive approach to sustainability that recalls and expand the triple bottom approach by establishing five main pillars, namely People,

⁴ <https://sdgs.un.org/goals>

Planet, Prosperity, Peace and Partnership⁵. In addition, the Agenda emphasizes even further the role of the private sector: “*Private business activity, investment and innovation are major drivers of productivity, inclusive economic growth and job creation. We acknowledge the diversity of the private sector, ranging from micro-enterprises to cooperatives to multinationals. We call upon all businesses to apply their creativity and innovation to solving sustainable development challenges.*” (UN General Assembly, 2015, pag. 29). It is now clear that the business is called to actively contribute to sustainable development.

From a stakeholder’s perspective, the SDGs offer a common guideline to report on corporate sustainability performance, which in turn allows easier communication with investors, governments, clients and employees (Huber, Comstock, & Smith, 2018). However, as I will discuss later, there is still not a standardized SDGs reporting system. From an investor’s perspective, investing in SDGs “*returns to society and returns to shareholders*” (Schramade, 2017, pag. 89). In fact, not only do SDGs set a detailed strategy on a global level to achieve a better future, but they are also estimated to open opportunities to business for \$12 trillion⁶. The Global Impact Investing Network has profiled several impact investors and has found that investors consider SDGs as a useful framework for communications with stakeholders, clients and other investors, as well as an attractive entry point to engage in impact investing (Global Impact Investing Network, 2016).

⁵

<https://www.unssc.org/news-and-insights/blog/sustainable-development-what-there-know-and-why-should-we-care/>

⁶

<https://www.undp.org/content/undp/en/home/blog/2017/8/25/More-than-philanthropy-SDGs-present-an-estimated-US-12-trillion-in-market-opportunities-for-private-sector-through-inclusive-business.html>

However, some questions still remain. How should a company report on SDGs? How to invest in SDGs? The following paragraphs will address these challenges by looking at the operationalization of the SDGs in corporate reporting and sustainability evaluation.

1.2. Corporate Sustainability Reporting

As previously mentioned, sustainable development also requires a specific legal framework to be put into practice. However, when it comes to reporting on non-financial performances, the regulatory context may vary significantly. In the European Union, the European Commission's Non-financial Reporting Directive in 2014 established that organizations with more than 500 employees should report on environmental, social and governance performance. More specifically, the directive indicates the issues that companies should report on in regards to environmental matters: "*details of the current and foreseeable impacts of the undertaking's operations on the environment, and, as appropriate, on health and safety, the use of renewable and/or non-renewable energy, greenhouse gas emissions, water use and air pollution*"; in regard to social matters: "*the information provided in the statement may concern the actions taken to ensure gender equality, implementation of fundamental conventions of the International Labour Organisation, working conditions, social dialogue, respect for the right of workers to be informed and consulted, respect for trade union rights, health and safety at work and the dialogue with local communities, and/or the actions taken to ensure the protection and the development of those communities*"; and in regards to human rights and bribery, "*the non-financial statement could include information on the prevention of human rights abuses and/or on instruments in place to fight corruption and bribery*" (Directive 2014/95/EU⁷).

⁷ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095>

However, the companies that will be analyzed belong to the S&P 500 index and, as such, are mostly American large organizations, for which there are not as strict legal constraints on non-financial reporting as those aforementioned. In fact, in 2019, proposals to improve Environmental, Social, and Governance disclosures were discussed in the House of Representatives⁸, and the US Securities and Exchange Commission published its rules on human capital reporting (pertaining, for instance, to management of workforces and number of employees)⁹, that were further modernized in August 2020¹⁰. However, ESG reporting in the United States remains mainly based on voluntary disclosure and is largely market-driven (Harper Ho, 2020) .

On the other hand, there are several frameworks that have committed to provide organizations with guidelines on how to report on progress in support of the SDGs. In 2018, almost 80% of the S&P 500 companies issued a sustainability report, 75% of which mentioned at least one reporting framework. The most used framework is the Global Reporting Initiative (GRI)¹¹, followed by the Carbon Disclosure Project (CDP)¹²; 12% provided an index for Sustainable Development Goals; smaller minorities referred to Sustainability Accounting Standards Board (SASB)¹³ and International Integrated Reporting Council (IIRC)¹⁴ frameworks (Investor Responsibility Research Center Institute (IRRCI), 2018). Such evidence will be further discussed in the third chapter of this thesis and compared with the results that emerged while investigating the most recent sustainability reports. What support do reporting frameworks offer in measuring corporate sustainability performance and how do they pose themselves with respect to the Sustainable Development Goals?

⁸ <https://docs.house.gov/Committee/Calendar/ByEvent.aspx?EventID=109770>

⁹ <https://www.sec.gov/comments/4-711/4-711.htm>

¹⁰ <https://www.sec.gov/news/press-release/2020-192>

¹¹ <https://www.globalreporting.org/>

¹² <https://www.cdp.net/en>

¹³ <https://www.sasb.org/>

¹⁴ <https://integratedreporting.org/>

The GRI was founded in late 1997 by the Coalition for Environmentally Responsible Economies (CERES) and it is currently considered as the leading tool in business reporting. It suggests a set of indicators to measure the ESG performance at the operational level. In 2017, GRI and the UN Global Compact launched a collaborative publication inventory of possible disclosures per SDG, at the level of the 169 targets, in which the 17 SDGs are mapped against the GRI Standards (Global Reporting Initiative, United Nations Global Compact, 2017; Global Reporting Initiative, United Nations Global Compact, WBCSD, 2016). Szennay and colleagues have conducted an independent study that confirms, through a cluster analysis, the robustness of the links between UN SDGs and GRI Standards, though indicating some goals that are covered better than others (Szennay, Szigeti, Kovács, & Szabó, 2019). The CDP, formerly Carbon Disclosure Project, is the most accredited organization for environmental disclosure, and it supports both the private and the public sector with reporting on this matter. The organization indicates how disclosure to CDP contribute to specific SDGs, namely those that address environmental issues¹⁵, and it is in turn indicated among the indicator sources suggested by the SDG Compass Inventory of Business Indicators¹⁶.

The SASB provides a framework for sustainability accounting, which was particularly crucial for building the score since it is aligned with the UN goals, as I will be extensively discussing over the next chapter. Moreover, the SASB adopts a specific investor-orientated perspective, or, in their words, “*SASB’s mission is to help businesses around the world identify, manage and report on the sustainability topics that matter most to their investors*”¹⁷. SASB does so by identifying material sustainability factors, namely those issues that are likely to have an impact on the financial performance of the company and are therefore of interest for investors, with materiality being defined by U.S. securities

¹⁵ <https://www.cdp.net/en/policy-and-public-affairs/sustainable-development-goals>

¹⁶ <https://sdgcompass.org/business-indicators/>

¹⁷ <https://www.sasb.org/>

law as “*a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the ‘total mix’ of information made available*” (as cited in Investor Responsibility Research Center Institute (IRRCI), 2018, pag. 21).

More specifically, SASB identifies 5 main sustainability areas, i.e. Environment, Social Capital, Human Capital, Business Model and Innovation, Leadership and Governance, that group 30 different disclosure topics (Sustainability Accounting Standards Board (SASB), 2017). It also defined the Sustainability Industry Classification Industry (SICS), which identifies 77 different industries, that group companies based on shared sustainability risks and opportunities¹⁸, and which are in turn clustered into 11 sectors, i.e. Consumer Goods, Extractives & Minerals Processing, Financials, Food & Beverage, Health Care, Infrastructure, Renewable Resources & Alternative Energy, Resource Transformation, Services, Technology & Communications, Transportation¹⁹. Having determined the sustainability issues and the industries, SASB Materiality Matrix indicates which issues are material for each industry (see Annex A). A regression analysis of stock returns conducted on more than 2000 organizations has proven that companies that perform well on material sustainability issues outperform companies that perform poorly on sustainability material issues, validating even further the thesis that investing in sustainability topics that are relevant for the industry is indeed profitable (Khan, Serafeim, & Yoon, 2015).

A possible mapping between SASB sustainability topics and UN Sustainable Development Goals was proposed by Betti, Consolandi and Eccles who indicated for each issue of the 30 SASB what SDGs are impacted - goal level- and also how many of the 169 SDGs targets are impacted - target level (Betti, Consolandi, & Eccles, 2018). To give an

¹⁸ <https://www.sasb.org/standards-overview/download-current-standards/>

¹⁹ <https://www.sasb.org/wp-content/uploads/2018/11/SICS-Industry-List.pdf>

example, within the SASB General Category “Environment”, the “GHG emissions” issue was considered as having an impact on two targets of SDG 7- *Affordable and Clean Energy*, one target of SDG 9 - *Industry, Innovation and Infrastructure*, and one target SDG 12 - *Responsible Production and Consumption*.

To sum up, different approaches have been proposed to report on corporate sustainability with specific reference to SDGs, hence companies are free to choose the frameworks that they find suitable to their needs to report on their sustainability performance. Luckily, the alignment of most of the renowned frameworks to SDGs allows to re-map the sustainability (or ESG) indicators to the UN Agenda, and therefore, in most of the cases, it is possible to identify the contributions of companies to SDGs, at least at a goal level.

1.3. Corporate Social Responsibility Assessment

Research has shown that both the several conflicting attempts to define sustainability and the confusion about how it should be practically operationalized in specific industries contribute to the current lack of a unique sustainability assessment system (Sroufe, Operationalizing Sustainability, 2016). In a survey research involving 17 multinational companies, when stakeholders were asked how sustainability is operationalized within their company, most of them referred to sustainability Key Performance Indicators (KPIs), such as GHG and CO₂ emission, energy efficiency, solid waste, water consumption and ecological footprint. Stakeholder engagement, sustainability reporting and auditing were largely mentioned as well, while the respondents agreed that compliance to regulations was not the right approach to tackle sustainability (Sroufe, 2017). Sustainability KPIs were extensively investigated in a recent study combining literature review and surveys, and grouped into Environmental, Social and Economic indicators, according to the Triple Bottom Line approach already mentioned (Hristov & Chirico, 2019). The authors propose

a set of goals for each line, identify specific KPIs to operationalize the strategic goals and indicate the measurement tools needed (such as questionnaire, cost analysis, financial reporting...). Other studies highlight the urgency to include into the sustainability performance assessment the context on which the company may have an impact. In other words, they suggest looking for relationships between the outputs of the corporate performance and the data available on the environment or on human beneficiary settings (Vörösmarty, et al., 2018). This context-focused approach indeed takes a step forward the SDG framework insofar as it measures the impact of an activity rather than the outputs of the activity. None of these studies, while mentioning the UN Sustainability Agenda, have attempted to propose an operationalization of the SDGs, nor have mapped the UN goals to measurable corporate performance indicators.

How is such data gathered and evaluated by sustainability assessment tools to guide decision-makers on corporate sustainability policies or investing strategies? While there is still not a standardized framework to assess corporate contributions to SDGs, several tools exist to measure Corporate Social Responsibility as a broader concept. Corporate Social Responsibility was defined by Bowen in 1953 as “*the obligations of businessmen to pursue those policies, to make those decisions, or to follow those lines of action which are desirable in terms of the objectives and values of our society*”. (Bowen, 1953 cited in Carroll, 1999). Since then, other concepts have been added to this first definition, such as stakeholder theory, business ethics theory and corporate citizenship (for a review see Carroll, 1999), that are now interchangeably used to refer to “business accountability to society” or to the “business responsibility for the wider societal good” (Crittenden, Crittenden, Ferrell, Ferrell, & Pinney, 2011). The existing CSR assessment tools can be grouped into three main categories, namely Indexes, Rankings and Ratings, and take into account different sustainability frameworks (Diez-Cañamero, Bishara,

Otegi-Olaso, Minguez, & Fernández, 2020). The term index refers to stock market indexes, which define a portfolio of investment holdings that belongs to a specific market or segment of the market (in this case determined by CSR performance).

The most famous index is the *Dow Jones Sustainability Indexes*, by S&P and RobecoSAM, and uses the ESG framework as reference. The DJSI World includes those companies that rank within the top 10% of their industries based on their Total Sustainability Score²⁰. Also the *MSCI World ESG Leaders Index* and the *STOXX Global ESG Leaders Index*, as the name suggests, refers to ESG. On the other hand, the *FTSE4Good Developed Index*, launched by FTSE Russell, assesses CSR based on the ESG framework aligned with the SDGs²¹. S&P and RobecoSAM has also issued a rating system, called the *Sustainability Yearbook*, defined according to the ESG framework, unlike the *ISS-oekom Corporate Rating*, which has declared to align with the SDGs²². Other famous ratings systems, which again are based on the ESG pillars, are *RepRisk Rating* and *Sustainalytics*, by Morningstar. Finally, among the several organizations issuing sustainability ranking systems, there are Fortune, Universum. Corporate Knights, Forbes and many others (for a review see Mooij, 2017).

While the CSR assessment systems mentioned above are based on data disclosed by companies, reports, KPIs and/or NGOs reports, it is worth mentioning that there is also a minority of private organizations currently employing machine learning algorithms and Natural Language Processing techniques to retrieve and analyze alternative data. These are *TrueValue Labs*²³, that adopts the ESG framework, *Sensefolio*²⁴ and *GlobalAI*²⁵, which

²⁰

<https://www.robecosam.com/en/media/press-releases/2019/dow-jones-sustainability-indices-review-results-2019.html#:~:text=After%20applying%20a%20orange%20of,best%2Din%2Dclass%20approach>

²¹ <https://research.ftserussell.com/products/downloads/FTSE4Good-brochure.pdf>

²² https://www.deka-etf.de/documents/iss_oekom_47_latest_de.pdf

²³ <https://truvaluelabs.com/trends/alternative-data>

²⁴ <https://sensefolio.com/about-sensefolio/>

²⁵ <https://www.globalai.co/oursolutions>

are the only organizations, together with FinScience, currently offering an ESG scoring system that measure corporate sustainability performance against the Sustainable Development Goals, by leveraging alternative data (such as social media posts and financial news).

Overall, these tools provide stakeholders, investors and corporations with the opportunity to compare sustainability performances and adopt investment strategies accordingly. On the one hand, since each tool is unique, investors can choose the CSR assessment system that best aligns with their sustainability strategies; on the other hand, corporations can acknowledge how well they do compared to other businesses in their industry and therefore be encouraged to put more efforts into sustainability strategies and ESG disclosure. From the next chapter, the discussion will move from the existing systems to the novel FinScience Sustainability Score.

2. FinScience Sustainability Score

While the first chapter proposed a critical literature review of corporate sustainability and its evaluation, the second and the third chapter will dive into the internship project that I took part in and that was aimed at building a novel sustainability score. The first part of this chapter will focus on the methodology behind the definition of the score and of its components. One of the main features of the score is in fact the dualistic approach to sustainability, as both data disclosed by the company and third-parties' perspectives are taken into account. The combination of traditional data and alternative data and the adoption of the UN SDG framework will be also presented as distinctive characteristics of the tool. In the second and third part of the chapter, I will explain the variables that make up the score and their corresponding collection process, with reference to the Python code that I wrote. The corporations that were selected to test the Score are the Standard and Poor 500 companies, the analysis of which will be presented in the last chapter.

1.1. Score definition and methodology

The construction of a novel corporate sustainability score has been the core of the project that I undertook from January to September 2020 at FinScience S.r.l. as a Data Science Intern. Finscience is a data-driven fintech company founded in 2017 by Google's former senior managers and Alternative Data experts, who have combined their digital and financial expertise. FinScience's technology applies machine learning algorithms to alternative data in order to extract relevant information for investment decisions and financial strategies, with alternative data being defined as “*non-traditional, big and often unstructured data coming from digital environments (i.e. blogs, forums, social or e-commerce platforms, maps, etc.)*”²⁶. There is general consensus that alternative data represents an attractive opportunity in investing decisions by offering a competitive

²⁶ <https://finscience.com/en/alternative-data/>

advantage over the use of traditional data only²⁷. This is why 70% of fund managers are currently monitoring alternative data or are planning to do so²⁸. Hence, the sustainability score designed by FinScience could not but leverage alternative data, while still including traditional data.

The Score was conceived as a modern tool to assess corporate contributions to the UN Sustainable Development Goals by comparing the information disclosed by the organization with the external perception of its commitment to sustainability. It was designed by a multidisciplinary team with a wide variety of competencies, as required by the nature of the problem. More specifically, the team was composed by ESG specialists, business analysts and data scientists . The Score emerges from the combination of 17 sub-scores that relate to internal data and 17 sub-scores that relate to external data. The steps taken towards the Score definition, before the actual data collection, are:

- 1) internal and external variables definition;
- 2) internal and external SDG sub-scores computation;
- 3) sub-scores discrepancy penalization;
- 4) sustainability score computation based on materiality;

I will now describe these steps in detail.

The FinScience team, in collaboration with partnered ESG specialists, defined a set of sustainability-related variables from internal and external data sources. For each variable, it was determined which SDGs are impacted (in a binary fashion: the variable either contributes to a Goal or it does not), the measurement level (continuous or binary), a weight that defines the specific importance of the variable, and whether the absence of data counts as 0 (in other words, the lack of disclosure is penalized). The final selection consists of 215 internal variables and 209 external variables. Also, for updating purposes,

²⁷ <https://www.economist.com/the-economist-explains/2016/08/22/why-investors-want-alternative-data>

²⁸ https://www.ey.com/en_gl/wealth-asset-management/when-focusing-on-the-future-where-do-you-look

other details were indicated, such as the availability of historical data and the update frequency. An extensive description of the variables will be provided in the next paragraphs and it is also available in tabular form in the Annexes.

The so-defined variables contribute to the 17 sub-scores according to their weights. More specifically, each SDG sub-score is computed as weighted mean of the variables that contribute to that Goal. Let us call $SDGI_j$ and $SDGE_j$ respectively the internal and the external score for the j -th SDG, N the number of internal variables that contribute to that SDG, M the number of external variables, and v the value taken on by the i^{th} variable. Then, the sub-scores are computed as follows:

$$SDGI_j = \sum_{i=1}^N w_{ij} v_{ij} ; \quad SDGE_j = \sum_{i=1}^M w_{ij} v_{ij}$$

The final score for a specific SDG, that I will be calling $SDGF$, is not just the average of the corresponding internal and external sub-scores. In fact, the difference between the internal and external score is part of the formula too: it contributes negatively to the sub-score computation. The introduction of a penalization term is to control for the discrepancy between what the organization communicates about its commitment on sustainability and third parties' perception of it. The idea is that, if sustainability performance is perceived as less virtuous than how the company describes it, the company might be doing SDG-washing²⁹, or other forms of deceptive marketing strategies that aim at improving the company image on ESG issues, such as social issues (social-washing)³⁰ and environmental themes (green-washing)³¹, often to increase sales³², while not acting

²⁹ <http://sdg.iisd.org/news/responsible-business-report-finds-high-risk-of-sdg-washing/>

³⁰

<https://www.bloomberg.com/news/articles/2020-04-09/social-washing-is-becoming-growing-headache-for-esg-investors>

³¹

<https://www.theguardian.com/sustainable-business/2016/aug/20/greenwashing-environmentalism-lies-companies>

³²

<https://www.nielsen.com/us/en/press-releases/2014/global-consumers-are-willing-to-put-their-money-where-their-heart-is/>

accordingly. On the other hand, if third parties' opinions are more encouraging than the company's self-disclosures on a specific SDG (the external score outperforms the internal score), it might be the case that corporate disclosure should be improved or that other factors are involved. In both cases, $SDGF$ is negatively affected. More specifically, the $SDGF$ for the j -th Goal is calculated as the weighted mean of the corresponding $SDGI_j$ and $SDGE_j$, with w being the weight assigned to the internal score, minus the absolute value of the difference between them, normalized by the $SDGE_j$ and multiplied by the penalty coefficient p :

$$SDGF_j = w \cdot SDGI_j + (1 - w) \cdot SDGE_j - p \frac{|SDGI_j - SDGE_j|}{SDGE_j}$$

Both w and p ranges between 0 and 1. As we believed that the internal and external scores should be equally contributing to the final score, we set $w = 0.5$; also, as we did not want the penalization term to weight too heavily on the final score, we set $p = 0.1$.

The step described above is repeated for each of the 17 couples of $SDGI$ and $SDGE$, to obtain 17 $SDGF$. How do these sub-scores contribute to the actual FinScience Sustainability Score? The SASB Materiality Matrix that was mentioned in the first chapter comes into play. Since the Score is meant as a tool to help investing strategies in the first place, the team decided to weigh more heavily those SDGs that are likely to have an impact on the financial performance of the company based on the Industry. Hence, each company was associated with its corresponding SASB SICS Industry, using the SICS Look-up tool³³. The $SDGF$ s covering those Goals that are material to the industry are valued as three times more impacting on the final Sustainability Score than the Goals that are not material. Let us call G_I the set of Goals material to the Industry I according to the

³³ <https://www.sasb.org/find-your-industry/>

SASB materiality matrix and R_I the remaining Goals. Then, the FinScience Sustainability Score for a company c belonging to the Industry I is computed as follows:

$$FSC_{(c,I)} = \frac{1}{3|G_I|+|R_I|} \cdot \left(\sum_{i \in G_I} 3 \cdot SDGF_i + \sum_{j \in R_I} SDGF_j \right)$$

where $|G_I|$ and $|R_I|$ denotes the number of Goals in the sets.

Having defined such a backbone, there is still an important piece missing: the actual data. Several data science tools were employed to implement the data extraction processes and to manage big data storage. All code was written in Python 3.6. A large amount of data was collected through web scraping, using Scrapy and Selenium libraries. Scrapy³⁴ is a popular open-source framework specifically designed for handling web scraping. It allows link extraction from websites, collection of tags and other HTML elements using CSS selectors and XPath expressions, and implementation of pipelines to be executed sequentially. Most importantly, it permits to do so by providing a series of distinct logical components which can be added to the main structure of the scraper. Because of such structure, Scrapy is particularly useful when it comes to recursively scraping web pages based on specific criteria, and it also makes possible to set the search strategy (either breadth first search or depth first search) and the crawl depth (the number of clicks needed to get to a page from the home page). Selenium³⁵ is an open-source web-based automation tool that supports Python. It is mainly used to automating tests on web applications, but it also offers multiple functions for web automation in general. For instance, it allows automatic interaction with input user elements, such as forms to be filled in. This came particularly handy during the data collection, as I will detail in the next paragraphs, since several websites offer the possibility to query their datasets by filling specific forms.

³⁴ <https://pypi.org/project/Scrapy/>

³⁵ <https://selenium-python.readthedocs.io/getting-started.html>

To manage and store the data, we adopted the Google Cloud Platform (GCP)³⁶, which offers several cloud computing services for data management, machine learning and artificial intelligence on the same infrastructure as Google. For file storage, we employed Google Cloud Storage³⁷, an online file storage web service for storing and accessing data on GCP infrastructure. The service combines the performance and scalability of Google's cloud with advanced security and sharing capabilities. Unlike Google Drive, it is more suitable for enterprises and organizations. Google Cloud Storage was mainly used to collect corporate sustainability reports in PDF format. To manage and query tabular data, Google BigQuery³⁸ was largely used. It allows the creation of tables based on a json-encoded schema, as well as the import of data encoded as csv or json. The creation of tables in BigQuery was implemented in the Scrapy projects pipelines together with the row insertion, as fundamental steps following from extraction of data from websites. At the end of the collection process, the data appeared organized in separate tables belonging to the same project. Another component of GCP infrastructure that was employed is the Google Compute Engine³⁹, which enables users to launch virtual machines on demand. Thanks to the capacities of such machines, this product was specifically appropriate for developing and running high-memory consuming Python scripts more efficiently.

The next paragraphs will focus on how I implemented the data extraction pipeline from almost 400 data sources on the 500 companies that constitute the S&P 500 index, namely the largest American companies, as in January 2020.

1.2. Internal Data

To build the SDG internal sub-scores (*SDGIs*) that rate the corporate sustainability performance according to the data disclosed by the company itself, 215 variables (or

³⁶ <https://cloud.google.com/gcp/>

³⁷ <https://cloud.google.com/storage>

³⁸ <https://cloud.google.com/bigquery>

³⁹ <https://cloud.google.com/compute>

metrics) were obtained from 4 different types of data sources: company website, CSR/sustainability report and report-related websites, memberships websites, balance sheet. As the bar chart above shows, most of the variables are extracted from the balance sheets, followed by the memberships, the report and lastly the corporate website (*Figure 2*).

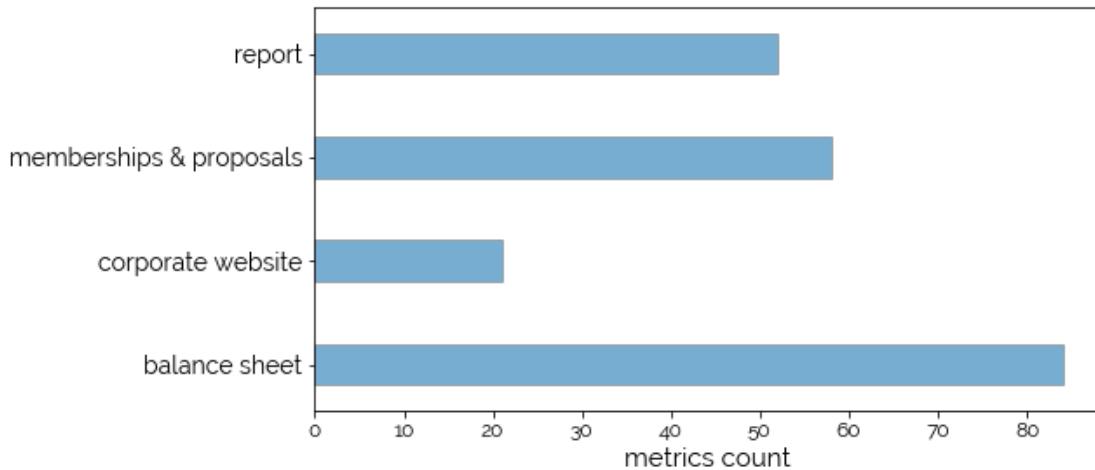


Figure 2: total metrics count by data source type

1.2.1. Corporate Website

To understand which sustainability issues companies tackle on the Internet, how relevant they consider SDGs and ESG themes, and, generally speaking, what is their online approach to sustainability development, we decided to explore the official company's website and associate 21 variables to it. To scrape the web pages, I created a Scrapy project inspired by the work carried out by Sozzi in her MSc final thesis, in which she scraped 100 corporate websites to extract SDG-related textual data⁴⁰ (Sozzi, 2014). As previously mentioned, Scrapy is a fast high-level web crawling and web scraping framework, providing specific functions for extracting structured data. It can be used for a wide range of purposes, from data mining to monitoring and automated testing. For each company, the crawler:

⁴⁰ <https://github.com/ONSBigData/Measuring-Sustainability-Reporting>

- 1) visits the company website homepage;
- 2) looks for SDG-related keywords in the textual content;
- 3) stores the textual content of the whole webpage if containing at least 3 keywords and extract the corresponding section;
- 4) extracts the links from the page;
- 5) visits the links if specific conditions are met;
- 6) repeats the process for each visited link from step 1) as if they were the homepage;
- 7) stops after 5 repetitions (*crawl depth*).

The links to the 500 companies' websites were retrieved from Bloomberg databases. Step 2) is based on a large open-source SDGs vocabulary available at Zenodo website, the open-access repository developed under the European OpenAIRE program and operated by CERN (Duran-Silva, Fuster, Masucci, & Quinquillà, 2019). As the SDG 17 was not covered by the original vocabulary, we added the corresponding keywords to it under the supervision of the FinScience ESG specialists. More than 100 keywords are associated with each Goal. The text extraction was performed using the jusText Python library⁴¹, which is suitable for removing boilerplate content from the HTML page, such as headers, footers and links. Both the keywords and the webpage content are normalized by lowering and stemming each word with the nltk package⁴², in order to control for plural forms and other inflected forms. If the condition at step 3) is met, the original text is then stored, in a Google Big Query table, together with the keywords found, the corresponding url and also the last keyword found in the url or in the url description. Such keyword determines the “section”. As for the extracted links, they are visited if the following conditions are met: the url or the description contains at least 1 keyword; the url domain is the same as the

⁴¹ <https://pypi.org/project/jusText/>

⁴² <https://www.nltk.org/>

original one or it contains specific words, i.e. “*investor*”, “*sustain*”, “*csr*”, “*environ*” (to avoid visiting external links but other relevant webpages); the url does not contain words such as “*news*”, “*contact*”, “*privacy*”. What follows is a row of the so-obtained table (*Table 1*).

timestamp	ticker	company	url	section	content	keywords
2020-02-16 13:18:31.090 168	MS	Morgan Stanley	https://www.morganstanley.com/Themes/climate-week-sustainability-impact-investing	sustainability	The New Investor Response to Climate Change Sep 18, 2019 Climate change issues are as wide-ranging as ever, from extreme weather, to water scarcity, to infectious diseases. Investors are heeding urgent challenges with a growing interest in funding solutions that offer real impact—and real returns.	disease water climate change

Table 1: example of data extracted from corporate website using web scraping

How is this information operationalized into quantitative indicators? We defined four distinct dichotomous variables relating to the section and based on all pages collected for each company: presence of a sustainability section; presence of a corporate governance section; presence of a remuneration section; presence of a diversity section. For instance, the company in *Table 1* is considered as having a sustainability section, as the webpage stored in Table 1 shows. In addition to these metrics, we also created 17 continuous variables relating to the actual content of each webpage, one for each Goal, and originating from a text classification process performed on each web page, as follows.

SDG Text Classifier. We built a supervised text classifier for assigning a Goal to pieces of text. This was done to better capture the SDG being addressed in webpages by taking as reference labelled texts rather than the keywords only, which were instead used as a filter to initially select relevant web pages. We first collected 4539 pieces of texts from websites specialized in SDG-related issues and manually labelled each of them with one Goal, as we wanted to build a single-label classification model. The so-obtained classes, which correspond to the 17 Goals, were unbalanced as the following chart shows (*Figure 3*).

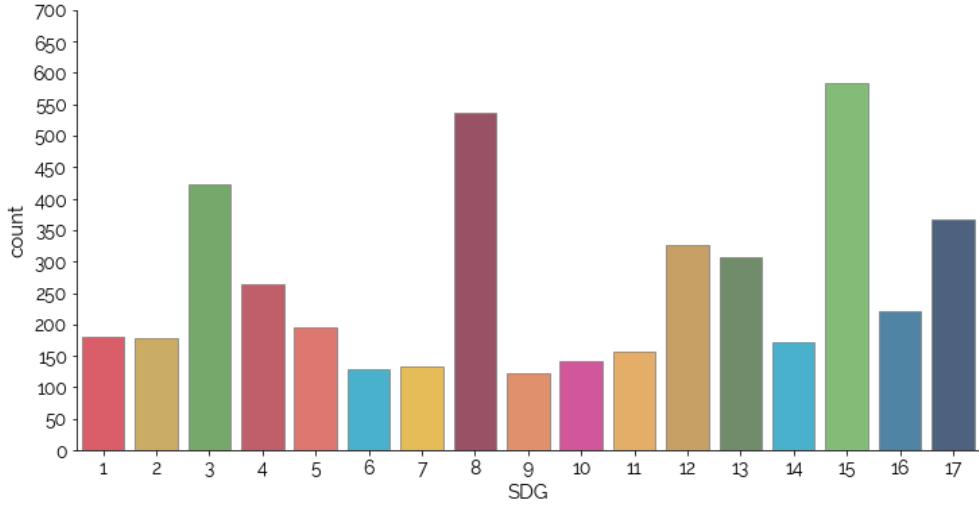


Figure 3: Goals distribution over SDG-labelled dataset for text classification

After cleaning the texts by lowering, stemming, removing stopwords and punctuation, the dataset was split into a training set (70% of the data) and a test set (30% of the data) using the `StratifiedShuffleSplit` function provided by the `scikit-learn`⁴³ library, in order to obtain similar class distributions in both sets. Several models were trained and tested, namely traditional classification methods (logistic regression, naïve Bayes, SVM) with TF-IDF representation and neural networks with word embeddings (word2vec and Glove). TF-IDF stands for Term Frequency — Inverse Document Frequency.

TF-IDF is a popular technique in text mining to measure the importance of a word to a document in a collection of documents. It is the acronym of two terms: Term Frequency and Inverse Document Frequency (Goyal, Pandey, & Jain, 2018). TF is the count of a specific word to the total number of words in a document, while IDF is the log ratio of the total number of documents to a document containing that word. That is to say, TF-IDF measures how salient a word is by controlling for its frequency in both each document and the whole collection of documents. In text classification, it allows to identify the most representative words for each class, which correspond to the most frequent words in the documents labelled with that class.

⁴³ https://scikit-learn.org/stable/model_selection.html#model-selection

To encode each document with its TF-IDF corresponding representation, I used the `TfidfVectorizer`⁴⁴ function of the scikit-learn package, which converts a collection of raw documents to a matrix of TF-IDF features, having as many rows as the number of documents and as columns the unique words of the collection. The `ngram_range` was set from 1 to 3 (which means that not only single words but also combinations of two and three consecutive words are taken into account) and the stopwords were excluded from the matrix. The algorithm is fit on the training set to build the vocabulary, that determines the number of columns of the matrix, and used to transform both the training and the test set.

After the transformation, the training set was passed to the models. Logistic regression, Naïve Bayes and Support Vector Machines algorithms were tuned on the training set using a 5-fold cross-validation, by taking F1-score as an evaluation metric. F1-score is the weighted average of the precision and recall values and was chosen in this context as it has been shown to be a solid evaluation metric when dealing with unbalanced datasets (Jeni, Cohn, & De La Torre, 2013).

In addition to the traditional classification models, neural network models were trained. Recently, word embeddings have become extremely popular and efficient in different NLP tasks. Unlike the TF-IDF vector representation, in which each feature is coded as a unique dimension in a one-hot representation, and documents are coded as sparse vectors, in word embedding each word is represented by a dense vector, based on a finite vocabulary, such that words with similar meanings are close to each other in the vector space (Huang, Qiu, & Huang, 2014). Different methods can be employed to extract such representation from text data, such as a neural network with an embedding layer, word2vec method and GloVe algorithm.

⁴⁴ https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

To build a neural network with an embedding layer, I used keras library⁴⁵. It requires that the input data is encoded so that each word is represented by a unique integer; this is done using the tokenizer module. The embedding layer, which is the first hidden layer of the network, takes as input the tokenized training set and uses it to learn the word embeddings. Again, the model was validated using a 5-folds cross validation. Word2vec (Mikolov, Chen, Corrado, & Dea, 2013) and GloVe (Pennington, Socher, & Manning, 2014) provide pre-trained models on large datasets, and therefore they do not require to be re-trained. The training set needs, instead, to be tokenized and transformed according to such pretrained models⁴⁶, in order to obtain a matrix representation, which is then passed to the hidden layer of the new neural net, as explained before. The table below summarizes the models' performances. As it can be seen, the Logistic Regression based on the TF-IDF representation is the best performing model on the test set, while the neural networks, as the results on the training set suggest, would overfit the training set (*Table 2*).

	model	Training set			Test set		
		accuracy	precision	F1 score	accuracy	precision	F1 score
Neural nets	Simple embedding	96.27%	97.36%	96.27%	81.98%	83.13%	82.32%
	Word2Vec	99.42%	99.47%	99.45%	73.42%	74.59%	73.68%
	GloVe	94.34%	94.86%	94.35%	86.86%	87.76%	86.99%
Linear Models	Logistic Regression	90.54%	91.34%	89.41%	90.99%	92.67%	89.92%
	Naïve Bayes	81.79%	83.61%	79.69%	82.28%	83.27%	80.38%
	Support Vector Machines	88.96%	91.61%	87.73%	90.76%	91.23%	89.46%

Table 2: classification models performance on training and test set

⁴⁵ <https://keras.io/>

⁴⁶ I was inspired by <https://realpython.com/python-keras-text-classification/>

Going back to the variables associated with the website, the results of the so-obtained classification model on the corporate websites were organized into 17 continuous variables, one for each Goal. More specifically, for each company, each of these variables represents the proportion of pages being assigned to that Goal by the model over the total web pages retrieved for that company. To sum up, 22 variables relate to the corporate website and the corresponding data are extracted through web scraping and Natural language processing techniques, and more specifically, with keywords research and supervised text classification (for a more extensive description of the variables mentioned, see Annex B2).

1.2.2. CSR/Sustainability Report

As shown in the first chapter, more and more corporations have been issuing sustainability and corporate social responsibility reports. Some researchers have analysed CSR reports using text mining and Natural Language Processing techniques to explore what sustainability themes are tackled (Liew, Adhitya, & Srinivasan, 2014; Székely & vom Brocke, 2017). In this context, we decided to analyse sustainability/CSR/corporate citizenship/environmental/corporate responsibility reports, which represent a fundamental source to understand the approach of companies to sustainability. More specifically, 51 variables relating to the sustainability report were formulated.

Reports collection. First of all, we wanted to investigate whether companies issue CSR reports (where, by CSR report, we mean also Sustainability report, Corporate Citizenship report, Corporate Responsibility report, ESG report, Integrated report, while annual reports are excluded) and, if so, since when they started to do so. Such information is provided by Corporate Register, “*a global online directory of corporate responsibility (CR) reports past and present*”⁴⁷. The website comes with a search bar that allows the user

⁴⁷ <https://www.corporateregister.com/>

to look up corporations and returns a list of the corresponding issued reports, if any. The second step is to look for the PDF file of the latest report issued by each company, based on such information. To find the corresponding file, Google Custom Search API⁴⁸ was employed and embedded in a new Scrapy project. For each company, the crawler:

- 1) looks up the company name;
- 2) stores in a Big Query table the data about each report issued by the company (report name, report type, report year);
- 3) looks up the name of the latest report, if any, using Google Custom Search API, which returns the first 100 results;
- 4) clicks the first link to a PDF file and downloads it to a Google Cloud Storage directory.

Based on such information, it is possible to state whether a company has ever issued a CSR report (dichotomous variable) and for how many years it has been reporting on sustainability issues (discrete/numerical variable).

Corporate Register is not the only data source considered to extract information about CSR reports. We also referred to GRI database on the organizations adopting a GRI Standards Report Registration System⁴⁹. Companies are looked up in the csv downloadable version of the dataset using a Python script. Similarly, the International Integrated Reporting Council (IIRC) provides an online database that contains a list of companies that have either produced a report that references the IIRC and/or the <IR> Framework or are influenced by the Framework through participation in <IR> Networks⁵⁰.

A Scrapy project was built to scrape the dataset and look up companies. GRI and IIRC data sources contribute to two dichotomous variables.

⁴⁸ <https://developers.google.com/custom-search/v1>

⁴⁹ <https://www.globalreporting.org/reportregistration/verifiedreports/>

⁵⁰ <http://examples.integratedreporting.org/organisations>

Certifications & Assurances. In Europe, organizations with more than 500 employees are required to report on ESG data (Directive 2014/95/EU⁵¹). Although it is not currently compulsory for such companies to consult an external assurance provider on this matter, the directive is now about to be modified. In this regard, during the consultation on non-financial reporting directive in 2020, “*67% of respondents believe that the EU should impose stronger audit requirements for non-financial information. If the EU were to introduce stronger audit requirements, respondents were evenly divided as to whether the requirement should be for limited assurance or reasonable assurance*” (Summary Report of the Public Consultation on the Review of the Non-Financial Reporting Directive, 2020, pag.3). In the United States, there are no specific regulations on external assurances, but more and more companies decide to voluntarily report on ESG matters. However, an external assurance is not the only way for companies to improve their reliability. In fact, the International Organization for Standardization’s certifications contribute to give more credibility to the reporting activity too and to express an explicit commitment to SDGs⁵².

For these reasons, both certifications and assurances are looked up in the reports, and this is done using regex expressions. More specifically, a list of ISO certifications was provided by the ISO sustainability team, with each of the Standard being mapped to the SDGs that it contributes to. A list of assurances-related keywords was provided by the FinScience ESG specialists. Each of the ISO standards correspond to a dichotomous variable (the certification is either mentioned or not in the report), while the assurances keywords converge into a single dichotomous variable, which takes value 1 if at least one keyword is mentioned. Since it may be the case that the organization mention assurances to claim that their reports have not received a third-party assurance, a sentiment analysis on the sentences containing assurance-related keywords was performed using the

⁵¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095>

⁵² <https://www.iso.org/sdgs.html>

SentimentAnalyzer⁵³ module provided by the nltk package and the company was considered as being externally assured if the sentiment was positive.

SDG text classification. We decided to perform text classifications on CSR/Sustainability reports as we did on corporate websites to identify the SDGs being addressed. A more detailed description of the SDG classification model is provided in the paragraph 2.2.1. For each company, the report retrieved as described previously is treated as follows. First, the report is divided into paragraphs. Then, with means of regex expressions, the SDG-keywords that are used for scraping and analysing the corporate website are also looked up in each paragraph. Eventually, each paragraph, if it contains at least 3 keywords, is classified using the SDG Classifier. Such classification results in 17 continuous variables, each of which expressing the ratio of paragraphs covering a specific Goal with respect to the total number of classified paragraphs (see Annex B1 for an exhaustive description of the variables).

1.2.3. Memberships & Shareholder Proposals

Organizations can commit to sustainable development and social responsibility by being a supporter/member/associate/signatory of associations, coalitions and initiatives that are sustainability oriented. After identifying more than 40 associations and their corresponding website and mapping them to the Goals that they contribute to, three types of approach are taken. For those websites that provide a list of their members, I built a Scrapy project that looks up companies in the HTML of each page. By looking at the whole HTML, instead of just the textual content, it is possible to detect a company's name even though it is mentioned in hidden tags, such as pictures captions (which might be the case of companies' logos). Both the HTML and the companies name are lowered;

⁵³ <https://www.nltk.org/api/nltk.sentiment.html>

companies' names are also stripped of acronyms and abbreviations such as "*corp*", "*ltd*", "*inc*" to facilitate the research.

In addition to this, we also considered two other data sources which are not properly memberships: CDP and Ceres. CDP, formerly Carbon Disclosure Project, provides a full set of publicly available climate change, forests and water security scores⁵⁴. In this context, we only considered whether or not the company decided to disclose its data on these issues, while, for the external scores, we will also consider the rating given by CDP to the disclosed performance. Ceres is a sustainability non-profit organization that provides a database of all shareholder proposals on proxy statements (a statement required of a firm when soliciting shareholder votes) relating to environmental, social, and sustainable governance (ESG) issues beginning with 2015⁵⁵. A novel Scrapy project was built to scrape the online database. For each company, the scraper looks up the company's ticker and stores the data into a Google Big Query. We grouped the proposals into 7 categories, i.e. human rights, environment, ethical finance, health, decent work, workplace diversity and sustainability, each of which corresponds to a dichotomous variable that takes value 1 if the company has been involved in a shareholder proposal of that type (see Annex B3 for an exhaustive description of the variables).

1.2.4. Balance Sheet

In addition to the alternative data described above, the SDG score could not neglect sustainability balance sheets, which contain traditional data on corporate sustainability performance. This data is provided by Bloomberg in csv format, and therefore no specific operation is needed to collect it. The variables that are collected regard the ESG corporate performance according to the traditional sustainability KPIs. Among the Environmental Indicators, we consider: total water use, total waste, waste recycled, % water recycled,

⁵⁴ <https://www.cdp.net/en/companies/companies-scores#446647786929955804cc9a3a08ef1eb4>

⁵⁵ https://engagements.ceres.org/?_ga=2.267874958.1878011541.1601282680-1884343248.1601282680

discharges to water, hazardous waste (tonnes). Among the Social indicators, we look at: ratio of women in total workforce to women in management, percentage of women employees, percentage of employee turnover, the lost time incident rate, total incident rate, fatalities per 1000 employees. Among the Governance indicators: % independent directors, independent chairperson, independent lead director, % women in the board, female chairperson or equivalent, director average age, % director meeting attendance. All data is eventually rescaled to a 0-1 range (for a complete list of the indicators, see Annex B4).

1.3. External Data

External data makes up the SDG external scores, or *SDGE*, which quantify how the corporate sustainability performance is perceived and talked about by third parties, such as NGOs, sustainability-committed institutions, employees, sustainability agencies and rankings. More specifically, 244 external variables were identified, and can be grouped into 5 main categories: rankings, indices, controversial activities, reviews and FinScience news. As the figure above shows, rankings and FinScience news represent the main external data sources (*Figure 4*).

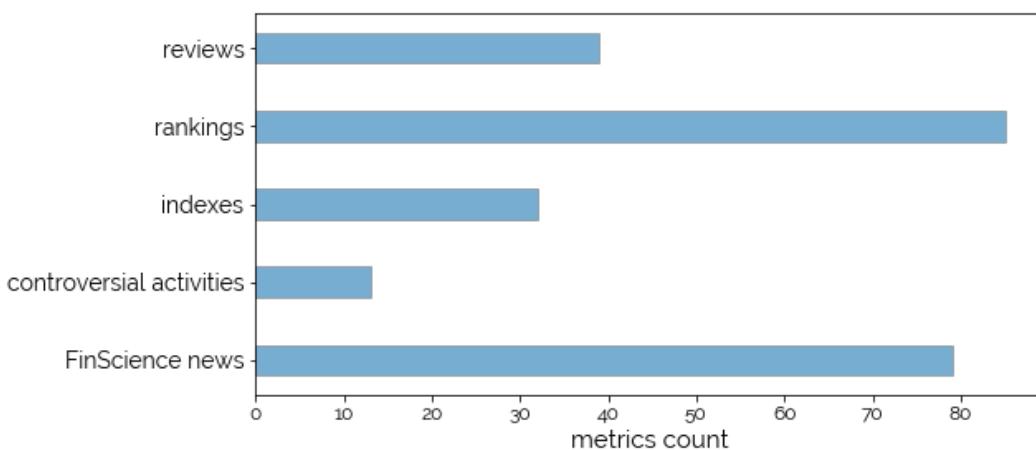


Figure 4: total external metrics count by data source type

1.3.1. FinScience news

To figure the external “perception” of corporate sustainability performance, we could not but look at news shared online. FinScience retrieves news content from the web and elaborates it using NLP tools every day. About 1.5 million web pages are visited on a daily basis on about 35.000 different domains. The content of these pages is extracted, interpreted and analysed to identify valuable information and sources. The FinScience’s news data pipeline can be divided into 3 steps⁵⁶. First, data gathering involves the collection of data from different web sources: websites, social network pages, news or blogs. The latter are identified mainly by following two different criteria that are not necessarily connected to each other: the level of sharing and visibility of a content and the identification of specific sources linked to a particular topic of interest. Then, the contents are extracted from the web pages and are pre-processed via a first level of data cleaning for the elimination of noise and the extraction of the main body: this is the input to the next phase of data processing. At the data processing stage, Natural Language Processing techniques are applied. The contents collected in the data gathering phase are subjected to an NLP analysis that allows to determine the objects (companies and SDG topics...) disseminated and discussed on the web. Once the topics covered have been identified, the data are analysed, normalized and enriched to obtain further metrics such as:

- 1) Digital popularity Value (DPV), a measure of the diffusion of a digital signal on the web. It is obtained by aggregating the diffusion metrics of the news mentioning the signal at hand and can take only positive values.
- 2) Sentiment: it measures how users feel about a specific company or information and can take values in the interval (-1, 1). This indicator seeks to quantify how current

⁵⁶

https://www.linkedin.com/posts/finscience-com_finscience-esg-methodology-activity-6716318325699317760-RTop

beliefs and positions affect future behaviours. The sentiment analysis algorithm is based on a semi-supervised method. The sentiment value - weighted by DPV - is finally assigned to the company in relation to each SDG to reflect the public perception of corporate sustainability efforts/impact.

For the scope of the present work, news collected are filtered based on three categories of data sources: vertical news, NGOs' news and mainstream news. Vertical news were extracted from a list of websites specialized in sustainability topics, and were organized into 17 indicators, one for each SDG; NGO's news were extracted from a list of 45 non-governmental organizations' websites, and therefore a single indicator was identified for each NGO; mainstream news refer to the remaining type of sources that mention the monitored companies together with at least one SDG-related theme, hence they contribute to 17 separate indicators, one for each SDG, in a way similar to vertical news (for an extensive description of the sources see Annex C1).

1.3.2. Rankings and Indexes

A list of 85 initiatives/organizations that provide sustainability-related rankings with their corresponding website was considered. Six of such websites offer PDF files where the members of the ranking are listed. In these cases, PDF files are downloaded, and then the companies' names, lowered and stripped, are looked up in the text. However, most of the rankings are available online and therefore a Scrapy project was built. The scraper simply looks up the companies' name in the whole HTML of each ranking page, as with the membership scraper described previously (for an extensive description of the metrics and the rankings monitored, see Annex C2). Similarly, a list of 13 sustainability indexes and their corresponding websites is taken into account too: 5 of them provide data into csv files, 2 provide PDF files, while for the remaining 6 sources, the index constituents are listed

on the webpages (for an extensive description of the metrics and the indexes monitored, see Annex C3). The same procedure described above is applied here.

1.3.3. Controversial Activities

Negative screening, as mentioned previously, is an investing strategy that looks at controversial activities to exclude industries or organizations from their portfolios. Hence, in addition to sustainability performance indicators, we included 7 negative rankings and 8 data sources on contentious issues, such as the use of hazardous chemicals and committed violations. Negative rankings are treated as described in the previous paragraphs. As for the other data sources, three of them provide data in csv format, while the remaining ones consist of databases, lists or texts available online. For the latter group, different steps are taken based on the website structure.

Among those, Good Jobs First provides an online database, called Violation Tracker, on corporate misconduct over banking, consumer protection, false claims, environmental, wage and hour, safety, discrimination, price-fixing, and other cases resolved by federal regulatory agencies⁵⁷. Since it is possible to query the database by company name, I built a Scrapy project that includes Selenium modules. For each company, the scraper fills the designed form area with the company name and stores the results in a Google Big Query table as soon as they are loaded. The information on the violations committed is mapped to 4 distinct variables, one for each violation type (environmental, consumer, governance and labour). The value taken by such variables is based on the ratio between the sanction amount and the company's market cap, to better compare the penalties. The remaining data sources, that relates to tobacco production and use of conflict minerals, were treated as rankings, by searching the companies' names in the lists provided by the corresponding websites and assigning 0 to the companies

⁵⁷ <https://www.goodjobsfirst.org/violation-tracker>

mentioned (in order to negatively affect the score). For an extensive description of the variables associated with controversial activities, see Annex C4.

1.3.4. Reviews

External scores could not ignore employees and customers' perspective and opinion and therefore we considered 4 of the most popular reviews websites: Glassdoor, Indeed, Payscale and Trustpilot. There is an increasing amount of research on such platforms as valuable data sources to monitor employee satisfaction. Glassdoor is one of the most popular review websites where current and former employees anonymously review companies. It has been claimed that, using Glassdoor data, "*organizations can effectively crowdsource their evaluations of leadership, looking at the link between employees' ratings and company performance*" (Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016, pag. 631). Luo and colleagues gave substance to this claim by proposing the first study leveraging text mining to explore the relation between employee satisfaction and company performance based on Glassdoor reviews, and found a positive correlation between general employee satisfaction and corporate performance (Luo, Zhou, & Shon, 2016). Another study focused instead on Glassdoor employer ratings on 2019 Fortune 500 companies and revealed a positive correlation to whether employees recommend their employer to their friends (Suen, Hung, & Tseng, 2020). Indeed.com is not only a company review website but also a job-search engine, the largest for number of unique users⁵⁸. Several studies were carried out to extract insights from Indeed job postings, focusing in particular on the skills required for specific job profiles, such as data scientists (Ho, Nguyen, Pafford, & Slater, 2019) and business analytics practitioners (Johnson, Albizri, & Jain, 2020). Less research is available on data collected from Payscale, but it is recognized as a valuable tool to retrieve information on average

⁵⁸ <https://techcrunch.com/2010/11/17/indeed-monster-largest-job-site/>

expected salary and, as such, a good database for graduate students⁵⁹. Finally, Trustpilot, a Danish consumer review website with almost million new reviews posted each month, was crawled to monitor customer satisfaction about UK retail energy suppliers (Littlechild, 2020) and consumers' fashion customization experiences (Lang, Xia, & Liu, 2020). I will now describe how each of these websites were treated.

Glassdoor database associates to each company an identification code, that appears in all the Glassdoor links about that company together with the corresponding Glassdoor name (for instance, Johnson & Johnson Company's Glassdoor ID is -EI_IE364.11,30, while its Glassdoor name is *Johnson-and-Johnson*⁶⁰). A Scrapy project was built to collect Glassdoor identification numbers of S&P 500 companies in the first place: the scraper looks up the companies' names on the Glassdoor bar research and stores them together with the corresponding Glassdoor company name. Then, the identifications codes are used to retrieve the Overview page and Reviews page and collect meaningful data. More specifically, Glassdoor provides ratings and pledges. Ratings relate to several aspects, such as Company culture, Work satisfaction, Management evaluation and Compensation, and are computed as the mean of users' ratings, on a scale from 1 to 5. Pledges express the company's commitment to specific issues or programmes, such as "Commitment to Diversity" and "Career Advancement Program". For each company, the scraper collects all ratings and pledges, which in turn correspond, in order, to continuous variables ranging from 0 to 1 and dichotomous variables taking value 0 or 1.

A different project was carried out to collect data from Indeed.com. As Glassdoor, it provides several rating items, which are specifically focused on work happiness. A Scrapy project was built to collect such ratings, which are then rescaled to a 0-1 range. On

⁵⁹

<https://www.nytimes.com/2013/09/14/business/economy/nice-college-but-whatll-i-make-when-i-graduate.html>

⁶⁰ https://www.glassdoor.com/Overview/Working-at-Johnson-and-Johnson-EI_IE364.11,30.htm,
<https://www.glassdoor.com/Reviews/Johnson-and-Johnson-Reviews-E364.htm>

the other hand, Payscale focuses on compensation and salary and, unlike the other review websites, it also offers a workplace gender breakdown, by showing the employment gender ratio and the minimum and maximum salary by gender. Ratings are also available, such as “Pay Transparency” and “Appreciation”. A Scrapy project was built to collect both ratings and gender-related data, which is then rescaled to a 0-1 range. Both Indeed and Payscale require the company to be searched by name, which is simply added to the main website link. To conclude, we also scraped Trustpilot, which provides reviews on companies’ products. A script mainly using Selenium looks up the companies by their website and collects the average reviews rating, that ranges from 1 to 5, which is then scaled to a 0-1 range.

3. Data analysis and findings on S&P 500 companies

In this last chapter, I will describe the analysis carried out on the internal and external data collected for the companies that constitute the Standard and Poor 500 index, namely the largest American corporations, by diving into trends and patterns emerging from the most salient variables. I will then move on presenting and analysing the scores obtained by combining such data, both the SDG sub-score and the final Sustainability score. How do S&P 500 companies deal with sustainability issues? What are the Sustainable Development Goals that they contribute to the most? Are there any differences between industries? These are some of the questions that I will aim to answer in this section, before drawing final conclusions on all the steps that have been taken so far.

2.1. Internal Data

Internal data sources were selected so as to assess corporate performance sustainability based on the data that the company discloses, through its website, reports and balance sheets. The S&P 500 companies on which data was collected are unevenly distributed over 11 GICS Sectors, with the Information Technology and the Industrials sectors being the most represented and the Energy and Communication Services the least represented (*Figure 5*)

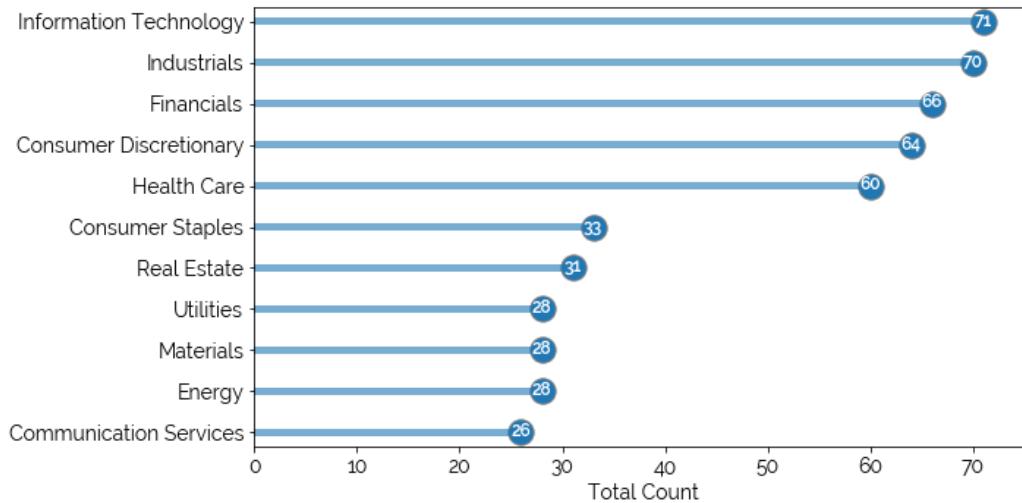


Figure 5: S&P500 companies count by sector

Corporate website. The algorithm that was designed to scrape the corporate websites managed to visit 447 websites out of 500. The remaining websites were not active or were protected against web scraping. Hence, the following results relate to the companies whose website could be scraped. In total, 5917 web pages were visited, as they contained salient keywords, according to the criteria described in the previous chapter. More than half of the companies have less than 10 web pages relating to SDG-related issues; on average, 14 webpages per company were found (*Figure 6*).

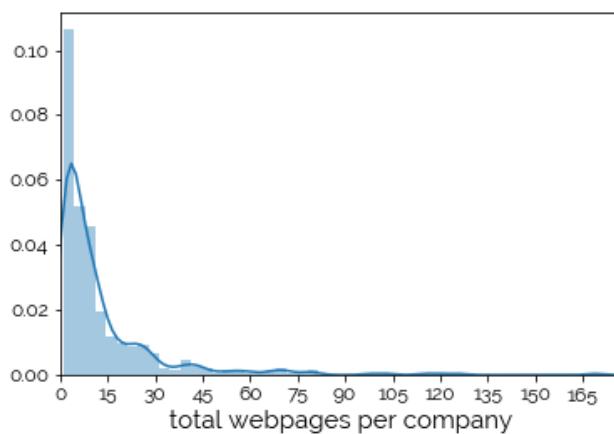


Figure 6: distribution of count of web pages retrieved per company

The most popular section, among the ones that we decided to monitor, is corporate governance (60% of the companies), followed by sustainability (56%) and diversity

(54%). On the other hand, only one company results in having a specific SDG page, with the acronym “sdg” being mentioned in its link (*Figure 7*).

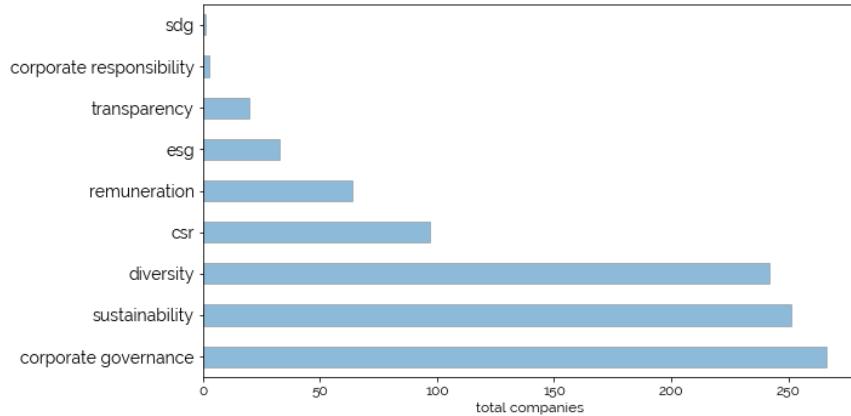


Figure 7: total count of sections by type

The most frequent SDG-related keyword of our list is *innovation*, followed by *sustainability*, *water*, *partnership* and *diversity* (*Figure 8*). As the wordcloud shows, most of the mentioned keywords relate to environment and social responsibilities (*Figure 9*).

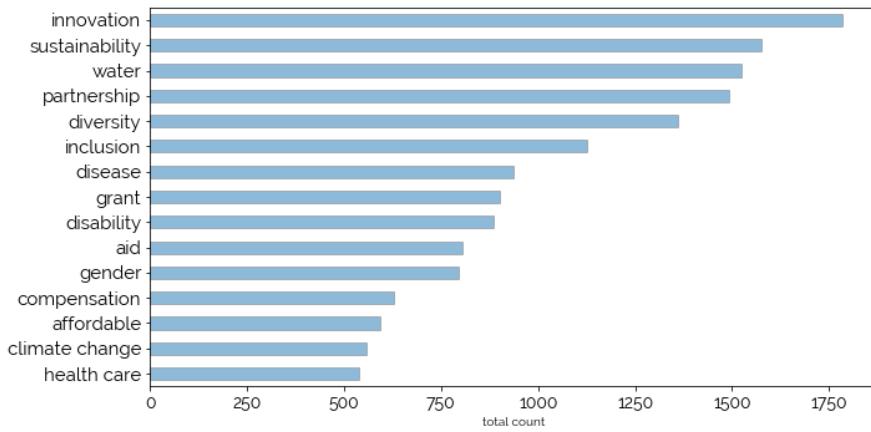


Figure 8: total count of the 15 most popular SDG-related keywords in corporate websites



Figure 9: most frequent SDG-related keywords in corporate websites

As for the classification performed on the website pages, the most covered SDG is *SDG 12 – Responsible Consumption and Production*, followed by *SDG 16 – Peace, Justice and Strong Institutions* and *SDG 4 – Quality Education*. The least addressed Goals are *SDG 17 – Partnerships for the Goals*, *SDG 1 – No Poverty* and *SDG 14 – Life below Water*. The table below shows the average presence of SGs computed as the average proportion of webpages addressing each SDG to the total number of webpages retrieved. For instance, 25% of the web pages covering SDG-related issues address the SDG 12 (*Table 3*).

	NAME	AVERAGE PRESENCE
SDG 12	Responsible Consumption and Production	25%
SDG 16	Peace, Justice and Strong institutions	16%
SDG 4	Quality Education	14%
SDG 7	Affordable & Clean Energy	7%
SDG 10	Reduced Inequalities	7%
SDG 3	Good Health & Well-being	6%
SDG 8	Decent work & Economic growth	6%
SDG 15	Life on Land	3%
SDG 9	Industry, Innovation & Infrastructure	3%
SDG 2	Zero hunger	2%
SDG 6	Clean Water & Sanitation	2%

SDG 13	Climate Action	1%
SDG 11	Sustainable Cities & Communities	<1%
SDG 5	Gender equality	<1%
SDG 14	Life below Water	<1%
SDG 1	No poverty	<1%
SDG 17	Partnerships for the Goals	<1%

Table 3: SDGs average presence in corporate websites ordered by frequency

Overall, these results suggest that most of the S&P 500 companies cover sustainability topics online, although there are topics that seem to be more addressed than others, as the keywords found, the extracted sections and the classification outputs demonstrate. More specifically, innovation and sustainability are often mentioned, with entire web pages being dedicated to sustainability. Classification has revealed a strong interest towards responsible consumption issues, which is about the resources being used in production activities, and therefore can be interpreted as a reasonable outcome.

CSR/Sustainability Report. The automatized research and extraction of sustainability reports resulted in a total of 447 companies (89%) having issued at least a sustainability report in the last years. After several manual checks, we concluded that, to our knowledge, the remaining companies have never issued a sustainability report. This result is consistent with the study carried out by the Governance and Accountability Institute, according to which 90% of S&P 500 companies published a sustainability report in 2019⁶¹. For each company, the most recent relevant report was downloaded. However, after several manual checks on the files extracted by the scraper, we were able to analyse 437 reports, because some reports were not available online (“reports not found”) or the files were damaged (*Figure 10*).

⁶¹ <https://www.ga-institute.com/research-reports/flash-reports/2020-sp-500-flash-report.html>

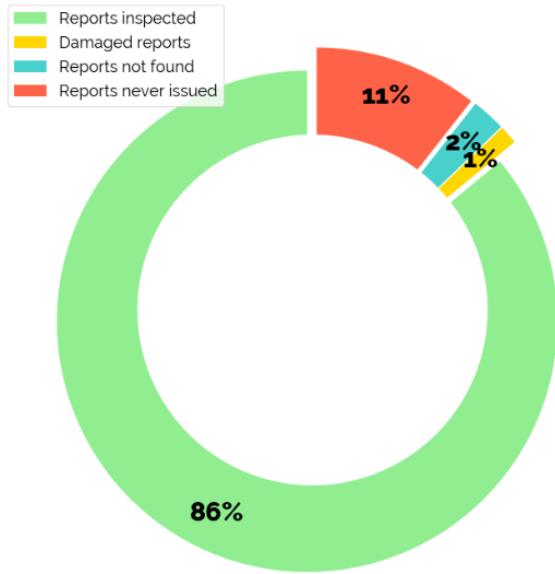


Figure 10: results of the reports files research

By scraping the reporting history, we also found that 353 have published at least one report in the last 3 years (between 2016 and 2019) but only 26 have been consecutively publishing a sustainability report in the last 3 years. As to the nature of the CSR/sustainability reports, we also found that only 21 companies out of 500 have published the report as an integrated report (that combines financial and non-financial data), which often refers to the reporting process based on the well-known International <IR> Framework⁶². The NLP analysis on the retrieved reports and, more specifically, the search for assurances-related keywords revealed that only 203 CSR/sustainability reports were assured by an external auditor. This does not come as a surprise, as external assurances on sustainability reports are not mandatorily required. The most frequent keywords in this sense are: “*external assurance*”, with 299 mentions, “*assurance statement*”, with 297 mentions, and “*limited assurance*”, with 81 mentions (*Figure 11*).

⁶² <https://integratedreporting.org/resource/international-ir-framework/>



Figure 11: most frequent assurances-related keywords

As for the certifications, that were looked up in the reports in the same way as the assurances, the most popular are: ISO 14001, mentioned 136 times, which is a family of standards related to environmental management⁶³; FSC, mentioned 68 mentions, that certifies responsibly sourced wood, paper and other forest products⁶⁴; TCFD, mentioned 57 times, refers to voluntary climate-related financial risk disclosures.

Let us look now at the SDG-related keywords. The most frequent expression in the 430 reports analysed among the list of keywords taken into account is “*environmental impact*”, followed by “*climate change*” and “*greenhouse gas*” (*Figure 12*). Overall, as the wordcloud shows, the dominant topic is the environment (*Figure 13*).

⁶³ <https://www.iso.org/iso-14001-environmental-management.html>

⁶⁴ <https://fsc.org/en/join-us/become-certified>

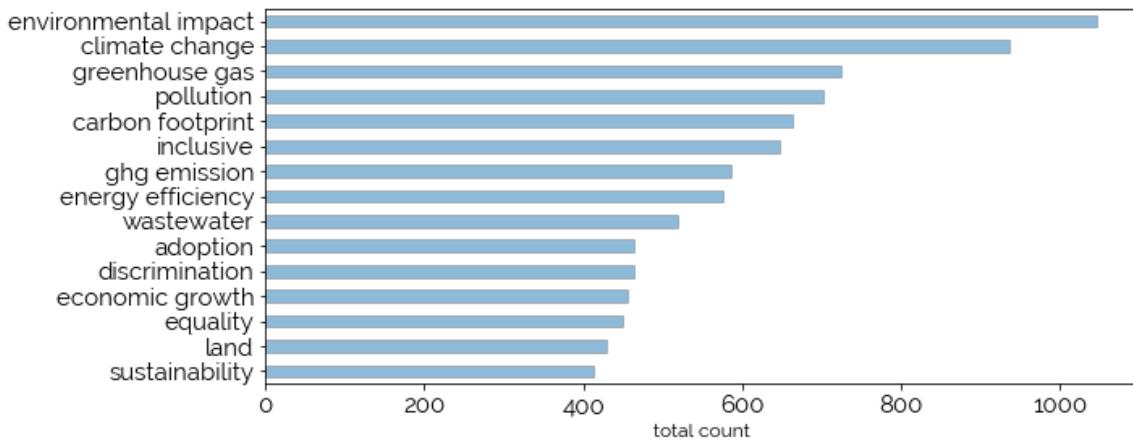


Figure 12: total count of the 15 most popular SDG-related keywords in CSR/sustainability reports

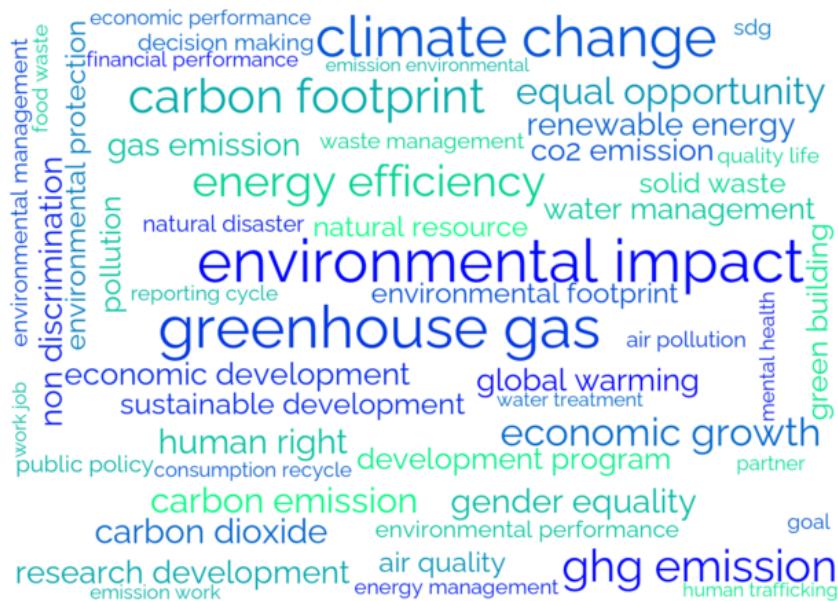


Figure 13: most frequent SDG-related keywords in CSR/sustainability reports

The classification performed on the paragraphs of the reports showed that the most addressed SDG is the *SDG 12 – Responsible Consumption and Production*, followed by *SDG 8 - Decent Work and Economic Growth* and *SDG 7 – Affordable and Clean Energy*. On the other hand, the least addressed Goal is *SDG 14- Life below Water*. The table below shows the average presence of each SDG computed as the average proportion of paragraphs covering that Goal to the total paragraphs related to sustainability (*Table 4*).

	NAME	AVERAGE PRESENCE
SDG 12	Responsible Consumption and Production	23%
SDG 8	Decent work & Economic growth	14%
SDG 7	Affordable & Clean Energy	10%
SDG 15	Life on Land	9%
SDG 16	Peace, Justice & Strong Institutions	9%
SDG 4	Quality Education	9%
SDG 6	Clean Water & Sanitation	4%
SDG 10	Reduced inequalities	4%
SDG 13	Climate Action	4%
SDG 3	Good Health & Well-Being	4%
SDG 9	Industry, Innovation & Infrastructure	2%
SDG 1	No poverty	2%
SDG 2	Zero Hunger	2%
SDG 11	Sustainable Cities & Communities	1%
SDG 17	Partnerships for the Goals	<1%
SDG 5	Gender equality	<1%
SDG 14	Life below Water	<1%

Table 4: SDGs average presence in sustainability reports ordered by frequency

Overall, it is possible to conclude that almost all S&P 500 companies address sustainability topics and, according to the keywords found, they are particularly focused on environmental issues. Such evidence is supported by the classification results too, as the most addressed Goal results to be *SDG 12 – Responsible Consumption and Production*, which relates to environmental issues in production activities, such as CO₂ emission and greenhouse gases emission. Similar topics are also covered by *SDG 7 – Affordable and Clean Energy*, which indeed is one of the top Goals according to our classification algorithm. In this regard, it is to be recalled that SDGs should not be seen as separated and independent clusters, but as deeply interconnected and interlinked by nature (Dörgő, Sebestyén, & Abonyi, 2018). Interestingly, corporate websites' top Goal is SDG 12 too, and in general Goals related to environmental issues are mainly covered. However,

unlike reports, corporate websites largely cover *SDG 16 – Peace, Justice and Strong Institutions*, with reference to corporate governance, as the keywords and section found suggest. Both reports and websites do not seem to extensively address diversity (SDG 10) nor gender equality (SDG 5). The chart below summarizes the differences between reports and websites in terms of classification results (*Figure 14*).

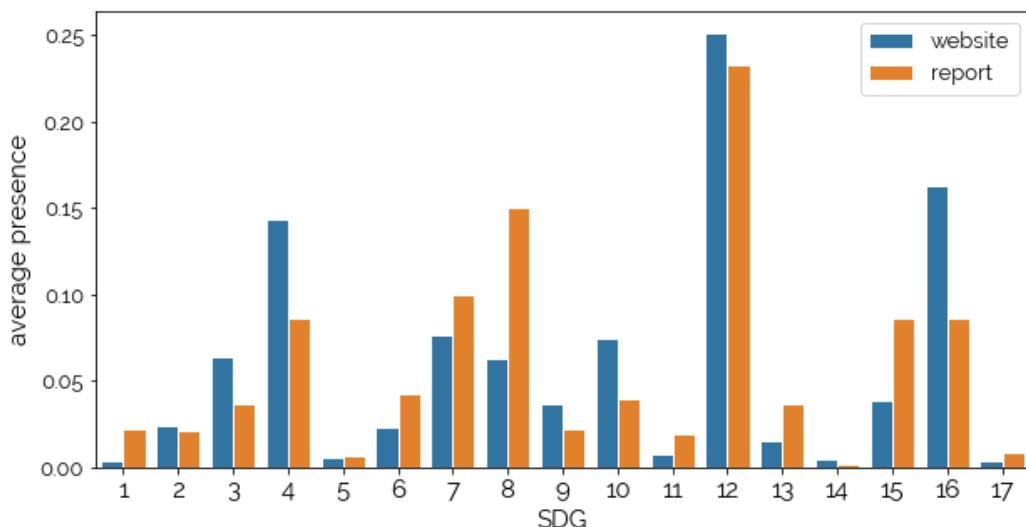


Figure 14: SDGs average presence in corporate websites and reports according to the SDG Classificator

Memberships & Proposals. More than 40 memberships were identified, and their corresponding websites were scraped to look up companies. On average, 14 S&P 500 companies are mentioned by each membership website. The memberships with the highest number of members belonging to S&P500 are CSRwire membership, the leading source of corporate social responsibility and sustainability news⁶⁵, with 103 members, and the Boston College Center for Corporate Citizenship (BCCCC) membership, “*the largest member services organization dedicated to advancing the work of CSR and sustainability professionals and their companies*”⁶⁶, with 93 members. On the other hand, 58% of the S&P 500 companies are members of at least 1 sustainability-committed organization. As for the sustainability-related shareholder resolutions, i.e. the proposals submitted by

⁶⁵ <https://www.csrwire.com/members>

⁶⁶ <https://ccc.bc.edu/content/ccc/membership.html>

shareholders for a vote at companies' meetings, we found 1481 proposals in total, with 5 proposals per company on average. As Ceres also allowed us to monitor ESG proposals from 2009 to present, we found that environmental and social proposals have been outnumbering governance proposals since 2015 (*Figure 15*). This is consistent with what was found by FrameworkESG in 2019⁶⁷. Since data was collected in March 2020, the information about 2020 is to be considered incomplete. By looking at the topics of the proposal, it emerges that the most addressed topic is climate change (80% of the environmental proposals, 30% of the total proposals), followed by political activity (62% of the social proposals, 24% of the total proposals).

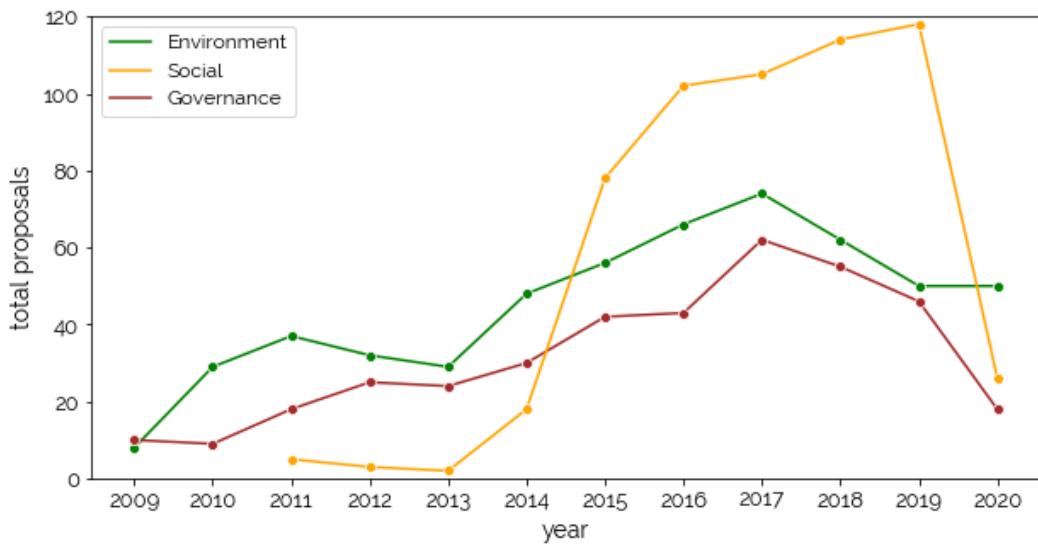


Figure 15: total ESG proposals by S&P500 shareholders over time

Balance Sheets. To conclude the internal data analysis, let us consider the traditional data provided by balance sheets and divided into Environment, Social and Governance data, which are usually employed to assess sustainability corporate performances. Data refers to 2019. The first evidence is that there is an abundance of missing data, which suggests that S&P 500 companies are not fully transparent on ESG disclosure. Taking all the variables together, on average, 50% companies disclose about

⁶⁷ <https://frameworkesg.com/insights/article/esg-for-cxos-2019-proxy-season-trends/>

environmental issues, 68% about social issues and 63% about governance issues. More specifically, as for the Environment performance, all S&P 500 companies are transparent about the total amount of waste produced, the amount of water used, the environmental quality management, the presence of a Green Building Policy, the use of sustainable packaging; on the other hand, there is small data about the amount of water recycled, the amount of discharges to water, the production of hazardous waste, the number of spills. Data about carbon dioxide (CO₂), nitrogen oxide (NO), sulphur oxide (SO) and GHG (greenhouse gases) emission is mostly incomplete.

As for the Social performance, all companies are transparent on social policies, such as the equal opportunity policy, the fair remuneration policy, the health and safety policy, the human rights policy, and the training policy, while there is very little disclosure on the ratio of women in total workforce to women in management, the gender gap, the total incident rate and the employee turnover.

As for the Governance measures, all companies are transparent about the presence of an executive and non-executive director responsible for CSR, the clawback provision for executive compensation, the executive compensation linked to ESG, to give some examples. More specifically, 97% of S&P 500 companies have an executive director with responsibility for CSR. On the other hand, there is no data about the director average age, the board age, the board size and the percentage of women in the board. In regard to gender employment data, not all S&P 500 companies seem to have reached a gender-balanced workplace. On average, in 2019, 38% of the employees were women, and only half of the companies have at least 40% of female employees (*Figure 16*). However, this data refers to only 227 S&P 500 companies, as the others did not disclose about gender diversity in the workforce, and therefore the actual scenario might be more positive or even more negative. Unfortunately, it is not possible to draw conclusion about the

number of women in the board as no data was disclosed. Gender inequalities in the workplace will be further discussed when analysing external data.

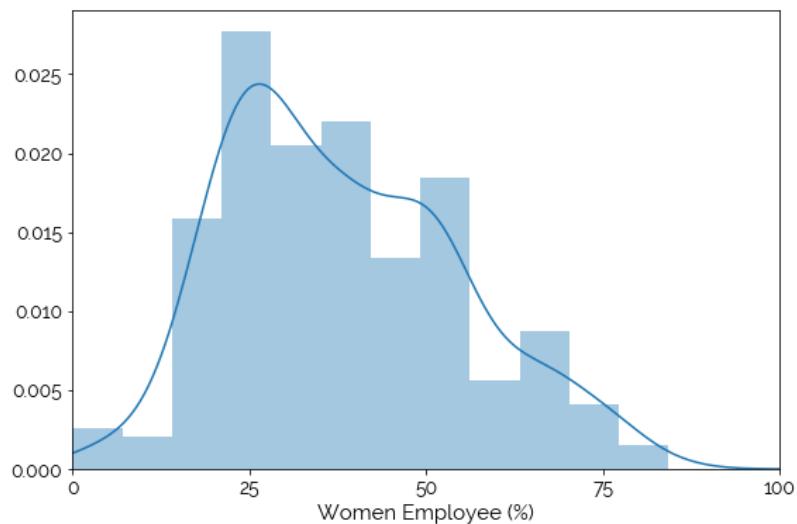


Figure 16: distribution of women employees on the available data (227 companies of S&P 500)

2.2. External Data

External data sources were selected so as to assess corporate performance sustainability based on the external “perceptions” of NGOs, organizations, employees, consumers and other third parties.

FinScience news. Data collected in the second half of 2019 was retrieved from FinScience database based on a list of data sources divided into three categories: NGOs, vertical websites and mainstream news. The extractions below show what mainstream news (*Table 5*) and NGOs news (*Table 6*) look like.

Company	Title	SDG	Sentiment	Popularity	Domain
Exxon Mobil Corp	<i>Minnesota Sues ExxonMobil, Koch Industries, & American Petroleum Institute / CleanTechnica</i>	SDG 7 – Affordable and Clean Energy	-0.34	0.10	cleantechnica.com
Duke Energy Corp	<i>New solar + storage microgrid will supply zero-emission energy for Hawaiian transportation company - Renewable Energy World</i>	SDG 13 – Climate Action	0.26	0.01	renewableenergyworld.com
General Electric Co	<i>Castlight Health intros new tool to help plan safe workplace reopenings / Healthcare IT News</i>	SDG 3 – Good Health and Well-being	-0.12	<0.000	healthcareitnews.com
Wells Fargo & Co	<i>\$40.2 million reminder about the importance of due diligence and monitoring / Federal Trade Commission</i>	SDG 16 – Peace, Justice and Strong Institutions	-0.34	0.001	ftc.gov

Table 5: a small extraction of mainstream news with corresponding SDGs

Company	Title	Sentiment	Popularity	NGO
Exxon Mobil Corp	<i>It's official: Reusables are safe during COVID-19 Grist</i>	-0.23	0.15	Greenpeace
Netflix Inc	<i>#AM_Equality: May 18, 2020 Human Rights Campaign</i>	-0.06	<0.000	Human rights campaign
General Electric Co	<i>'Anything But Secure': Advocates Decry USDA Rule Allowing Big Ag to Set Its Own Regulations on GMOs Common Dreams News</i>	-0.09	0.005	Food And Water Watch
Netflix Inc	<i>Dominican poor struggle for water as coronavirus... Daily Mail Online</i>	-0.21	<0.000	Oxfam International
Dow Inc	<i>Michigan: threat of toxic contamination looms after dam failures trigger flooding US news The Guardian</i>	-0.42	0.01	Sierra Club

Table 6: a small extraction of NGOs news with corresponding source

News collection employs an ontology-based entity recognition, and therefore companies can be recognized and found if they have entities associated. This is the case for 453 of the S&P 500 companies, while the remaining ones remain excluded from the analysis that follows. We found that 342 companies were mentioned in at least one sustainability-specialized website for at least one SDG (vertical data). On average, sentiment of both mainstream and vertical news is slightly positive, with the only exception of mainstream news about the *SDG 6 – Clean Water and Sanitation*, whose average sentiment is somewhat below 0. Mainstream news covering *SDG 16 – Peace, Justice and Strong Institutions* are those mentioning the largest number of companies and having, on average, the highest sentiment (weighted by popularity), which suggests that those companies tackled well governance and justice issues according to the news retrieved (*Figure 17*).

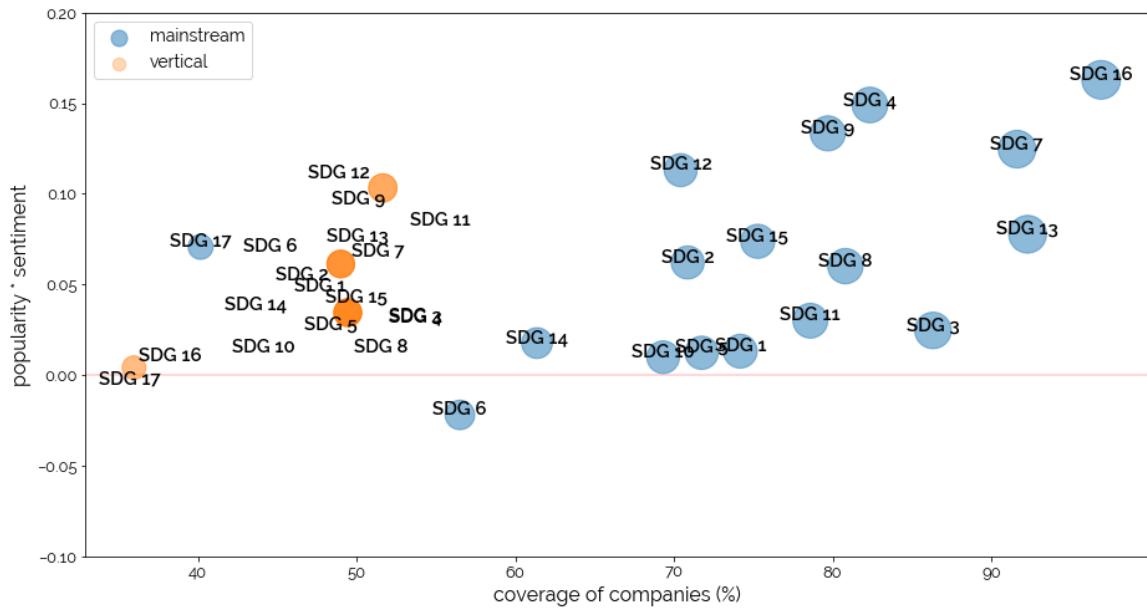


Figure 17: mainstream and vertical news sentiment and coverage of companies by SDG

On the other hand, almost all companies were mentioned in at least one article mentioning SDGs (mainstream data). Among the 47 NGOs monitored, 30 mentioned at least one of the S&P 500 companies. The NGO mentioning the largest number of S&P 500 companies is *Greenpeace* (110 companies), the popular non-governmental organization committed to global environmental challenges⁶⁸. Overall, the sentiment of such news is slightly positive (0.02). On the other hand, news of NGOs such as Amnesty International and FAO are slightly negative. The top NGOs for coverage are shown in the chart below (*Figure 18*).

⁶⁸ <https://www.greenpeace.org/international/explore/about/vision/>

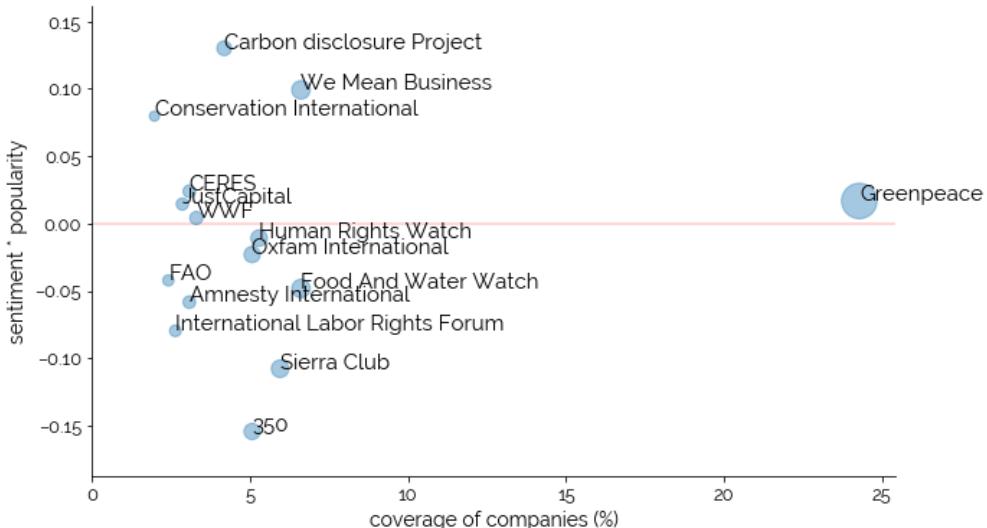


Figure 18: news sentiment and coverage of companies of the 10 top NGOs for coverage

Rankings and Indexes. Through web scraping, text mining on PDF files and csv files analysis we found which of the S&P 500 companies are included in rankings and indexes. On average, each index includes 80 companies, while each ranking 16 companies. The index including the largest number of corporations is Calvert US Large-Cap Core Responsible Index Fund⁶⁹, with 295 S&P 500 members, which is composed of companies that meet the Calvert's Principles for Responsible Investment, based on the ESG pillars, and selected from the 1000 largest US companies by market capitalization⁷⁰. With 74 members, it is followed by Just 100 ranking by Forbes partnered with Just Capital, that takes into account worker treatment, customer treatment, community support, environmental impact and leadership behavior⁷¹. The richest in terms of members is the one provided by the Hispanic Network, with 103 S&P 500 companies, a growing magazine that promotes the advancement of Hispanics in education, business and employment to ensure equal opportunity⁷². In second position, with 288 companies, there is the Corporate Equality Index 2020, the American benchmarking tool on corporate

⁶⁹ <https://www.calvert.com/Calvert-US-Large-Cap-Core-Responsible-Index-Fund-CISIX.php>

⁷⁰ <https://www.calvert.com/media/34498.pdf>

⁷¹ <https://www.forbes.com/just-companies/#59f4bc632bf0>

⁷² <https://hnmagazine.com/recognition-lists/#1494965575890-0b7e873b-129j>

policies, practices and benefits pertinent to lesbian, gay, bisexual, transgender and queer employees⁷³. Taken together, these results suggest a large part of S&P 500 companies is perceived as committed to social responsibilities.

Controversial Activities. Following from the indexes and rankings analysis, it is to be recalled that we also monitored negative rankings, namely those lists that refer to a corporate strategy or performance with a negative impact. The Political Economy Research Institute (PERI) of the University of Massachusetts provides lists of the top water and air polluters and greenhouse producers. We found that 8% of S&P 500 companies belong to the water polluters ranking and 8% to the air polluters ranking, while only 4% of companies are mentioned as top greenhouse gas producers, and they operate mainly in the Energy, Industrials, Utilities and Materials sectors. In the chart above, the width of the flow is proportional to the number of companies belonging to the sector indicated on the left that negatively contribute to the environmental issue on the right (*Figure 19*).

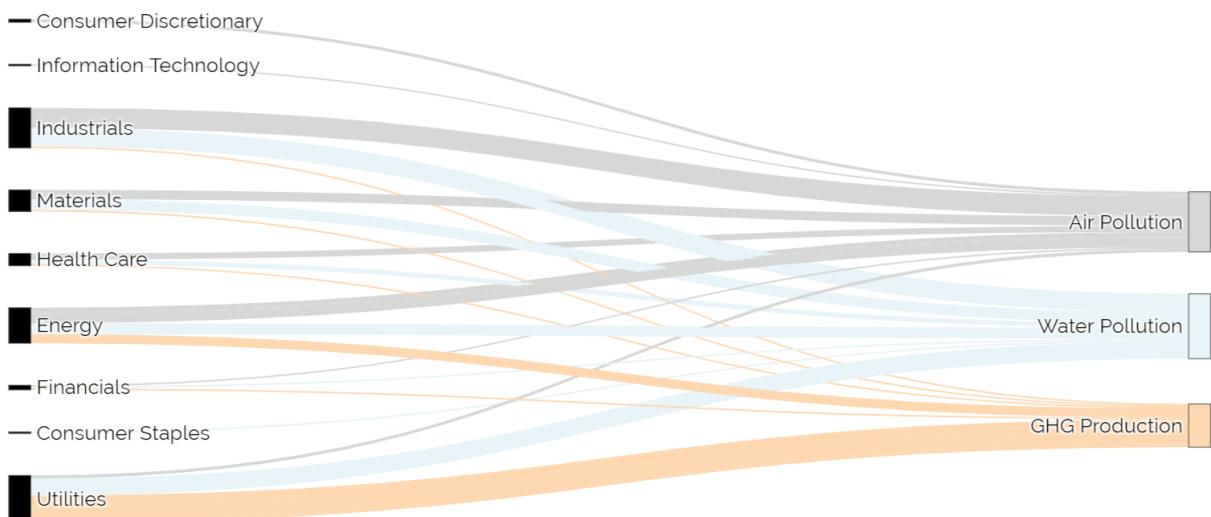


Figure 19: polluters by categories and sectors based on PERI data

⁷³ <https://www.hrc.org/resources/corporate-equality-index>

Among the S&P 500 companies, we also found that 55 are using hazardous chemicals, most of which belong, not surprisingly, to the Materials sector (17 companies), followed by the Energy sector (11 companies). In total, 892 substances are being used by these 55 companies, corresponding to 300 distinct compounds, which in turn can be clustered into different chemicals groups, as indicated by the non-profit organization ChemSec⁷⁴. The most common are the petroleum substances (56% of the total) followed by Highly reactive compounds (12%). The company using the largest number of hazardous chemicals is an Oil, Gas & Consumable Fuels producer, for a total of 136 different substances. It is to specify that for 16% of such substances, less toxic alternatives are available in the market. The chart above shows the total number of hazardous chemicals by sectors, with the count of companies per sector using them (*Figure 20*).

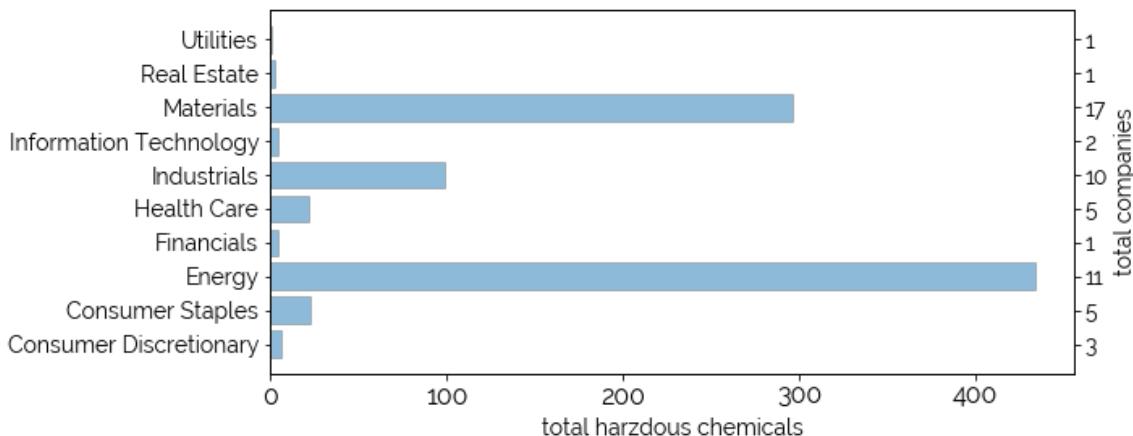


Figure 20: total number of hazardous chemicals by sectors and number of companies per sector

Let us look now at corporate misconducts, as collected by scraping Good Job First's database. Violations are divided into 4 main categories: Governance, Consumer, Labour and Environment. From 2019 until March 2020, 35% of S&P 500 companies committed at least one violation, with a maximum of 3 violations committed by the same organization.

⁷⁴ <https://sinlist.chemsec.org/chemical-groups/>

Overall, it emerges that most of the corporate misconducts relate to labour rights (55% of the total violations), followed by governance (27%), consumer rights (12%) and, at last, environmental issues (6%). In terms of sectors, the top “violator” is the Industrials sector; the Financials sector is the one that committed the largest number of misconducts related to Governance; violations related to Environment and Consumer issues are spread over almost all sectors. These results are summarized by the chart above, with the size of each squared node being proportional to the number of incoming edges (*Figure 21*).

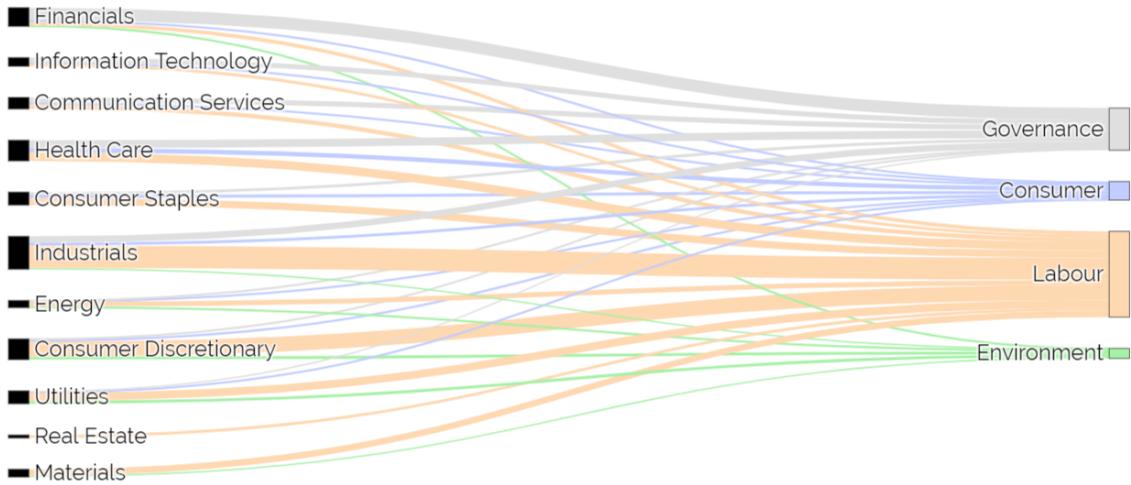


Figure 21: violations committed in 2019-2020 by S&P 500 companies by type and sector

An additional criterion that investors may want to consider when adopting socially responsible investing strategies is whether or not, and to what extent, companies are involved in arms sales. We found that 3% of S&P 500 companies sell arms, most of which belong to the Industrials sector, and more specifically, and not surprisingly, to the Aerospace & Defense subsector. The chart below shows the distribution of the Sectors, and their corresponding subsectors, that were involved in arms sales in 2018 (*Figure 22*).

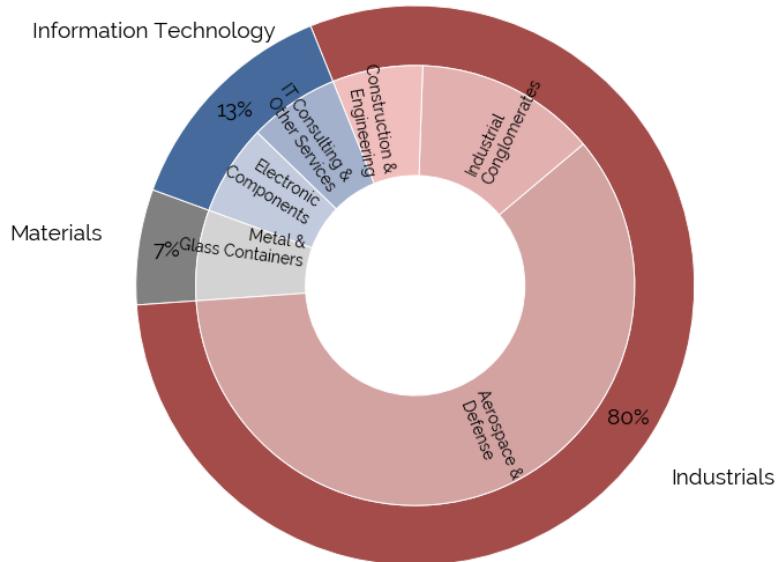


Figure 22: sectors and subsectors involved in arms sales in 2018

Taken together, these results show that S&P 500 companies are involved in several controversial activities that relate not only to environmental pollution, due to gas emission and use of toxic chemicals, but also to governance issues, consumer and labour rights. Overall, Energy, Materials and Industrials sectors are the ones that seem to negatively impact the most sustainable development.

Reviews. What do employees and former employees think about their companies? And what about their products? Glassdoor, Indeed, Payscale and Trustpilot were investigated to answer these questions. The first evidence is that these websites do not provide the same coverage of the S&P 500 companies, with Glassdoor being the only one covering all companies, followed by Indeed and Payscale, with almost a full coverage, and at last Trustpilot, with almost a 50% coverage (*Figure 23*).

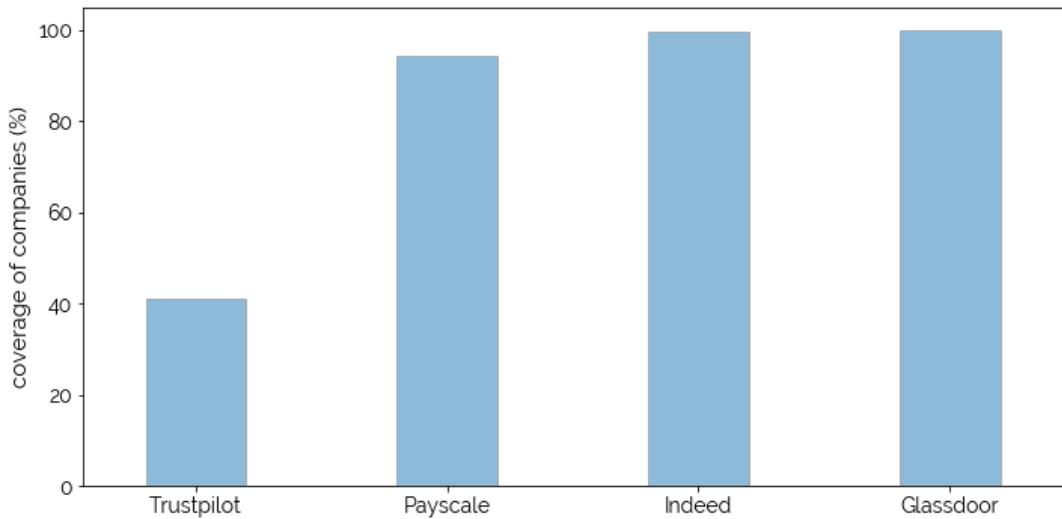


Figure 23: coverage of S&P500 companies by reviews websites

Glassdoor will be analysed first. It provides three main types of data: reviews, ratings and pledges. Pledges can be seen as boolean variables, as they can either be present or not in the Glassdoor page of a company, based on specific programmes or commitment declared by the employers. Not all S&P 500 companies have a pledge on their Glassdoor page. The most popular is the Diversity Commitment pledge, which is exhibited by almost 35% of the companies, followed by the Social Responsibility Pledge (28%). On the other hand, less than 5% of corporations seem to offer returnship opportunities (internships for adults who have been out of the workforce). The barchart shows the percentages of S&P 500 companies exhibiting pledges (*Figure 24*).

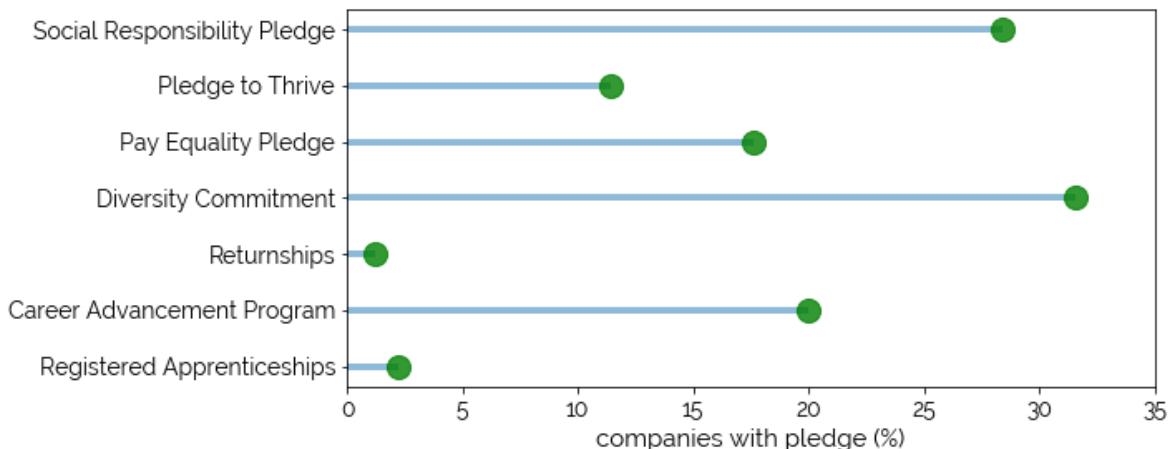


Figure 24: pledges distribution over S&P companies' Glassdoor pages

As for the ratings, they are expressed either as stars (from 1 to 5) or percentages (from 0 to 100). The ratings available regard the work-life balance, compensations and benefits, career opportunities, senior management, culture and values, and an overall rating, while the percentages are related to recommendations to friends, positive interview experience and positive perspectives. As the heatmap below illustrates (*Figure 25*), the total number of reviews is not correlated with any other measurement. It is then possible to distinguish 3 main blocks: the ratings and the percentage of recommendations to a friend, that are highly positively correlated between them, which suggest that if a company receives a good rating for a specific issues, it will probably perform well in the other ratings too, and will be also recommend to other people; the remaining percentages, namely the positive interview experiences and positive perspectives, are not correlated between them nor they are with other metrics; the third block is represented by the pledges, the ones of which relate to diversity, inclusion and pay equality are positively correlated between them, while apprenticeships and returnships (which are also quite rare) are not correlated with any measurement.

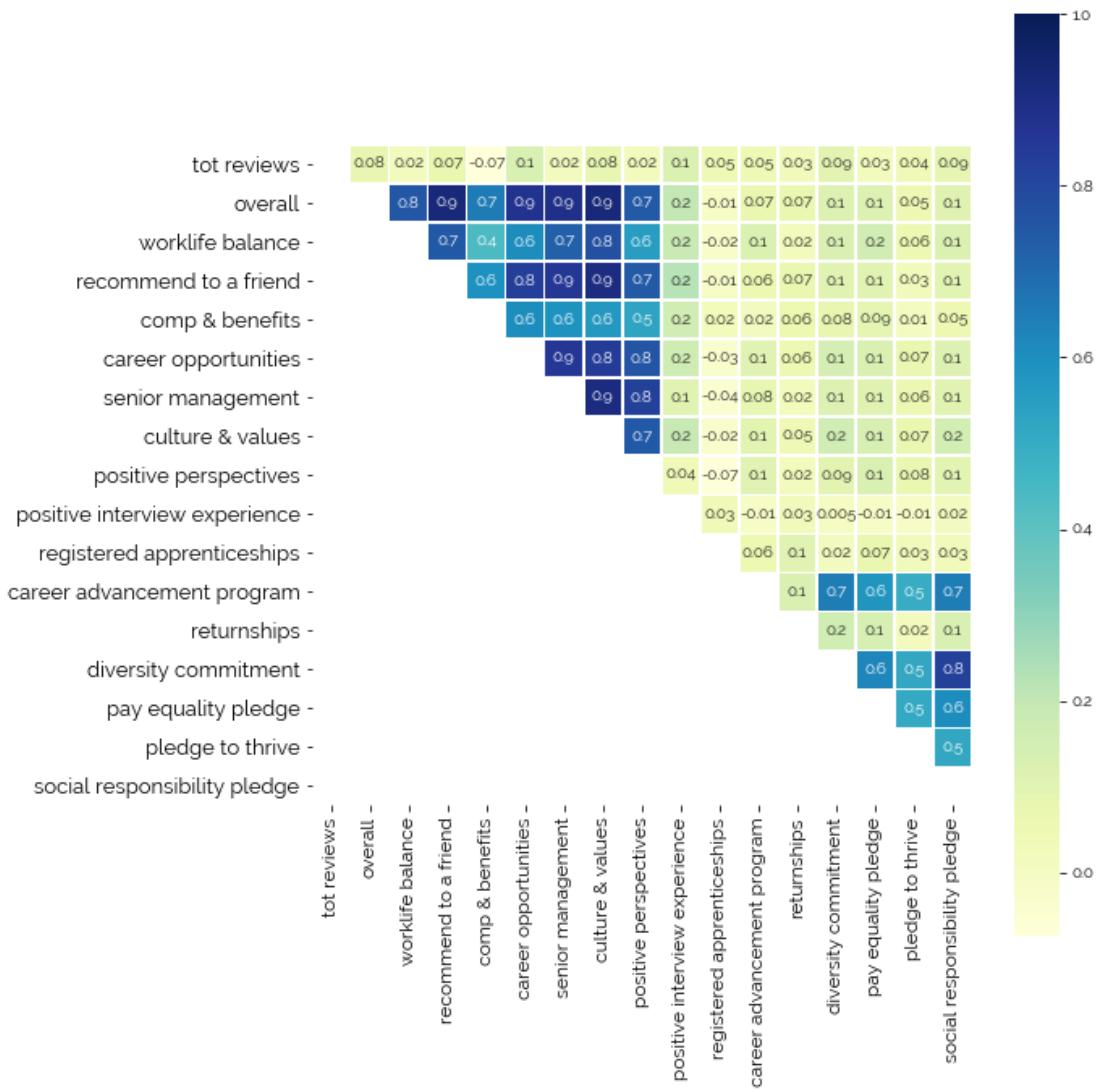


Figure 25: heatmap of Glassdoor measurements on S&P500 companies with correlation values

While there are not significant differences between sectors in terms of average ratings, it is possible to detect significant differences between some of the ratings, by performing t-tests on their corresponding distributions (with $\alpha=0.01$). For instance, Senior Management is rated as significantly lower than Culture & Values (t-statistics=-13.080, $p=0.000$, *Figure 26*), and Worklife Balance (t-statistics=12.441, $p=0.000$, *Figure 27*). Culture and Values are perceived as significantly more positively than Career Opportunities too (t-statistics=6.563, $p=0.000$, *Figure 28*).

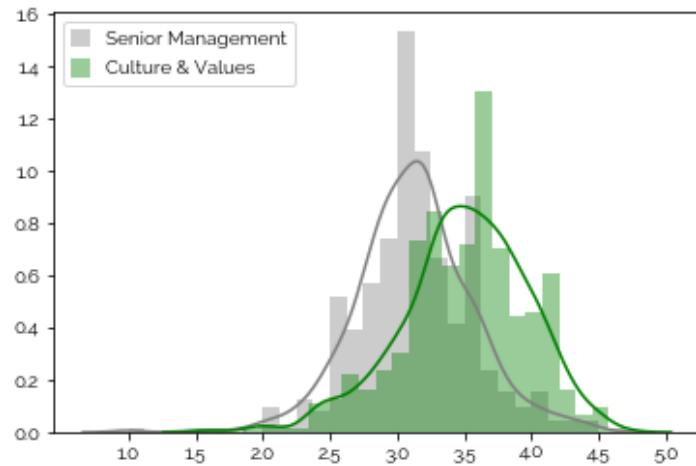


Figure 26: senior management and culture & values ratings distributions

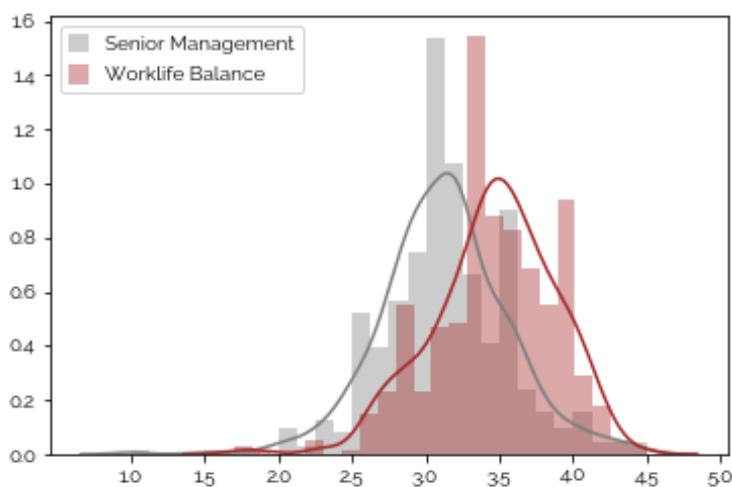


Figure 27: senior management and worklife balance ratings distributions

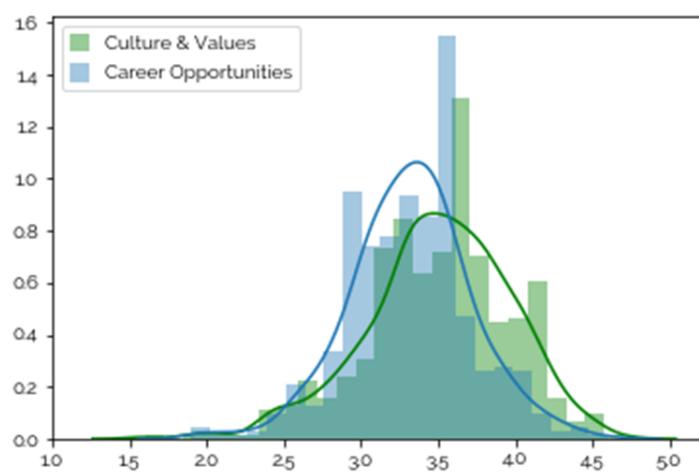


Figure 28: culture & values and career opportunities ratings distributions

These results, together with the previous findings about the most common pledges, suggest that S&P 500 companies distinguish themselves for cultures and beliefs, with a special attention to social inclusion (see the abundance of social responsibility and diversity pledges as well as the ratings discussed). On the other hand, senior management and career opportunities tend to receive lower ratings; in addition to this, there is an overall scarcity of returnships, career advancement programme and apprenticeships pledges.

Another very popular review website for employees is Indeed. It provides 17 ratings, some of which recall the Glassdoor items (worklife balance, management happiness, compensation and benefits). Once again, ratings are highly positively correlated between them, with correlation values ranging from 0.5 to 1; the only exception is Compensation and Benefits, whose correlation values are not as high as the others, as they range from 0.3 to 0.6 (*Figure 29*).

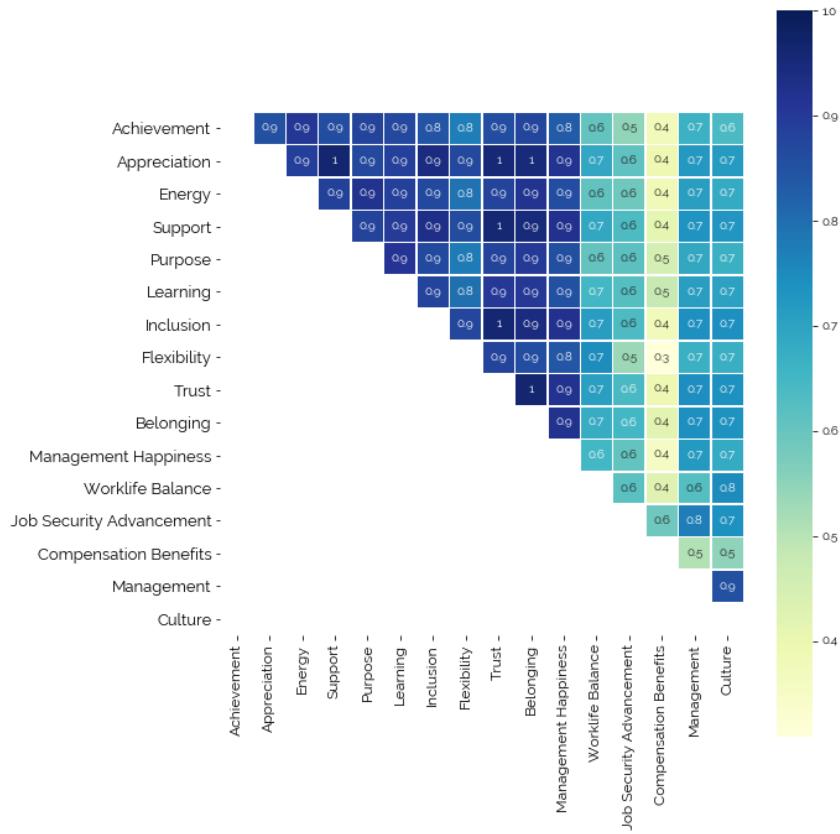


Figure 29: heatmap Indeed measurements on S&P500 companies with correlation values

Such metrics were in turn divided into 4 main categories: work satisfaction, management evaluation, compensation and company culture (*Figure 30*). S&P 500 companies are better perceived in terms of compensation and work satisfaction, while company culture and management evaluation scores are on average lower (*Figure 31*).

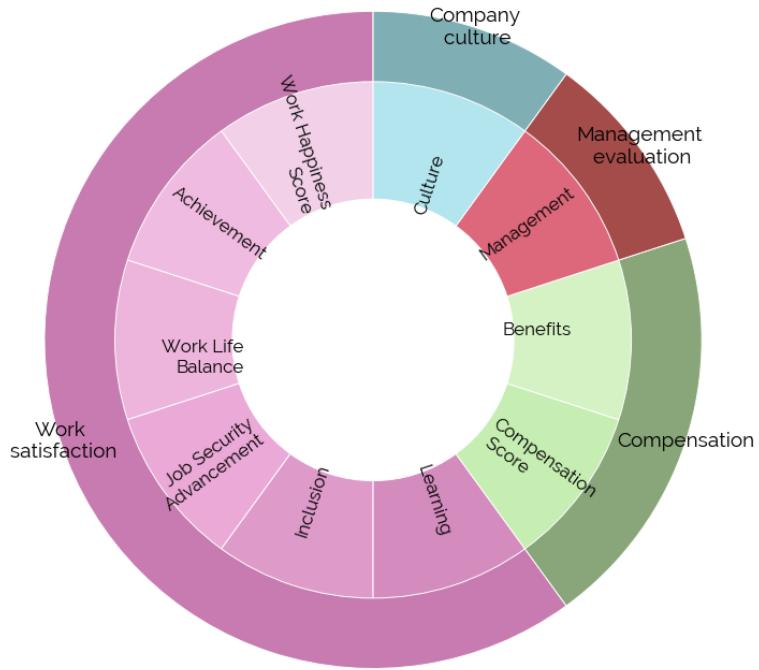


Figure 30: *Indeed ratings distribution by category*

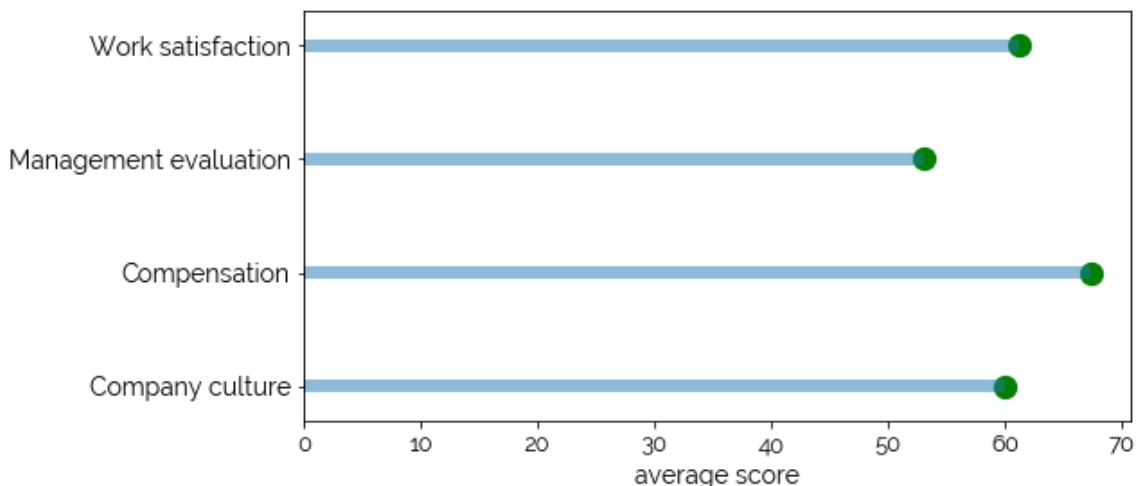


Figure 31: *S&P500 average Indeed ratings by category*

If Glassdoor and Indeed are specifically specialized in work satisfaction and happiness as experienced by employees and former employees, Payscale draws attention to salary and remuneration. Interestingly, ratings related to pay are in fact the lowest, compared to other indicators (*Figure 32*). This result is not consistent with Indeed data, that would in fact

suggest a general satisfaction with S&P 500 compensation. However, Payscale is more specific in this sense, as it breaks down pay into three items, namely fair pay, fair policy and pay transparency, which also address justice and governance issues.

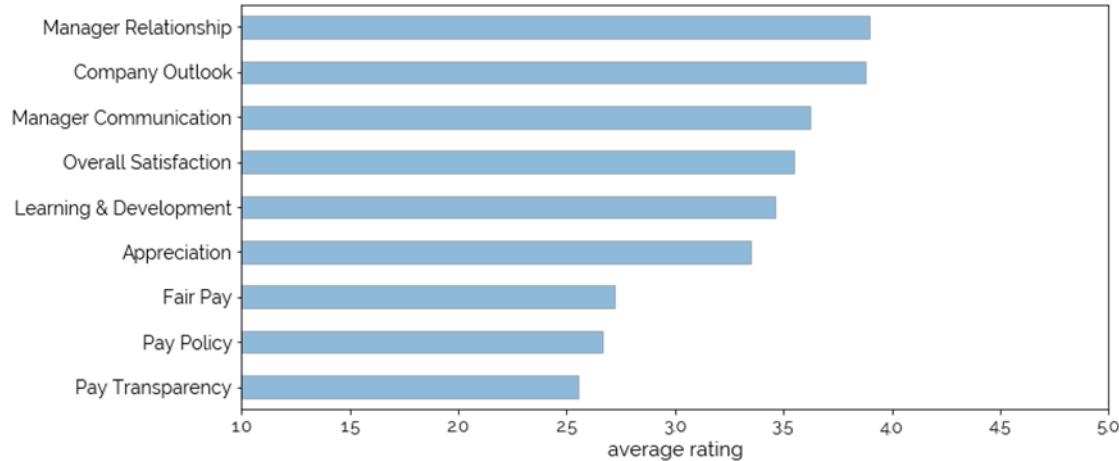


Figure 32: average Payscale ratings on S&P500 companies

Unlike the previous websites, Payscale also offers information on gender diversity in the workplace, by providing the proportion of women to the total employees and the average salary by gender (women and men are considered). It is evident that women are underrepresented in S&P 500 companies and underpaid compared to their male colleagues. Industrials, Information Technology and Materials area the sectors with the lowest percentage of female employees (around 30%). On the other hand, Real Estate, Consumer Discretionary and Consumer Staples companies are closer to reaching gender parity, with slightly less than 50% female employees (*Figure 33*).

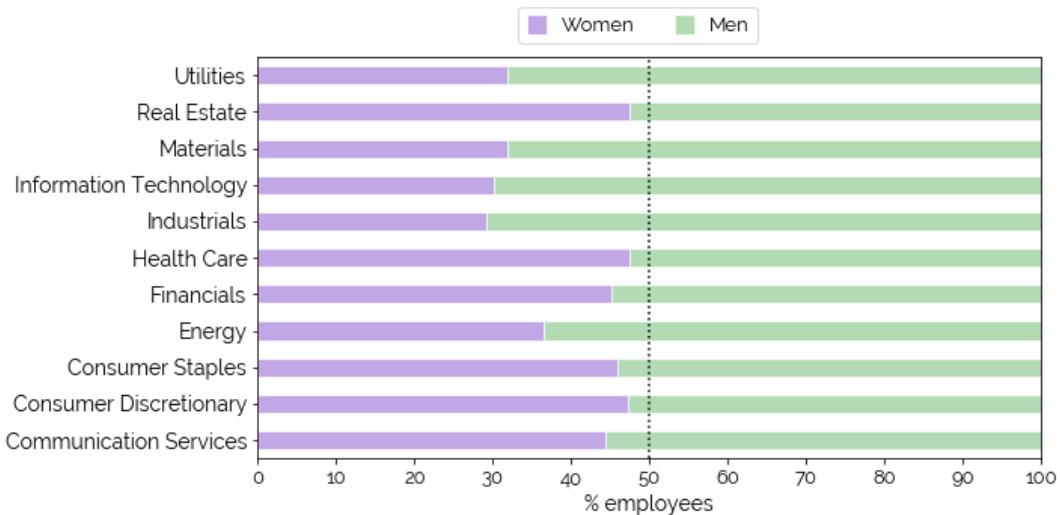


Figure 33: gender disparity in S&P500 companies by sector

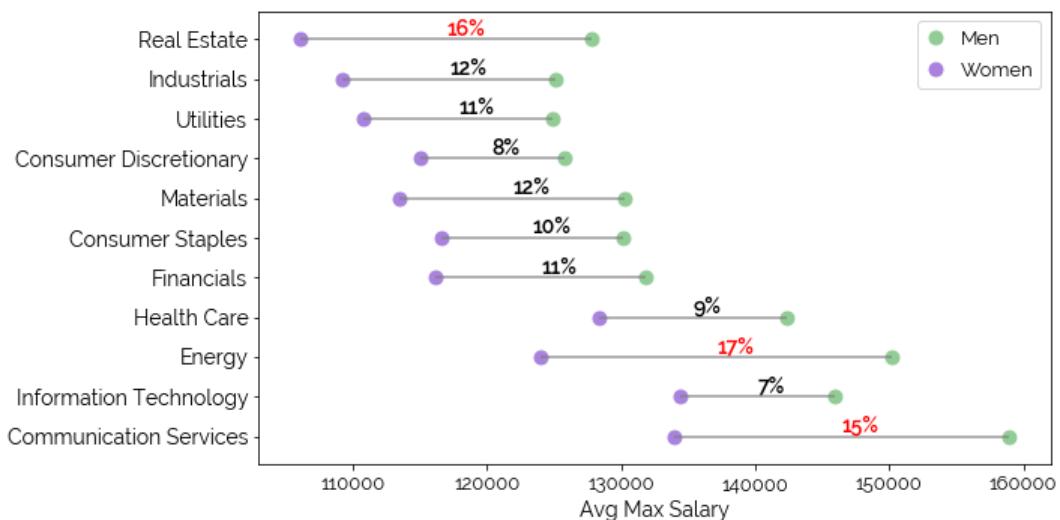
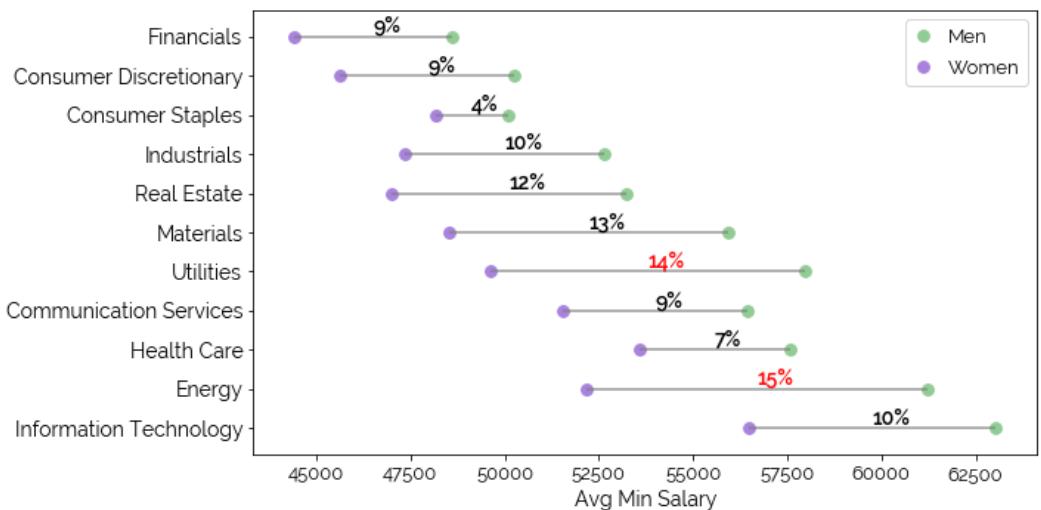


Figure 34: S&P500 gender wage gap based on Payscale data

Between January 2018 and January 2020, nearly 1.6 million people took PayScale's online salary survey, providing information about compensable factors⁷⁵, and that is where the strongest disparity emerges. Data related to the minimum and maximum salary, as disclosed by female and male employees, is in fact available for 83% of the S&P 500 companies. We computed the wage gap as the difference between the annual salary of men and the annual salary of women salary as percentage of male earnings⁷⁶. In all sectors, women earn significantly less than men, with both the minimum and the maximum salary being lower than the male counterparts. Interestingly, the gap is not constant within the same sector, but it changes if the minimum or the maximum salary is considered. If we look at the minimum salary, the widest gap is reached by the Energy sector, in which a woman earns 0.85\$ for each dollar earned by her male colleague. More specifically, at minimum, a woman makes 52174\$ per year in this sector, while a man 61208\$. However, the gap is even wider if the maximum salary is considered, as men make 137103\$ per year, while a woman 17% less. Similarly, the gap increases in the Real Estate sector, moving from a 12% difference between minimum salaries to 16% between maximum salaries. This evidence suggests that at higher positions, corresponding to higher salaries, gender disparity tends to become more severe. On the other hand, the "fairest" sector is the Consumer Staples industry, which mainly regards food and beverage, household goods, and hygiene products, where women earn on average 7% less than men. These findings are summarized by the two charts in *Figure 34*, which illustrates the gender wage gap by sector, ordered by minimum and maximum salary. The longer the horizontal line, the wider the gap; the more shifted to the right, the higher the salary. If minimum and maximum wages are considered altogether, by computing the mean of the two, it emerges that on average, women in S&P 500 companies 90 cents for every dollar paid to

⁷⁵ <https://www.payscale.com/data/gender-pay-gap>

⁷⁶ https://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics

their male counterparts, or in other words, they make 10% less than men. This result is more encouraging than the findings reported by the American Association of University Women, according to which women make 82% of what men earn in the United States⁷⁷.

To conclude the analysis of reviews websites, let us consider now Trustpilot. Unlike the previous websites, Trustpilot is a consumer review website, and therefore ratings and opinions relate to the company's products, rather than to the work environment. As mentioned previously, less than half of the S&P500 companies are covered by Trustpilot, hence the analysis that follows is not fully comprehensive. Users can provide a rating from 1 to 5 together with a review. Overall, ratings tend to be low, with the average being 2.6/5. Such value is much lower than the average ratings described so far, which were above 3/5. This general trend, together with a check on the content of the reviews, suggest that either consumers tend to use Trustpilot to mainly report negative experiences⁷⁸ or their experiences with reviewed companies' products are in fact bad in general. Interestingly, the review ratings are indeed significantly negatively correlated with the log number of reviews ($r = -0.5$, $p=0.000$), as *Figure 30* suggests. By looking at the colors of the dots, with each dot representing a company, it is also possible to detect a prevalence of Consumer Discretionary (22% of the total), Information Technologies (21%) and Financials (17%) companies. In terms of ratings, the worst performing sector is Communication Services, with 2.1/5 stars on average.

⁷⁷ <https://www.aauw.org/resources/research/simple-truth/>

⁷⁸ See for instance <https://www.marketingcharts.com/digital-28628>

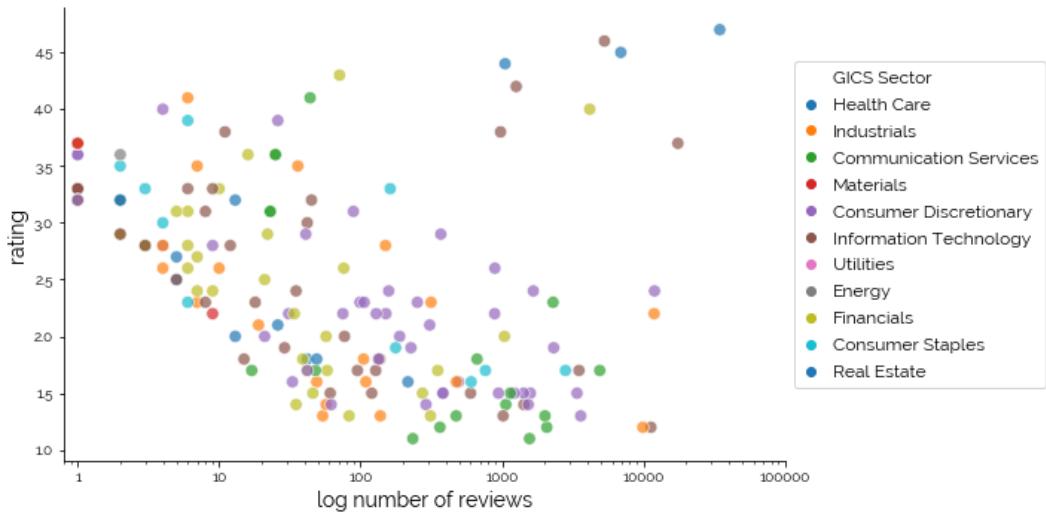


Figure 35: Trustpilot review ratings over log number of reviews per company

To conclude, the analysis on external data reveals several aspects of S&P 500 commitment to sustainable goals. Overall, there is a strong interest in social issues, as the inclusion in several rankings and indexes, employees' reviews and proposals show. However, based on the information available, diversity and inclusion are still goals to be achieved, specifically in terms of gender equality (while there is not much data about ethnicity, sexual orientation nor disabilities). For some industries, the social performance is also affected by the participation in controversial activities, such as arms sales and misconducts related to labour rights. On the other hand, environmental performance is affected by the high production of air and water pollutants and the use of hazardous substances. The next paragraph will finally connect the dots.

2.3. Scores

Having collected and explored the data as discussed in the previous paragraphs, let us move on building the actual FinScience Sustainability Score. The first step, as described in the methodology, is to compute the SDGs internal and external sub-scores, by averaging, for each SDG, the re-scaled data (from 0 to 1) that contribute to the Goal weighted by their relevance. It is also to be recalled that missing data is handled differently based on the information that it conveys and on the importance of its disclosure. For this reason, most of the balance sheet data (internal data) weights on the internal SDGs scores as their absence suggests a lack of transparency. On the other hand, the data scraped from the corporate websites does not come into play if missing, as the websites might not be visited because of scraping limitations that the company is not responsible for. Similarly, we decided to exclude external variables, if missing, from the score computation, as they provide viewpoints voluntarily expressed by third parties, and as such not strictly required. Although all the variables, sub-scores and the final scores were designed to range from 0 to 1, the scores will be represented and described as multiplied by 100 to conform with scoring systems.

Let us start by looking at correlations between scores. As predictable, most of the scores are correlated between them. In this regard, it is to be recalled that UN SDGs are interlinked by design, and so are the variables discussed so far, as the same variable can contribute to several Goals. The internal scores for SDGs 1, 2, 3, 4, 5, 6 are highly positively correlated, which means that companies that positively impact hunger, well-being, education, gender equality or water sanitation are likely to do so on all these issues at the same time. This is also the case for SDGs 7 (clean energy), 11 (sustainable cities), 12 (responsible production) and 13 (climate action), which are indeed strictly interconnected. On the other hand, the Internal SDG 17 score (partnerships for the goals)

is not correlated with any other scores (*Figure 36*). However, the same relationships are not found when considering the External SDGs scores. As the second heatmap clearly shows (*Figure 37*), the correlation values are indeed much lower, with a few exceptions, such as SDG 1 (poverty) and SDG 2 (hunger), and SDG 5 (gender equality) and SDG 10 (inequalities). What can be said about the Internal and External Score computed for the same Goal? Is it possible to conclude that if a company's internal score is good, so will be the corresponding external score? This is not what emerges from the computed scores. In fact, there is no significant correlation within any of the SDGs Scores pairs. In other words, the contributions of a company belonging to the S&P 500 index to a specific SDG is not informative of how the company is externally perceived in relation to that Goal. A possible interpretation of these companies is that companies might tend to communicate on several sustainable issues at the same time, while, at the external level, it is possible to better discern between sustainable topics.

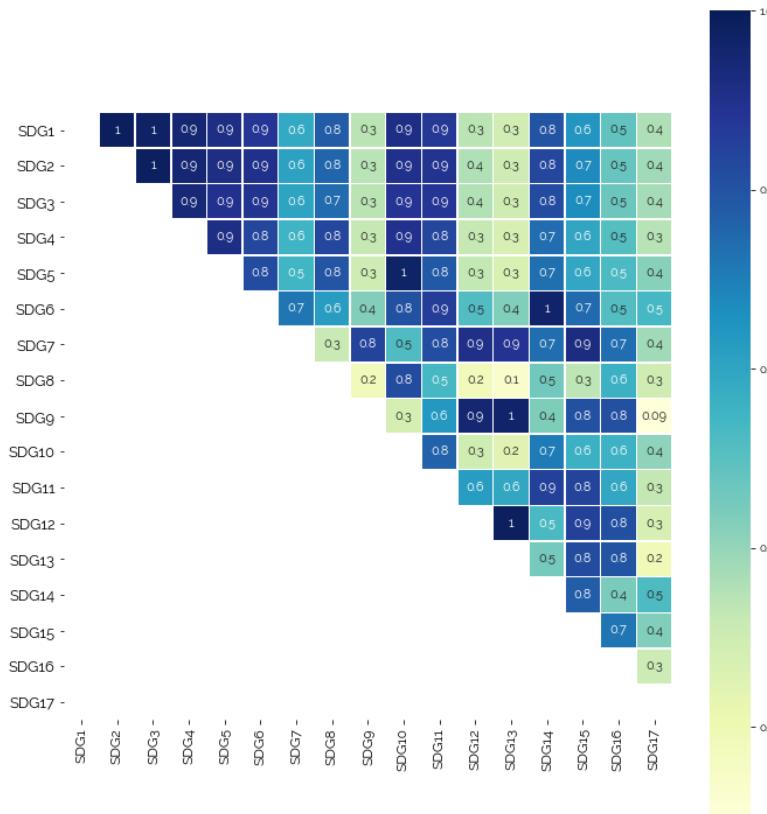


Figure 36: heatmap of Internal SDGs Scores with correlation values

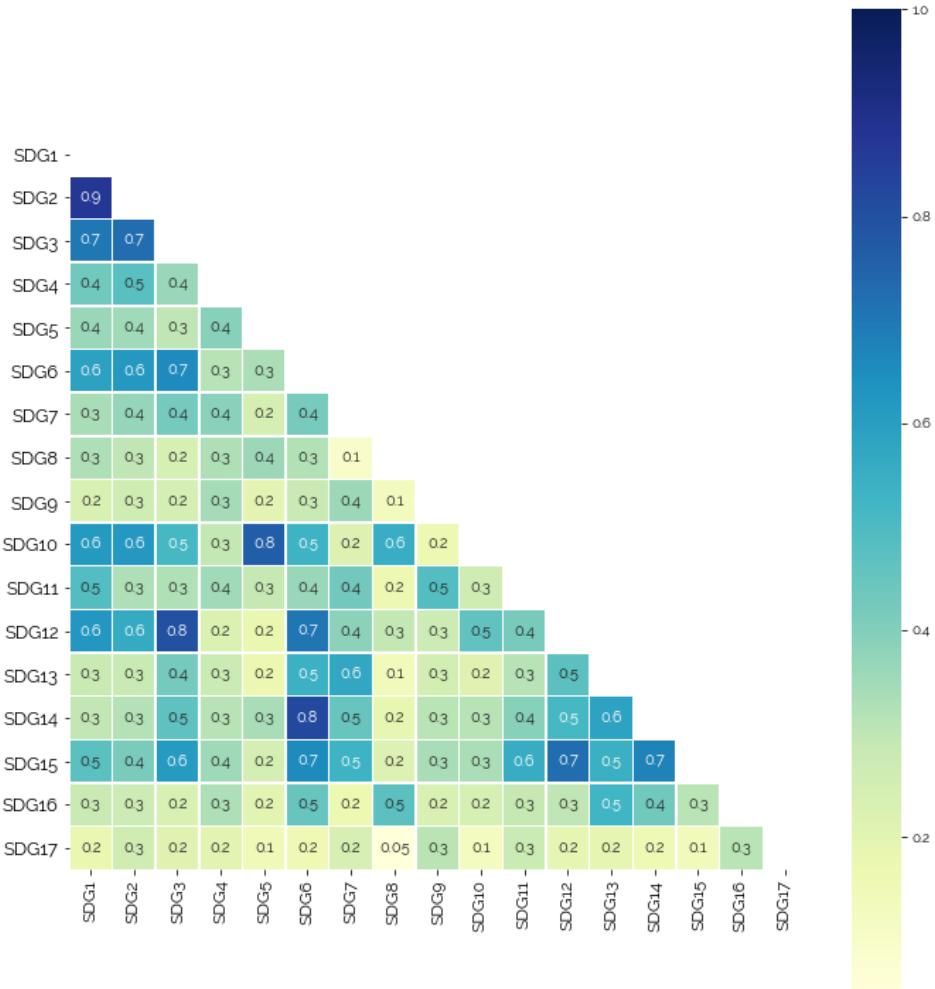


Figure 37: heatmap of External SDGs Scores with correlation values

What are then the Goals that S&P 500 companies contribute to the most? And what is the least addressed sustainability issue? On average, S&P 500 companies' best internal score is achieved for *SDG 16- Peace, Justice and Strong Institutions* (78 on average). From an ESG perspective, this means that, based on the corporate self-disclosed data that were considered, the largest American companies are more committed to Governance issues. On the other hand, the lowest scores, on average, relate to *SDG 14 – Life Below Water* (59 on average). *Figure 38* shows the distributions of the two SDGs Scores, which are indeed significantly different (t-statistics=-29.118, p=0.000). *Figure 44* shows the average Internal SDGs Scores for all the 17 Goals. Besides the SDG 14, it is easy to notice that the smaller “slices” of the pies, with an average score lower than 60, are the ones

corresponding to *SDG 7 – Affordable and Clean Energy*, *SDG 12 – Responsible Consumption and Production*, *SDG 13 – Climate Action* and *SDG 6 – Clean Water and Sanitation*. As all of these Goals relate to the Environment pillar, these findings suggest that S&P 500 companies' contributions towards a green economy should be further strengthened (*Figure 39*).

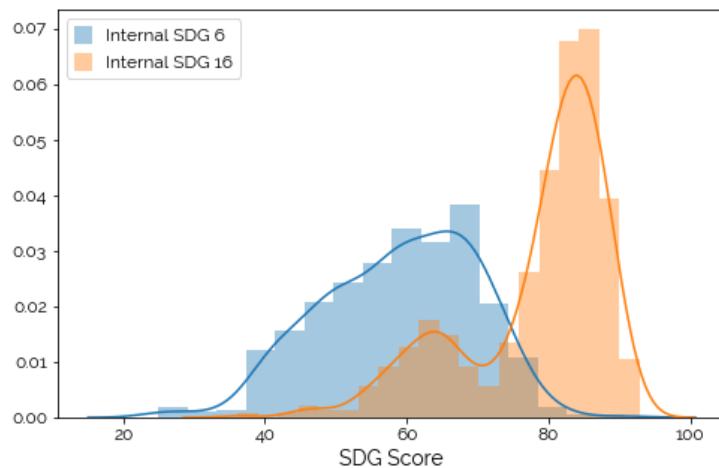


Figure 38: lowest and highest internal scores distributions

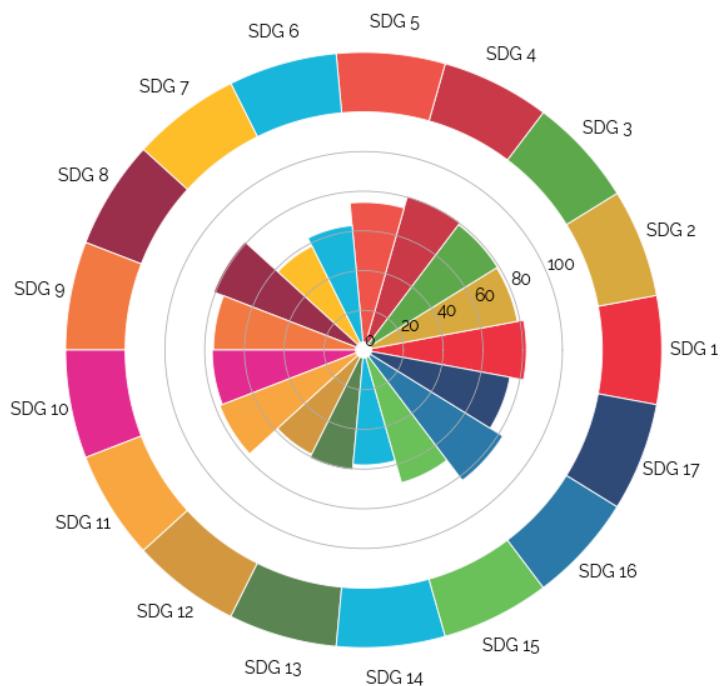


Figure 39: average Internal SDGS Scores

How are S&P 500 companies externally perceived? Do employees, NGOs, specialized shareholders and academics believe that the largest US companies are more committed to Governance and Social matters than to Environmental issues? On average, External SDGs Scores are slightly lower than their internal counterparts. On the other hand, External SDGs Scores result to be more homogenous, meaning that there are no evident differences between the scores. Compared to the previous pie radar chart (*Figure 39*), the following pie looks indeed more regular (*Figure 41*). Nonetheless, *SDG 4- Quality Education* can be recognized as the highest External SDG Score, with an average of 73, and *SDG 12- Responsible Consumption and Production* as the worst-performing SDG, with an average of 63 (*Figure 40*).

To better understand the differences between internal and external scores, the ESG perspective can be applied to group the SDGs scores in broader categories, by mapping each Goal to one ESG pillar. In this case, the company's sector was not included in the computation, and therefore the SDGs are weighted equally. This will not be the case for the Final Scores, as I will illustrate later on. What emerges from the ESG grouping is that, on average, Environment Internal Score is higher than the corresponding External Score, which suggests that the external perception about corporate commitment to Environment outperforms the company's self-disclosure on this matter. On the other hand, the overall Governance External Score is lower than the corresponding Internal Score, which can be interpreted as if companies tend to be perceived on Governance issues as less virtuous than what they declare. Finally, there is not a significant difference, on average, between the Internal and External score on Social performance, and therefore self-disclosure and third-parties' opinions and perception on social matters tend to match (*Figure 42*).

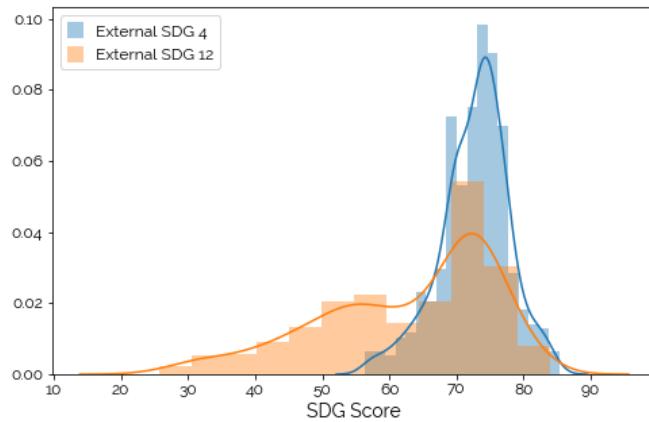


Figure 40: lowest and highest external scores distributions

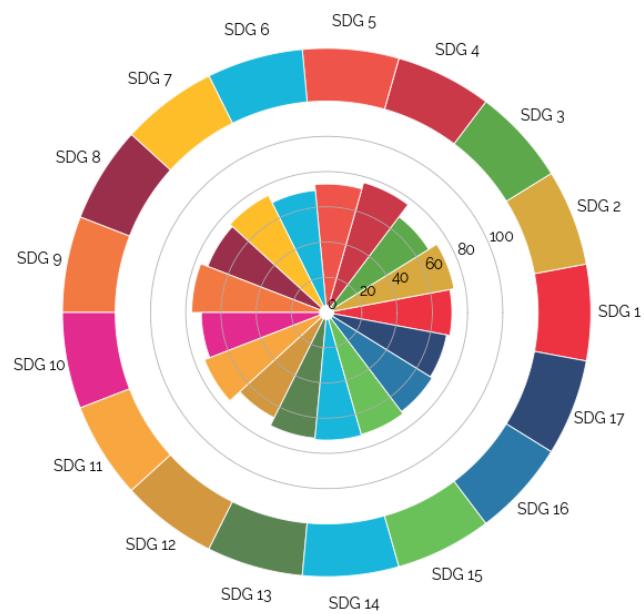


Figure 41: average external SDGS Scores

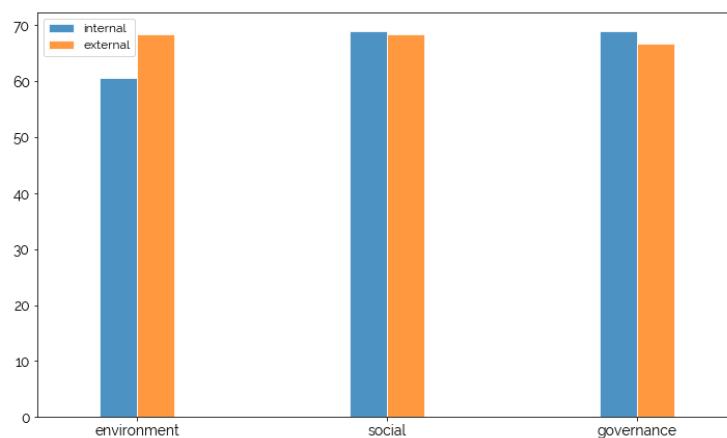


Figure 42: average SDGs Scores aggregated by ESG categories

From the combination of so-computed Internal SDGs Scores it is then possible to compute the overall Internal Score. This is calculated as the weighted average of the internal sub-scores, with the weights being determined by the materiality analysis presented in the first chapter. In other words, for each company, the sub-scores of the SDGs material to the Sector that the company belongs to are weighted more heavily than the others. Similarly, the overall External Score is computed, with the only difference being that the 17 External SDGs Scores are taken into account. *Figure 43* shows the average Internal and External Score by sector, with respect to the average between such values (that is 68). In general, External and Internal Scores do not match, and the discrepancy between the two scores varies based on the sector. More specifically, the External Score significantly outperforms the Internal Score in Communication Services, Financials, Health Care, Information Technology and Real Estate sectors. In other words, for these sectors, the external perception of the corporate contributions to sustainable development is more encouraging than the corporate self-disclosed data itself. On the other hand, Materials (manufacture chemicals, construction materials), Utilities (electric, gas and water utilities), Energy (oil, gas, coal and consumable fuels), Consumer Staples (foods, beverage, household goods, and hygiene products), Industrials (aerospace and defense, building products, electrical equipment) and Consumer Discretionary (automotive, household durable goods, leisure facilities) sectors tend to be perceived by external sources, such as consumers, employees, news, specialized organizations and other third parties as less virtuous than what companies self-report (*Figure 43*). The highest discrepancy is reached by the Materials sector, while the smallest gap between external and internal scores is in Real Estate sector (*Figure 44*)

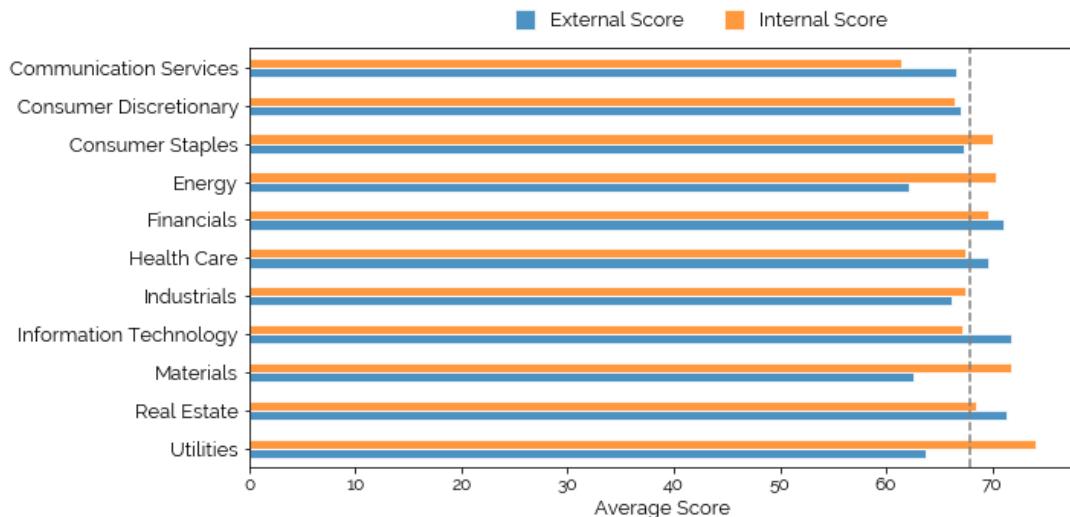


Figure 43: average Internal and External Score by Sector

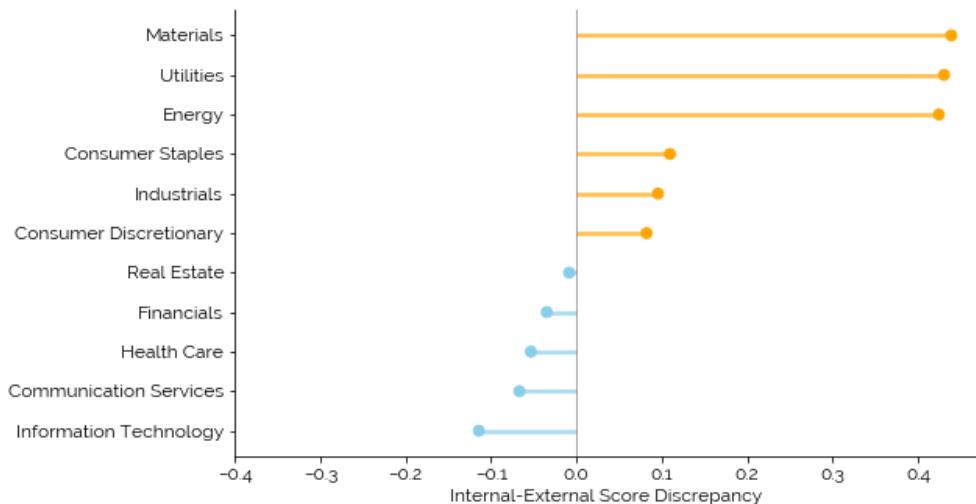


Figure 44: average difference between internal and external score by Sector

To conclude, let us look at the final FinScience Sustainability Score that emerges from the combination of the so-discussed sub-scores. More specifically, as detailed in the methodology chapter, the Score is computed as the weighted average of the 17 overall SDGs scores, based on the Sector that the company operates in and according to the materiality analysis. The 17 overall SDGs scores are in turn obtained as the average of the internal and external score for each Goal, penalized by the difference between the two. The Score ranges from 42 to 80, with an average of 66 and a standard deviation of 6

points. Because of missing data, it was not always possible to compute the Score for *SDG 17 – Partnerships for the Goals*, which comes as no surprise as the declaration of specific agreements for reaching UN Sustainable Development Goals is not compulsory nor common and also not easy to detect. More specifically, the overall SDG 17 Score was calculated for half of the companies. On average, the highest SDG Score, including the penalization term, is reached by *SDG 4 – Quality Education*, while the lowest is the score for *SDG 12 – Responsible Consumption and Production*. In general, based on the resulting scores and on the external and internal data sources, S&P 500 companies tend to perform better in Social and Governance matters than on Environmental issues (*Figure 45*).

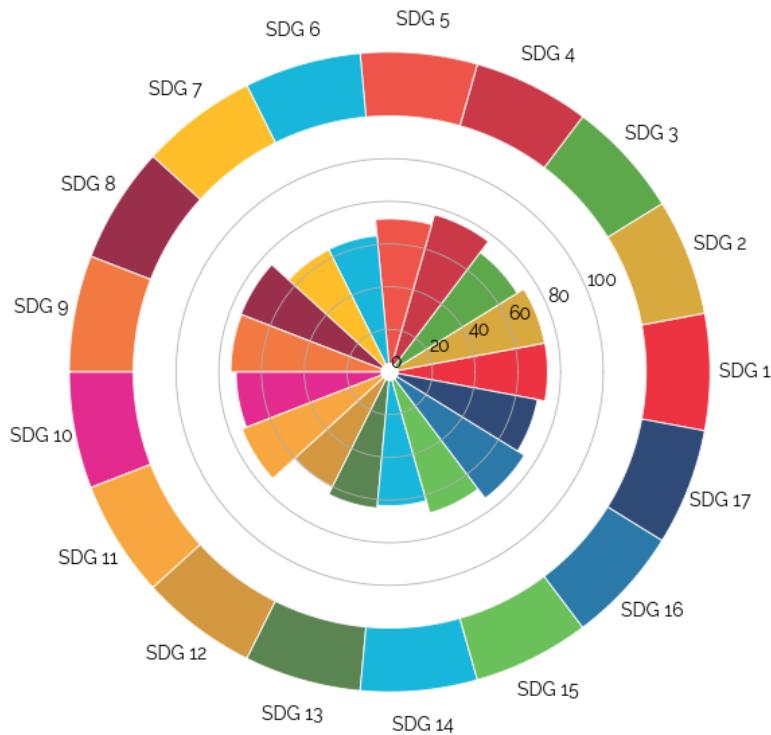


Figure 45: average Final SDGs Scores on S&P500 companies

The following tables show the 5 corporations at the bottom (*Table 7*) and the 5 corporations at the top (*Table 8*) of the S&P 500 companies based on the final score obtained.

Company	Sector	FinScience Sustainability Score
Corteva Inc	Materials	42
Lennar Corp	Consumer Discretionary	45
Charter Communications Inc	Communication Services	46
Essex Property Trust Inc	Real Estate	48
Copart Inc	Industrials	49

Table 7: bottom S&P500 companies according to FinScience Sustainability Score

Company	Sector	FinScience Sustainability Score
NVIDIA Corp	Information Technology	79
salesforce.com Inc	Information Technology	79
Principal Financial Group Inc	Financials	79
Mosaic Co/The	Materials	79
Prudential Financial Inc	Financials	78

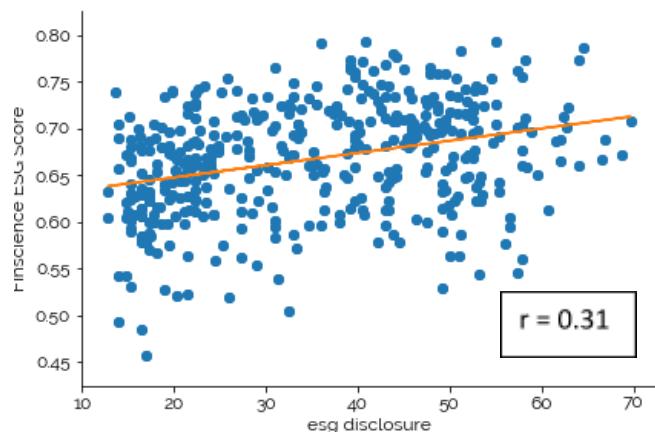
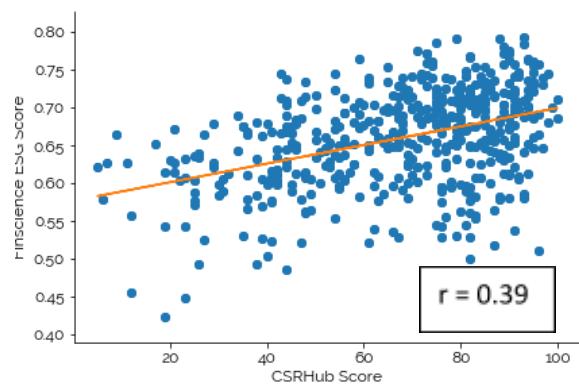
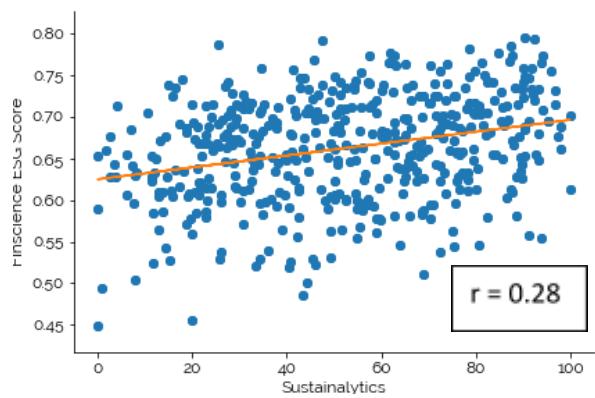
Table 8: top S&P500 companies according to the FinScience Sustainability Score

To further evaluate the Score, we decided to compare it to other existing sustainability rating systems, namely Sustainalytics⁷⁹, by Morningstar, CSRHub Rating⁸⁰, by CSRHub, and ESG Disclosure⁸¹, by Bloomberg. We found slightly positive correlation values between FinScience Sustainability Scores and such ratings on S&P 500 companies (*Figure 46*). On the one hand, these results suggest that the brand-new Score presented in this thesis and obtained from the data collected through several automatized processes has generated results somehow similar to the findings reported by three famous sustainability evaluation providers. On the other hand, the lack of high correlation values reinforces the idea that the FinScience Sustainability Score provides a novel approach to sustainability assessment, by taking the UN Sustainable Development Goals as main reference and by including a broad range of data sources. In this regard, we found that the overall Internal Score is indeed more correlated to the aforementioned rating systems than the final Score. This indicates that the inclusion of external data sources has in fact added value, by shading lights on factors and variables that are excluded by traditional tools.

⁷⁹ <https://www.sustainalytics.com/esg-ratings/>

⁸⁰ <https://www.csrhub.com/>

⁸¹ <https://www.bloomberg.com/company/press/bloomberg-launches-proprietary-esg-scores/>



Conclusion

The presented work was intended to illustrate the corporate sustainability performance evaluation tool that was developed during my 9-month internship as a data scientist at FinScience. Several tools are already available in the market to assess the Environmental, Social and Governance performance of organizations and to support investors with Socially Responsible Investing strategies. However, the novel tool that was detailed in this thesis is specifically designed to measure corporate contributions to the Sustainable Development Goals defined by the United Nations in 2015, and therefore it aims at providing a more granular and exhaustive approach to sustainability, with a specific attention to global challenges. In addition, one of the main innovative features of the Score is the duality of the sources considered, as they relate to both the information disclosed by the companies themselves and the data provided by third parties. This distinction was drawn to monitor at the same time the image that companies want to convey on sustainable issues and the public reception of such activities. What is more, the data considered not only derives from different types of sources, but it also pertains to two main categories: traditional and alternative data. While traditional data, such as gas emission, adoption of specific ESG policies, incident rates and so on, carry indispensable knowledge for sustainability assessment, alternative data, such as news sentiment, public rankings, consumer and employees reviews, text classification of CSR reports and corporate websites sections, shed lights on hidden facts that might in fact reveal a different picture.

My role as an intern was to manage the collection processes needed to retrieve data from hundreds of data sources in an automatic fashion for the 500 companies that constitute the S&P 500 index, namely the largest American corporations. Web scraping was a central part of the job, together with data manipulation. It was indeed necessary to

transform large amounts of textual data with means of Natural Language Processing tools and text mining. An example of this work was the construction of a text classifier to identify the SDGs being addressed in CSR reports and corporate websites. My role was also to analyze the data extracted and to combine it into final scores that would summarize the corporate performance on each Goal, as illustrated throughout the last chapter of the thesis. From the numerous findings that emerged from the analysis, it can be concluded that S&P 500 companies' commitment to sustainable development varies in terms of Goals, self-disclosure, public reception and type of source, with some evident differences between sectors. In general, the worst performances relate to the Environmental topics, and specifically to the production and consumption patterns (SDG 12), but also to climate (SDG 13) change and renewable energy production (SDG 7). It is important to emphasize that environment-related keywords are yet the most frequent in corporate websites and reports, which suggest that while companies are aware of the importance of these issues and interested in these themes, there is still a need for more tangible action. On the other hand, companies tend to perform well on social issues, such as training and education (SDG 4) and well-being (SDG 3), and governance (SDG 16), although there is still a long way to go on diversity and inclusion subjects, as the discussion on the gender wage gap illustrated (SDG 5, 10).

It should not be ignored that this study presents some limitations that might have affected the presented results. For instance, as for the underlying methodology, it has to be recalled that each variable has a binary relationship with each SDG, meaning that it can either contribute or not to that Goal, with the contribution being expressed by its weight. However, if a variable contributes several Goals, then the same weight applies to all of them, as if it contributed evenly and to the same extent to different issues. As a consequence, some Goals might have been treated equally with respect to a specific

variable although this may not be the case in reality. Moreover, it has to be recalled that the score was built to measure corporate performance with respect to SDGs at a Goal level, and therefore it does not provide a breakdown of corporate contributions to the 169 targets. However, the descriptions of the targets were considered when defining the methodology and selecting the variables. Lastly, it might be argued that the score does not consider the specific socio-economical aspects of the geographical area in which organizations are located, and therefore it does not allow to monitor and quantify the business impact on such factors. However, such an approach would better suit governmental organizations' needs and is therefore beyond the scope of the presented tool, which was specifically designed for investors and stakeholders in the first place. In addition to this, all the companies considered are American, and therefore it would have not been possible to notice any difference in this sense. To conclude, the most challenging part was the automatization of the data collection processes. Although specific pipelines were developed for different sources and data underwent cleaning processes, statistical analysis and also manual checks, the data collection was often made in bulk, with criteria that might leave behind relevant data or include incorrect information. This might be the case for keywords extraction as well as text classification, that cannot reach a perfect accuracy by nature. Also, a more extensive analysis and manipulation shall be performed in future to understand whether, and to what extent, the sub-scores are in fact comparable and to detect what contributes the most to their final values.

While these themes might be subject of further discussion and there might be place for more in-depth analysis, the scoring system as a whole and the results presented can be considered reliable and valid as they reflected a methodology that was based on the most recent literature and was extensively reviewed by FinScience's team of ESG specialists, data scientists and investment specialists.

References

- Betti, G., Consolandi, C., & Eccles, R. G. (2018). The Relationship between Investor Materiality and the Sustainable Development Goals: A Methodological Framework. *Sustainability*, 10(7). Tratto da <https://doi.org/10.3390/su10072248>
- Carroll, A. B. (1999). Corporate Social Responsibility: Evolution of a Definitional Construct. *Business & Society*, 38(3), 268-295. Tratto da <https://doi.org/10.1177/000765039903800303>
- CFA Insititute. (2015). *Environmental, soocial and governance issues in investing. A Guide for Investment Professionals*. Tratto da <https://www.cfainstitute.org/-/media/documents/article/position-paper/esg-issues-in-investing-a-guide-for-investment-professionals.ashx>
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New Talent Signals: Shiny New Objects or a Brave New World? *Industrial and Organizational Psychology*, 9(3), 621-640.
- Crittenden, V., Crittenden, W. F., Ferrell, L., Ferrell, O., & Pinney, C. (2011). Market-oriented sustainability: A conceptual framework and propositions. *Journal of the Academy of Marketing Science*, 39(1), 71-85. Tratto da <https://doi.org/10.1007/s11747-010-0217-2>
- Diez-Cañamero, B., Bishara, T., Otegi-Olaso, J., Minguez, R., & Fernández, J. (2020). Measurement of Corporate Social Responsibility: A Review of Corporate Sustainability Indexes, Rankings and Ratings. *Sustainability*, 12(5). Tratto da <https://doi.org/10.3390/su12052153>
- Dörgő, G., Sebestyén, V., & Abonyi, J. (2018). Evaluating the Interconnectedness of the Sustainable Development Goals Based on the Causality Analysis of Sustainability Indicators. *Sustainability*, 10(3766). Tratto da <https://doi.org/10.3390/su10103766>
- Du Pisan, J. (2007). Sustainable development – historical roots of the concept. *Environmental Sciences*, 83-96. doi:10.1080/15693430600688831
- Duran-Silva, N., Fuster, E., Masucci, F., & Quinquillà, A. (2019, December). *A controlled vocabulary defining the semantic perimeter of Sustainable Development Goals*. doi:10.5281/zenodo.3567769
- Elkington, J. (1992). Towards a Sustainable Corporation: Win–Win–Win Business Strategies for Sustainable Development. *California Management Review*, 36(2), 90–100. Tratto da <https://doi.org/10.2307/41165746>
- European Commission. (2020). *Summary Report of the Public Consultation on the Review of the Non-Financial Reporting Directive*.
- Fehling, M., Nelson, B. D., & Venkatapuram, S. (2013). Limitations of the Millennium Development Goals: a literature review. *Global Public Health*, 8(10), 1109-1122. doi:10.1080/17441692.2013.845676

- Global Impact Investing Network. (2016). *Achieving the Sustainable Development Goals: The Role of Impact Investing*. Tratto da
[https://gsgii.org/reports/achieving-the-sustainable-development-goals-the-role-of-impact-i
nvesting/](https://gsgii.org/reports/achieving-the-sustainable-development-goals-the-role-of-impact-investing/)
- Global Reporting Initiative, United Nations Global Compact, WBCSD. (2016). *Linking the SDGs and GRI - GRI Standards*. Tratto da
https://www.globalreporting.org/Documents/ARCHIVES/resource%20library/SDG_GRI_LInkage.pdf
- Global Reporting Initiative, United Nations Global Compact. (2017). *Business Reporting on the SDGs: An Analysis of the Goals and Targets*. Tratto da
https://www.globalreporting.org/resourcelibrary/GRI_UNGC_Business-Reporting-on-SDGs_Analysis-of-Goals-and-Targets.pdf
- Global Sustainable Investment Alliance. (2018). *Global Sustainable Investment Review*. Tratto da
http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf
- Goh, C. S., Chong, H.-Y., Jack, L., & Mohd Faris, A. F. (2020). Revisiting triple bottom line within the context of sustainable construction: A systematic review. *Journal of Cleaner Production*, 252. Tratto da <https://doi.org/10.1016/j.jclepro.2019.119884>
- Goyal, P., Pandey, S., & Jain, K. (2018). *Deep learning for natural language processing: Creating neural networks with Python*. Berkeley, CA: Apress.
- Harper Ho, V. (2020). Non-Financial Reporting & Corporate Governance: Explaining American Divergence & Its Implications for Disclosure Reform. *Accounting, Economics, and Law: A Convivium*.
- Hill, R. P., Ainscough, T., Shank, T., & Manullang, D. (2007). Corporate Social Responsibility and Socially Responsible Investing: A Global Perspective. *Journal of Business Ethics*, 70, 165–174.
- Ho, A., Nguyen, A., Pafford, J. L., & Slater, R. (2019). A Data Science Approach to Defining a Data Scientist. *SMU Data Science Review*, 2(3). Tratto da
<https://scholar.smu.edu/datasciencereview/vol2/iss3/4>
- Hristov, I., & Chirico, A. (2019). The Role of Sustainability Key Performance Indicators (KPIs) in Implementing Sustainable Strategies. *Sustainability*, 11(20). doi:10.3390/su11205742
- Huang, C., Qiu, X., & Huang, X. (2014). Text classification with Document Embeddings. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data* (p. 131-140). Sun M., Liu Y., Zhao J. Tratto da
https://doi.org/10.1007/978-3-319-12277-9_12
- Huber, B. M., Comstock, M., & Smith, H. (2018, October 4). UN Sustainable Development Goals - The Leading ESG Framework for Large Companies. *Harvard Law School Forum on Corporate Governance*. Tratto da
<https://corpgov.law.harvard.edu/2018/10/04/un-sustainable-development-goals-the-leading-esg-framework-for-large-companies/>

- Investor Responsibility Research Center Institute (IRRCI). (2018). *State of Sustainability and Integrated Reporting 2018*. Tratto da <https://www.weinberg.udel.edu/IIRCIResearchDocuments/2018/11/2018-SP-500-Integrated-Reporting-FINAL-November-2018-1.pdf>
- Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing Imbalanced Data. Recommendations for the Use of Performance Metrics. *International Conference on Affective Computing and Intelligent Interaction and workshops : [proceedings]. ACII (Conference)*, 245-251.
- Johnson, M. E., Albizri, A., & Jain, R. (2020). Exploratory Analysis to Identify Concepts, Skills, Knowledge, and Tools to Educate Business Analytics Practitioners. *Decision Sciences Journal of Innovative Education*, 18(1). Tratto da <https://doi.org/10.1111/dsji.12195>
- Khan, M., Serafeim, G., & Yoon, A. (2015). Corporate Sustainability: First Evidence on Materiality. Tratto da <https://dash.harvard.edu/bitstream/handle/1/14369106/15-073.pdf>
- Lang, C., Xia, S., & Liu, C. (2020). Style and fit customization: a web content mining approach to evaluate online mass customization experiences. *Journal of Fashion Marketing and Management*. Tratto da <https://doi.org/10.1108/JFMM-12-2019-0288>
- Liew, W., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry*, 65(3), 393-400. Tratto da <https://doi.org/10.1016/j.compind.2014.01.004>
- Littlechild, S. (2020). Online reviews and customer satisfaction: The use of Trustpilot by UK retail energy suppliers and three other sectors. *Cambridge Working Papers in Economics*.
- Lomazzi, M., Borisch, B., & Laaser, U. (2014). The Millennium Development Goals: experiences, achievements and what's next. *Global Health Action*, 7(1). doi:10.3402/gha.v7.23695
- Luo, N., Zhou, Y., & Shon, J. J. (2016). Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.com. *Proceedings of the 37th International Conference on Information Systems*. Tratto da <https://aisel.aisnet.org/icis2016/DataScience/Presentations/15/>
- Mikolov, T., Chen, K., Corrado, G., & Dea, J. (2013). Efficient Estimation of Word Representations in Vector Space. *CoRR*. Tratto da <https://arxiv.org/abs/1301.3781>
- Mooij, S. (2017). The ESG Rating and Ranking Industry; Vice or Virtue in the Adoption of Responsible Investment? *Social Science Research Network*.
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: global vectors for word representation. *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Tratto da <http://dx.doi.org/10.3115/v1/d14-1162>
- Scheyvens, R., Banks, G., & Hughes, E. (2016). The Private Sector and the SDGs: The Need to Move Beyond 'Business as Usual'. *Sustainable Development*. Tratto da <https://doi.org/10.1002/sd.1623>
- Schramade, W. (2017). Investing in the UN Sustainable Development Goals: Opportunities for Companies and Investors. Tratto da <https://doi.org/10.1111/jacf.12236>

- Sozzi, A. (2014). *Sustainability Reports [MSc Data Science Thesis]*. Royal Holloway University of London, Department of Computer Science, Egham, Surrey TW20 0EX, UK.
- Sroufe, R. (2016). Operationalizing Sustainability. *Journal of Sustainability Studies*.
- Sroufe, R. (2017). Integration and organizational change towards sustainability. *Journal of Cleaner Production*, 162(20), 315-319. Tratto da <https://doi.org/10.1016/j.jclepro.2017.05.180>
- Suen, H.-Y., Hung, K.-E., & Tseng, F.-H. (2020). Employer Ratings through Crowdsourcing on Social Media: An Examination of U.S. Fortune 500 Companies. *Sustainability*, 12(6), 6308. Tratto da <https://doi.org/10.3390/su12166308>
- Sustainability Accounting Standards Board (SASB). (2017). SASB Counceptual Framework. Tratto da <https://www.sasb.org/wp-content/uploads/2019/05/SASB-Conceptual-Framework.pdf>
- Székely, N., & vom Brocke, J. (2017). What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. *PLoS ONE*, 12(4). Tratto da <https://doi.org/10.1371/journal.pone.0174807>
- Szennay, Á., Szigeti, C., Kovács, N., & Szabó, D. R. (2019). Through the Blurry Looking Glass—SDGs in the GRI Reports. 8(2). Tratto da <https://doi.org/10.3390/resources8020101>
- Tinbergen , N. (1952). "Derived" Activities; Their Causation, Biological Significance, Origin, and Emancipation During Evolution. *The Quarterly Review of Biology*, 27(1), 1-32. doi:<https://doi.org/10.1086/398642>
- UN General Assembly. (2015). Transforming our world : the 2030 Agenda for Sustainable Development. Tratto da <https://www.refworld.org/docid/57b6e3e44.html>
- UN Global Compact. (2004). *Who Cares Wins. Connecting Financial Markets to a Changing World*. United Nations Department of Public Information.
- United Nations, The Global Compact. (2004). *Who cares wins: Connecting financial markets to a changing world*. Geneva, Switzerland. Tratto da https://www.unglobalcompact.org/docs/issues_doc/Financial_markets/who_cares_who_wins.pdf
- Vörösmarty, C., Rodríguez Osuna, V., Koehler, D., Klop, P., Spengler, J., Buonocore, J. J., . . . Sánchez, R. (2018). Scientifically assess impacts of sustainable investments. *Science*, 523-525. doi:[10.1126/science.aoa3895](https://doi.org/10.1126/science.aoa3895)
- Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable Development: A Bird's Eye View. *Sustainability*, 3(10), 1637-1661. Tratto da <https://doi.org/10.3390/su3101637>
- WCED. (1987). *Report of the World Commission on Environment and Development: Our Common Future (Brundtland Report)*. Tratto da https://sswm.info/sites/default/files/reference_attachments/UN%20WCED%201987%

Annexes

Annex A. SASB Standards – SDGs Mapping

		S D G 1	S D G 2	S D G 3	S D G 4	S D G 5	S D G 6	S D G 7	S D G 8	S D G 9	S D G 0	S D G 1	S D G 2	S D G 3	S D G 4	S D G 5	S D G 6	S D G 7
Environment	GHG Emissions									1		1			1			
	Air Quality				1									1	1			
	Energy Management			1						1		1			1	1		
	Water & Wastewater Management		1	1				1					1	1		1	1	
	Waste & Hazardous Materials Management				1				1				1	1		1	1	
	Ecological Impacts		1						1				1	1		1	1	
Social Capital	Human Rights & Community Relations	1							1		1		1	1				1
	Customer Privacy																	1
	Data Security																	1
	Access & Affordability	1		1		1			1		1		1	1	1	1		
	Product Quality & Safety		1	1	1													1
	Customer Welfare		1	1	1			1										1
	Selling Practices & Product Labeling	1	1	1	1													1
Human Capital	Labor Practices	1		1		1	1			1			1	1				1
	Employee Health & Safety			1		1				1								
	Employee Engagement, Diversity & Inclusion				1	1				1		1						
Business Model & Innovation	Product Design & Lifecycle Management		1	1			1	1	1	1			1	1	1	1	1	1
	Business Model Resilience														1			1
	Supply Chain Management	1		1		1	1	1	1		1		1	1	1	1	1	1
	Materials Sourcing & Efficiency			1		1	1		1				1	1		1	1	1
	Physical Impacts of Climate Change		1				1	1					1	1	1	1	1	
Leadership & Governance	Business Ethics																	1
	Competitive Behavior	1							1		1	1						
	Management of the Legal & Regulatory Environment		1					1					1	1				1
	Critical Incident Risk Management	1		1			1		1				1		1	1	1	
	Systemic Risk Management							1		1	1							1

Annex B1. INTERNAL DATA: Report

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Report	Presenza sdg 1 nel report (classificazione)	sdg_1_report	4	False	1
Report	Presenza sdg 2 nel report (classificazione)	sdg_2_report	4	False	2
Report	Presenza sdg 3 nel report (classificazione)	sdg_3_report	4	False	3
Report	Presenza sdg 4 nel report (classificazione)	sdg_4_report	4	False	4
Report	Presenza sdg 5 nel report (classificazione)	sdg_5_report	4	False	5
Report	Presenza sdg 6 nel report (classificazione)	sdg_6_report	4	False	6
Report	Presenza sdg 7 nel report (classificazione)	sdg_7_report	4	False	7
Report	Presenza sdg 8 nel report (classificazione)	sdg_8_report	4	False	8
Report	Presenza sdg 9 nel report (classificazione)	sdg_9_report	4	False	9
Report	Presenza sdg 10 nel report (classificazione)	sdg_10_report	4	False	10
Report	Presenza sdg 11 nel report (classificazione)	sdg_11_report	4	False	11
Report	Presenza sdg 12 nel report (classificazione)	sdg_12_report	4	False	12
Report	Presenza sdg 13 nel report (classificazione)	sdg_13_report	4	False	13
Report	Presenza sdg 14 nel report (classificazione)	sdg_14_report	4	False	14
Report	Presenza sdg 15 nel report (classificazione)	sdg_15_report	4	False	15
Report	Presenza sdg 16 nel report (classificazione)	sdg_16_report	4	False	16
Report	Presenza sdg 17 nel report (classificazione)	sdg_17_report	4	False	17
Report	Presenza di report di sostenibilità	sustainability_report	8	True	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
Report	Anni di storico del report di sostenibilità	corp_reg_history	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
Report	Presenza di un rapporto integrato	Integrated	9	False	16
Report	ISO 8528	ISO 8528	8	False	11, 12, 13
Report	DIRECTIVE 2008/98/EC	DIRECTIVE 2008/98/EC	8	False	11, 12, 14, 15
Report	ISO 14044	ISO 14044	8	False	12, 13
Report	ISO 55001	ISO 55001	8	False	6, 7, 8, 9, 11, 12, 13
Report	ISO 14046	ISO 14046	8	False	6, 13, 14
Report	ISO 14021	ISO 14021	8	False	13

Report	ISO 22716	ISO 22716	8	False	3, 8
Report	FOE	FOE	8	False	2, 12, 13, 15
Report	ISO 16128	ISO 16128	8	False	3, 13
Report	H&S POLICIES	H&S POLICIES	8	False	3, 8
Report	INTERNATIONAL FINANCE CORPORATION	INTERNATIONAL FINANCE CORPORATION	8	False	8
Report	TCFD	TCFD_certification	8	False	13
Report	MARINE STEWARDSHIP COUNCIL	MSC	8	False	2, 6, 14
Report	ISO 31000	ISO 31000	8	False	3, 8, 9, 11, 14, 15, 16
Report	ISO 26000	ISO 26000	8	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
Report	ISO 9001	ISO 9001	8	False	1, 9
Report	ISO 50001	ISO 50001	8	False	7, 11, 12, 13
Report	ISO 37001	ISO 37001	8	False	8, 10, 11, 16
Report	FSC	FSC/PEFC	8	False	12, 13, 14, 15
Report	ISO 14001	ISO 14001	8	False	1, 2, 3, 4, 6, 7, 8, 9, 12, 13, 14, 15
Report	ISO 22301	ISO 22301	8	False	6, 7, 8, 9, 11, 16
Report	ISO 14040	ISO 14040	8	False	13
Report	ISO 20121	ISO 20121	8	False	3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16
Report	ISO 13485	ISO 13485	8	False	3, 10
Report	ISO 22000	ISO 22000	8	False	2, 3, 12
Report	ISO 45001	ISO 45001	8	False	3, 5, 8, 9, 10, 11, 16
Report	ISO 10993	ISO 10993	8	False	3
Report	ISO 14025	ISO 14025	8	False	13
Report	ISO 14064	ISO 14064	8	False	9, 13
Report	ISCC PLUS	ISCC PLUS	8	False	2, 6, 7, 9, 12, 13
Report	Third-party assurance on SR	assurances	9	True	16
Report	GRI application (look at GRI table and levels)	GRI	8	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16

Annex B2. INTERNAL DATA: Corporate Website

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Website	Presenza sdg 1 nel sito (classificazione)	sdg_1_website	4	False	1
Website	Presenza sdg 2 nel sito (classificazione)	sdg_2_website	4	False	2
Website	Presenza sdg 3 nel sito (classificazione)	sdg_3_website	4	False	3
Website	Presenza sdg 4 nel sito (classificazione)	sdg_4_website	4	False	4
Website	Presenza sdg 5 nel sito (classificazione)	sdg_5_website	4	False	5
Website	Presenza sdg 6 nel sito (classificazione)	sdg_6_website	4	False	6
Website	Presenza sdg 7 nel sito (classificazione)	sdg_7_website	4	False	7
Website	Presenza sdg 8 nel sito (classificazione)	sdg_8_website	4	False	8
Website	Presenza sdg 9 nel sito (classificazione)	sdg_9_website	4	False	9
Website	Presenza sdg 10 nel sito (classificazione)	sdg_10_website	4	False	10
Website	Presenza sdg 11 nel sito (classificazione)	sdg_11_website	4	False	11
Website	Presenza sdg 12 nel sito (classificazione)	sdg_12_website	4	False	12
Website	Presenza sdg 13 nel sito (classificazione)	sdg_13_website	4	False	13, 14
Website	Presenza sdg 14 nel sito (classificazione)	sdg_14_website	4	False	14
Website	Presenza sdg 15 nel sito (classificazione)	sdg_15_website	4	False	15
Website	Presenza sdg 16 nel sito (classificazione)	sdg_16_website	4	False	16
Website	Presenza sdg 17 nel sito (classificazione)	sdg_17_website	4	False	17

Website	Presenza sezione sul sito corporate responsabi...	sustainability_csr_section	8	True	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Website	Presenza di una sezione corporate governance	corporate_governance_section	3	False	5, 16
Website	Presenza su organizzazione e politiche remuner...	remuneration_section	8	False	16
Website	Presenza di sezione diversity	diversity_section	8	False	5, 10, 16

Annex B3. INTERNAL DATA: Memberships & Proposal

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Sito CDP	adesione CDP Climate Change	CDP_climate_change	9	False	7, 12, 13, 15, 16, 17
Sito CDP	adesione CDP Water Security	CDP_water_security	9	False	6, 12, 14, 16, 17
Sito CDP	adesione CDP Forest	CDP_forests	9	False	12, 13, 15, 16, 17
AAFA/FLA	AAFA/FLA Apparel & Footwear Industry Commitmen...	AAFA/FLA	7	False	8, 17
Accord on Fire and Safety in Bangladesh	Accord on Fire and Safety in Bangladesh - Sign...	NaN	7	False	None
Afirm	AFIRM Group Members	Afirm	7	False	3, 9, 12, 17
RCMS	American Chemistry Responsible Care Measuremen...	RCMS	7	False	6, 7, 8, 12, 13, 16, 17
ACGA	Asian Corporate Governance Association (ACGA)	ACGA	7	False	16, 17
B Corporation	B Corporation GIIRS Ratings	NaN	7	False	None
BSR Member	BSR Member	BSR Member	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
CBSR	Canadian Business for Social Responsibility	NaN	7	False	None
Carbonfund	Carbonfund Partners	Carbonfund	7	False	4, 7, 13, 17
CERES	Ceres - Network Members	CERES	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...

Circular Economy 100	Circular Economy 100 - Members	Circular Economy 100	7	False	4, 11, 12, 13, 17
BCCCC	Boston College Center for Corporate Citizenship...	BCCCC	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
CECP	CECP (Chief Executives for Corporate Purpose)	CECP	7	False	16, 17
Climate registry	Climate Registry (anche dati raw)	Climate registry	7	False	7, 13, 17
CHRB	Corporate Human Rights Benchmark 2019	CHRB	7	False	1, 4, 5, 8, 10, 16, 17
CSR Europe	CSR Europe - corporate members	CSR Europe	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
CSRwire	CSRwire member companies	CSRwire	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Global Compact	Global Compacts - Participants	Global Compact	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
IAG	Investors Against Genocide	IAG	7	False	1, 2, 4, 5, 8, 10, 16, 17
EDF	EDF Climate Corps 2019	EDF	7	False	7, 13, 17
EMAS	EMAS	NaN	7	False	None
EP	Equator Principles	EP	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
FLA	Fair Labor Association	FLA	7	False	1, 2, 3, 4, 5, 16, 17
FSC	Forest Stewardship Council	FSC_membership	7	False	15, 17
GBC	Global Business Coalition	GBC	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
GLOBAL G.A.P	Global G.A.P. Member	GLOBAL G.A.P (Associate)	7	False	2, 3, 12, 17
GRI	Global Reporting Initiative 2019	NaN	7	False	None
GRESB	GRESB (Global Real Estate Sustainability Bench...	GRESB	7	False	9, 11, 17
ImpactSpace	ImpactSpace	ImpactSpace	7	False	16, 17
NAEM	NAEM Corporate Member	NAEM	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Oceana	Oceana Partners	Oceana	7	False	12, 14, 17

EP100	The Climate Group EP100	EP100	7	False	7, 9, 13, 17
POP	Pledge One Percent	POP	7	False	11, 17
TCFD	TCFD supporters	TCFD_membership	7	False	7, 13, 16, 17
RE100	RE100	RE100	7	False	7, 13, 17
RBA	Responsible Business Alliance	RBA	7	False	16, 17
SPLC	Sustainable Purchasing Leadership Council	SPLC	7	False	11, 12, 16, 17
SPC	Sustainable Packaging Coalition	SPC	7	False	12, 13, 15, 17
SB	Sustainable Brands Members	SB	7	False	12, 16, 17
SAC	Sustainable Apparel Coalition	SAC	7	False	1, 5, 8, 9, 10, 11, 12, 17
SASB	SASB Industry Group Members	SASB	7	False	17
SASB	Reporting secondo i SASB Standards	NaN	7	False	None
SC	The Sustainability Consortium member	SC	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
We Mean Business	We Mean Business - partners	NaN	7	False	None
WEConnectInternational	WEConnect International - Members	WEConnectInternational	7	False	5, 10, 17
WBENC	Women's Business Enterprise National Council (...)	WBENC	7	False	5, 10, 17
WBCSD	World Business Council for Sustainable Development...	WBCSD	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
WCAM	World Coal Association Members	WCAM	7	False	7, 13, 17
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 (Ethi...)	ceres_fin	7	False	16
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 (Climate...	ceres_env	7	False	6, 12, 13, 14, 15
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 \n(Human...	ceres_hum	7	False	1, 2, 3, 4, 12

Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 \n(He...	ceres_health	7	False	3
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 \n(Dec...	ceres_work	7	False	3, 8
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 \n(Wor...	ceres_div	7	False	5, 8, 10
Ceres Engagement Tracker	Ceres Climate & Energy Resolutions 2018 \n(Sus...	ceres_sust	7	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Annex B4. INTERNAL DATA: Balance Sheet

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Bloomberg ESG (E) 2	Direct CO2 Emissions	E_co2	8	True	7, 12, 13
Bloomberg ESG (E) 4	Renewable energy use	E_renew	8	False	7, 12, 13
Bloomberg ESG (E) 5	Electricity used	E_elec	8	False	7, 12, 13
Bloomberg ESG (E) 6	Total GHG Emissions (tonnes)	E_ghg	8	False	7, 12, 13
Bloomberg ESG (E) 7	Methane Emissions (Th Tonnes)	E_methane	8	True	7, 12, 13
Bloomberg ESG (E) 8	Nitrogen Oxide Emissions (Tonnes)	E_no_emiss	8	True	7, 12, 13
Bloomberg ESG (E) 9	Sulphur Dioxide Emissions (Tonnes)	E_so_emiss	8	True	7, 12, 13
Bloomberg ESG (E) 10	VOC Emissions (Tonnes)	E_voc_emiss	8	True	7, 12, 13
Bloomberg ESG (E) 11	Gas Flaring (Tonnes)	E_gas	8	False	7, 12, 13
Bloomberg ESG (E) 12	Total Energy Consumption	E_energy	8	False	7, 12, 13
Bloomberg ESG (E) 13	Total water use	E_water	8	False	6, 12, 14
Bloomberg ESG (E) 14	Total Waste	E_waste	8	False	12, 15
Bloomberg ESG (E) 15	Waste recycled	E_waste_rec	8	False	12, 15

Bloomberg ESG (E) 16	% Water recycled	E_water_rec	8	False	6, 12, 14
Bloomberg ESG (E) 17	Discharges to Water	E_disch	8	False	6, 12, 14
Bloomberg ESG (E) 18	Hazardous Waste (Tonnes)	E_haz_waste	8	False	12, 13
Bloomberg ESG (E) 19	Number of Spills	E_spill	8	False	6, 12, 13, 14
Bloomberg ESG (E) 20	Amount of Spills (Tonnes)	E_spill_am	8	False	6, 12, 13, 14
Bloomberg ESG (E) 22	Environmental Fines	E_fine	8	False	12, 13, 16
Bloomberg ESG (E) 24	Investments in operational sustainability	E_inv_op	8	True	9, 12, 13, 15, 16
Bloomberg ESG (E) 24bis	Presence of Investments in operational sustain...	E_inv_op_pres	8	True	9, 12, 13, 15, 16
Bloomberg ESG (E) 25	Renewable Electricity Target Policy	E_elec_pol	8	True	7, 12, 13
Bloomberg ESG (E) 26	Energy Efficiency Policy	E_ener_pol	8	True	7, 12, 13, 16
Bloomberg ESG (E) 27	Emission Reduction Initiatives Policy	E_emiss_pol	8	True	7, 9, 12, 13, 16
Bloomberg ESG (E) 28	Environmental Supply Chain management	E_supply_pol	8	True	12, 13, 16
Bloomberg ESG (E) 29	Environmental Quality Management	E_env_pol	8	True	12, 13, 15, 16
Bloomberg ESG (E) 30	Green Building Policy	E_build_pol	8	True	9, 11, 12, 13, 16
Bloomberg ESG (E) 31	Sustainable Packaging	E_pack_pol	8	True	9, 12, 13, 16
Bloomberg ESG (E) 32	Waste reduction policy	E_waste_pol	8	True	9, 12, 13, 15, 16
Bloomberg ESG (E) 33	Water policy	E_water_pol	8	True	6, 9, 12, 14, 16
Bloomberg ESG (E) 34	Biodiversity Policy	E_bio_pol	8	True	9, 12, 13, 15, 16
Bloomberg ESG (E) 35	Climate Change Policy	E_cc_pol	8	True	7, 9, 12, 13, 16

Bloomberg ESG (E) 36	New Products - climate change	E_new_pol	8	True	9, 12, 13, 16
Bloomberg ESG (E) 37	Climate Change opportunities discussed	E_cc_disc	8	True	13, 16
Bloomberg ESG (E) 38	Risks of climate change discussed	E_cc_risk	8	True	7, 13, 16
Bloomberg ESG (E) 40	Carbon Dioxide Intensity per EBITDA	E_c02_ebidta	8	False	7, 12, 13
Bloomberg ESG (S) 1	Ratio of Women in Total Workforce to Women in ...	S_ratio_women	8	False	5, 10
Bloomberg ESG (S) 3	Women employees %	S_women_emp	8	True	5, 10
Bloomberg ESG (S) 4	Employee Turnover %	S_turnover	8	False	8
Bloomberg ESG (S) 5	Employees Unionized %	S_union	8	True	8
Bloomberg ESG (S) 6	Lost Time incident rate	S_inc_time	8	False	3, 8
Bloomberg ESG (S) 7	Total incident rate	S_inc_rate	8	False	3, 8
Bloomberg ESG (S) 8	Fatalities per 1000 employees	S_fatal	8	False	3, 8
Bloomberg ESG (S) 9	Community Spending/Profit Before Tax	S_community	8	False	1, 10, 11
Bloomberg ESG (S) 10	Median Gender Pay Gap	S_gender_pay_gap	8	False	8, 10
Bloomberg ESG (S) 11	Training spending per Employee	S_training	8	False	4, 8
Bloomberg ESG (S) 12	Responsible Supply Chain	S_supply_mgmt	8	False	8, 16
Bloomberg ESG (S) 13	Sustain Sup Guidelines Encomp ESG	S_esg_guide	8	False	8, 16
Bloomberg ESG (S) 14	Anti-bribery Ethics Policy	S_antibribery	8	True	8, 16
Bloomberg ESG (S) 15	Business Ethics Policy	S_ethics	8	True	8, 12, 16
Bloomberg ESG (S) 16	Employee CSR Training	S_CSR_training	8	True	4, 8
Bloomberg ESG (S) 17	Employee Protection/Whistleblower policy	S_whistlebl	8	True	8, 16

Bloomberg ESG (S) 18	Equal opportunity policy	S_equal_opp	8	True	5, 8, 10, 16
Bloomberg ESG (S) 19	Fair remuneration policy	S_remun_pol	8	True	5, 8, 10, 16
Bloomberg ESG (S) 20	Health and safety policy	S_heal_pol	8	True	8, 16
Bloomberg ESG (S) 21	Human rights policy	S_hr_pol	9	True	1, 2, 3, 5, 6, 8, 10, 16
Bloomberg ESG (S) 22	Training policy	S_train_pol	8	True	4, 8, 16
Bloomberg ESG (S) 23	Policy against child labor	S_child_pol	9	True	1, 8, 10, 16
Bloomberg ESG (S) 24	Consumer Data Protection Policy	S_cons_data_pol	8	True	16
Bloomberg ESG (S) 25	UN Global Compact Signatory	S_un_pol	8	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Bloomberg ESG (S) 26	SDG Target Policy	S_sdg_pol	9	False	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Bloomberg ESG (G) 1	Employee representatives on Board	G_empl	8	False	8, 10
Bloomberg ESG (G) 2	CEO Duality (fattore negativo)	G_duality	8	True	8, 16
Bloomberg ESG (G) 3	Independent directors %	G_indep_per	8	False	16
Bloomberg ESG (G) 4	Independent Chairperson	G_indep_chair	8	True	16
Bloomberg ESG (G) 5	Independent Lead Director	G_indep_dir	8	True	16
Bloomberg ESG (G) 6	Women in the board %	G_per_women	8	False	5, 10
Bloomberg ESG (G) 8	Female Chair Person or equivalent	G_female_chair	8	True	10
Bloomberg ESG (G) 9	Director average age (the lower the better)	G_dir_age	8	False	10, 16
Bloomberg ESG (G) 10	Director Meeting Attendance %	G_meeting	8	False	16
Bloomberg ESG (G) 11	Board Size (the lower the better)	G_board_size	8	False	16
Bloomberg ESG (G) 12	Board Age range (the higher the better)	G_board_age	8	False	16

Bloomberg ESG (G) 13	CSR/Sustainability Committee	G_sust_comm	9	True	16
Bloomberg ESG (G) 14	Non-executive Director with responsibility for...	G_nonex_csr	9	True	16
Bloomberg ESG (G) 15	Executive Director with Responsibility for CSR	G_ex_csr	9	True	16
Bloomberg ESG (G) 16	Executive compensation linked to ESG	G_comp_esg	9	True	16
Bloomberg ESG (G) 17	ESG Linked Compensation for Board	G_board_comp_esg	9	True	16
Bloomberg ESG (G) 18	Clawback Provision for Executive compensation	G_clawback	8	True	16
Bloomberg ESG (G) 19	Chg of Ctrl Benefits/Golden Parachute Agreements	G_agreem	8	True	16
Bloomberg ESG (G) 21	Say on pay provision	G_sayonpay	8	True	16
Bloomberg ESG (G) 22	R&D Expenditure per Cash Flow	G_rd_cash	8	False	9
Bloomberg (G)	Executive pay as percentage of total personnel...	G_exec_ratio	8	False	16
Bloomberg ESG (G) 26	Poison Pill Plan	G_poison	8	True	16
Bloomberg ESG (G) 27	Sustainable investment/capital expenditures	G_sust_invest	9	False	9, 11, 12, 16

Annex C1. EXTERNAL DATA: FinScience news

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Twitter/NGO	Carbon disclosure Project:Sentiment pesato per...	Carbon_disclosure_Project_pop_sent	7.0	No	6, 12, 13, 14, 15
Twitter/NGO	CERES:Sentiment pesato per DPV	CERES_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Twitter/NGO	Greenpeace:Sentiment pesato per DPV	Greenpeace_pop_sent	7.0	No	6, 7, 11, 12, 13, 14, 15

Twitter/NGO	WWF:Sentiment pesato per DPV	WWF_pop_sent	7.0	No	12, 14, 15
Twitter/NGO	Oxfam International:Sentiment pesato per DPV	Oxfam_International_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Twitter/NGO	Ocean Conservancy:Sentiment pesato per DPV	Ocean_Conservancy_pop_sent	7.0	No	14
Twitter/NGO	Food And Water Watch:Sentiment pesato per DPV	Food_And_Water_Watch_pop_sent	7.0	No	1, 2, 3, 6, 12
Twitter/NGO	Conservation International:Sentiment pesato per...	Conservation_International_pop_sent	7.0	No	12, 13, 14, 15
Twitter/NGO	INTERNATIONAL FEDERATION \nFOR HUMAN RIGHTS:Se...	International_Federation_for_Human_Rights_pop_...	7.0	No	1, 2, 4, 5, 8, 10
Twitter/NGO	Amnesty International:Sentiment pesato per DPV	Amnesty_International_pop_sent	7.0	No	1, 2, 3, 4, 5, 8, 10, 16
Twitter/NGO	Human Rights Watch:Sentiment pesato per DPV	Human_Rights_Watch_pop_sent	7.0	No	1, 2, 3, 4, 5, 8, 10, 16
Twitter/NGO	International Labor Rights Forum:Sentiment pes...	International_Labor_Rights_Forum_pop_sent	7.0	No	1, 3, 4, 5, 8, 10, 16
Twitter/NGO	Points of light:Sentiment pesato per DPV	Points_of_lights_pop_sent	7.0	No	11
Twitter/NGO	Great Place to work:Sentiment pesato per DPV	Great_place_to_work_pop_sent	7.0	No	8
Twitter/NGO	Human rights campaign:Sentiment pesato per DPV	Human_rights_campaign_pop_sent	7.0	No	5, 10
Twitter/NGO	Ethical Consumer:Sentiment pesato per DPV	Ethical_Consumer_pop_sent	7.0	No	12
Twitter/NGO	Fairtrade:Sentiment pesato per DPV	Fairtrade_pop_sent	7.0	No	1, 3, 8, 9, 10, 11, 12
Twitter/NGO	Forest500:Sentiment pesato per DPV	Forest500_pop_sent	7.0	No	12, 13, 14, 15
Twitter/NGO	InfluenceMap:Sentiment pesato per DPV	InfluenceMap_pop_sent	7.0	No	5, 8, 10
Twitter/NGO	JustCapital:Sentiment pesato per DPV	JustCapital_pop_sent	7.0	No	5, 8, 10, 12
Twitter/NGO	Non-GMO P:Sentiment pesato per DPV	Non-GMO_pop_sent	7.0	No	6, 7, 11, 12, 13, 14, 15
Twitter/NGO	Sipri:Sentiment pesato per DPV	Sipri_pop_sent	7.0	No	1, 16
Twitter/NGO	BSR:Sentiment pesato per DPV	BSR_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Twitter/NGO	IAG:Sentiment pesato per DPV	IAG_pop_sent	7.0	No	1, 2, 4, 5, 8, 10, 16, 17

Twitter/NGO	FSC:Sentiment pesato per DPV	FSC_pop_sent	7.0	No	15, 17
Twitter/NGO	OCEANA:Sentiment pesato per DPV	Oceana_pop_sent	7.0	No	12, 14, 17
Twitter/NGO	POP:Sentiment pesato per DPV	POP_pop_sent	7.0	No	1, 2, 3, 8, 10, 11, 12, 16, 17
Twitter/NGO	We Mean Business:Sentiment pesato per DPV	We_Mean_Business_pop_sent	7.0	No	7, 12, 13, 17
Twitter/NGO	WEConnectInternational:Sentiment pesato per DPV	WEC_pop_sent	7.0	No	5, 10, 17
Twitter/NGO	Heifer International:Sentiment pesato per DPV	Heifer_International_pop_sent	7.0	No	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 16
Twitter/NGO	The Nature Conservancy:Sentiment pesato per DPV	The_Nature_Conservancy_pop_sent	7.0	No	11, 12, 13, 14, 15
Twitter/NGO	Sierra Club:Sentiment pesato per DPV	Sierra_Club_pop_sent	7.0	No	7, 13, 14, 15
Twitter/NGO	Slow Food:Sentiment pesato per DPV	Slow_Food_pop_sent	7.0	No	2, 3, 11, 12, 15
Twitter/NGO	Global Green:Sentiment pesato per DPV	Global_Green_pop_sent	7.0	No	6, 7, 9, 11, 12, 13, 14, 15
Twitter/NGO	World Resource Institute:Sentiment pesato per DPV	World_Resource_Institute_pop_sent	7.0	No	2, 6, 7, 9, 11, 12, 13, 14, 15
Twitter/NGO	WCBSD:Sentiment pesato per DPV	WCBSD_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Twitter/NGO	Environmental defense fund:Sentiment pesato pe...	Environmental_defense_fund_pop_sent	7.0	No	2, 3, 6, 7, 11, 13, 14, 15
Twitter/NGO	Friends of the Earth:Sentiment pesato per DPV	Friends_of_Earth_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 1...
Twitter/NGO	350:Sentiment pesato per DPV	350_pop_sent	7.0	No	7, 13
Twitter/NGO	Global Witness:Sentiment pesato per DPV	Global_Witness_pop_sent	7.0	No	1, 2, 3, 12, 15, 16
Twitter/NGO	UNICEF:Sentiment pesato per DPV	UNICEF_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 10, 13, 16, 17
Twitter/NGO	UNFCCC:Sentiment pesato per DPV	UNFCCC_pop_sent	7.0	No	7, 13, 17
Twitter/NGO	FAOclimate:Sentiment pesato per DPV	Fao_climate_pop_sent	7.0	No	1, 2, 3, 7, 13
Twitter/NGO	FAO:Sentiment pesato per DPV	FAO_pop_sent	7.0	No	1, 2, 3
Twitter/NGO	IISD SDGs:Sentiment pesato per DPV	IISD_SDGs_pop_sent	7.0	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Twitter/Mainsteam	Individuazione tramite finscience di sentiment...	main_SDG_1_pop_sent	6.0	No	1

Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_2_pop_sent	6.0	No	2
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_3_pop_sent	6.0	No	3
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_4_pop_sent	6.0	No	4
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_5_pop_sent	6.0	No	5
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_6_pop_sent	6.0	No	6
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_7_pop_sent	6.0	No	7
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_8_pop_sent	6.0	No	8
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_9_pop_sent	6.0	No	9
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_10_pop_sent	6.0	No	10
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_11_pop_sent	6.0	No	11
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_12_pop_sent	6.0	No	12
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_13_pop_sent	6.0	No	13
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_14_pop_sent	6.0	No	14
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_15_pop_sent	6.0	No	15
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_16_pop_sent	6.0	No	16
Twitter/Mainstream	Individuazione tramite finscience di sentiment...	main_SDG_17_pop_sent	6.0	No	17
Twitter/Vertical 1	Individuazione sulla base dei siti verticali d...	vert_SDG_1_pop_sent	6.0	No	1
Twitter/Vertical 1	Individuazione sulla base dei siti verticali d...	vert_SDG_2_pop_sent	6.0	No	2
Twitter/Vertical 1	Individuazione sulla base dei siti verticali d...	vert_SDG_3_pop_sent	6.0	No	3
Twitter/Vertical 1	Individuazione sulla base dei siti verticali d...	vert_SDG_4_pop_sent	6.0	No	4
Twitter/Vertical 1	Individuazione sulla base dei siti verticali d...	vert_SDG_5_pop_sent	6.0	No	5

Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_6_pop_sent	6.0	No	6
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_7_pop_sent	6.0	No	7
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_8_pop_sent	6.0	No	8
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_9_pop_sent	6.0	No	9
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_10_pop_sent	6.0	No	10
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_11_pop_sent	6.0	No	11
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_12_pop_sent	6.0	No	12
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_13_pop_sent	6.0	No	13
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_14_pop_sent	6.0	No	14
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_15_pop_sent	6.0	No	15
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_16_pop_sent	6.0	No	16
Twitter/Vertica 1	Individuazione sulla base dei siti verticali d...	vert_SDG_17_pop_sent	6.0	No	17

Annex C2. EXTERNAL DATA: Rankings

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Corporate Knights Global	Corporate Knights Global 100	Corporate_Knights_Global_rank	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
BWC	Best Workplace for Commuters 2020	BWC_rank	NaN	No	8, 10
Comparably	Best Places to Work for Culture 2019	Comparably_Culture_rank	NaN	No	8
Comparably	Best Places to Work for Women 2019	Comparably_Women_rank	NaN	No	5, 8, 10
Comparably	Best Places to Work for CEOs 2019	Comparably_CEOs_rank	NaN	No	8, 16
Comparably	Best Places to Work for Work-life Bilance 2019	Comparably_Work_life_rank	NaN	No	3, 8

Comparably	Best Places to Work for Perks & Benefits 2019	Comparably_Benefits_rank	NaN	No	3, 8
Comparably	Best Places to Work for Compensation 2019	Comparably_Compensation_rank	NaN	No	1, 8
Comparably	Best Places to Work for Happiest employees 2019	Comparably_Happiest_rank	NaN	No	3, 8
Comparably	Best Places to Work for Diversity 2019	Comparably_Diversity_rank	NaN	No	5, 8, 10
Comparably	Best Places to Work for Professional Developme...	Comparably_Development_rank	NaN	No	4, 8
Computerworld	Computer World 100 Best Places to Work in IT 2019	Computerworld_rank	NaN	No	8, 9
GPW	Best Workplaces in Asia 2020	GPW_Workplace_Asia_rank	NaN	No	3, 5, 8
GPW	World's Best Workplaces 2019	GPW_World_rank	NaN	No	8
GPW	Best Workplaces in Europe 2019	GPW_Workplace_Europe_rank	NaN	No	8
GPW	Best Workplaces in Latin America 2019	GPW_Workplace_Latin_America_rank	NaN	No	8
GPW	2019 Best Big Companies to Work Forâ„¢\n	GPW_Big_rank	NaN	No	8
GPW	100 Best Workplaces for Diversity 2019 in the ...	GPW_Diversity_US_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Parentsâ„¢ 2019 in the Uni...	GPW_Parents_US_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Womenâ„¢ 2019 in the Unite...	GPW_Women_US_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Millennialsâ„¢ 2019 in the...	GPW_Millenials_rank	NaN	No	8
GPW	Best Small & Medium Workplacesâ„¢ 2019\r\nin t...	GPW_Small_Medium_rank	NaN	No	8
GPW	Best Workplaces for Women in Argentina	GPW_Women_Argentina_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Colombia	GPW_Women_Colombia_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Ecuador	GPW_Women_Ecuador_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in PerÃ¹	GPW_Women_Peru_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Irlanda	GPW_Women_Ireland_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Italia	GPW_Women_Italy_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in UK	GPW_Women_UK_rank	NaN	No	5, 8, 10

GPW	Best Workplaces for Women in India	GPW_Women_India_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Giappone	GPW_Women_Japan_rank	NaN	No	5, 8, 10
GPW	Best Workplaces for Women in Sri Lanka	GPW_Women_SriLanka_rank	NaN	No	5, 8, 10
Futurescape	CSR India 2019 Top 100	Futurescape_rank	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Fortune 100	Fortune 100 Best Companies 2019 to work for	Fortune_100_rank	NaN	No	5, 8, 10
Forbes	Just 100 2020	Forbes_100_rank	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Hispanic Network	Top 100 Diversity Supplier Programs for Hispanics	Hispanic_Network_rank	NaN	No	1, 8, 10
Training Magazine	Training Magazine Top 125 for Employee Develop...	Training_Magazine_rank	NaN	No	4, 8
Forbes	The World's Most Innovative Companies	Forbes_Innovative_rank	NaN	No	9
Fast Company	The World's 50 Most Innovative Companies	Fast_Company_rank	NaN	No	9
Blueandgreentomorrow	10 Green Companies With Amazing Environmental ...	Blueandgreentomorrow_rank	NaN	No	7, 12, 13, 14, 15
PRN	PRNews CSR Awards 2019	PRN_rank	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Omni50	Omni50 2019	Omni50_rank	NaN	No	8, 10, 11, 12, 16
Ethisphere	Ethisphere 2019	Ethisphere_rank	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
JustCapital	JUST Jobs Policy Tracker	JustCapital_rank	NaN	No	5, 8, 10, 16

Annex C3. EXTERNAL DATA: Indexes

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
CEI	Corporate Equality Index 2020	CEI_index	NaN	No	5, 8, 10
Bloomberg Gender Equality	Bloomberg Gender Equality Index	Bloomberg_Gender_Equality_index	NaN	No	5, 8, 10
Nasdaq CELS	Nasdaq Clean Edge Green Energy	Nasdaq_CELS_index	NaN	No	7, 9, 11, 12, 13
Nasdaq GWE	ISE Clean Edge Global Wind Energy Index	Nasdaq_GWE_index	NaN	No	7, 9, 11, 12, 13
Nasdaq HHO	ISE Clean Edge Water Index	Nasdaq_HHO_index	NaN	No	6, 14
Nasdaq OMX	Nasdaq OMX Clean Edge Smart Grid Infrastructur...	Nasdaq_OMX_index	NaN	No	9, 11
RENIXX	Renewable Energy Industrial Index	renixx_index	NaN	No	7, 9, 12, 13
Calvert	Calvert US Large Cap Social Index Fund	Calvert_index	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
FTI	Fashion Transparency Index 2020	FTI_index	NaN	No	1, 3, 8, 10, 12
iShares MSCI	iShares MSCI KLD 400 Social Index Fund	MSCI_KLD_index	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Bloomberg	S&P Global Clean Energy Index	SP_Clean_Energy_index	NaN	No	7, 9, 12, 13
Bloomberg	S&P Global Eco Index	SP_Global_eco_index	NaN	No	6, 7, 12, 13, 14, 15
Bloomberg	S&P 500 ESG indexes (S&P 500 ESG Index, S&P Eu...	SP_ESG_Global_Europe_index	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Bloomberg	MSCI ESG Leaders indexes (US, Europe, Asia & P...	MSCI_ESG_index	NaN	No	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...
Index	Conteggio degli indici	sdg_1_indexes	8.0	No	1

Index	Conteggio degli indici	sdg_2_indexes	8.0	No	2
Index	Conteggio degli indici	sdg_3_indexes	8.0	No	3
Index	Conteggio degli indici	sdg_4_indexes	8.0	No	4
Index	Conteggio degli indici	sdg_5_indexes	8.0	No	5
Index	Conteggio degli indici	sdg_6_indexes	8.0	No	6
Index	Conteggio degli indici	sdg_7_indexes	8.0	No	7
Index	Conteggio degli indici	sdg_8_indexes	8.0	No	8
Index	Conteggio degli indici	sdg_9_indexes	8.0	No	9
Index	Conteggio degli indici	sdg_10_indexes	8.0	No	10
Index	Conteggio degli indici	sdg_11_indexes	8.0	No	11
Index	Conteggio degli indici	sdg_12_indexes	8.0	No	12
Index	Conteggio degli indici	sdg_13_indexes	8.0	No	13
Index	Conteggio degli indici	sdg_14_indexes	8.0	No	14
Index	Conteggio degli indici	sdg_15_indexes	8.0	No	15
Index	Conteggio degli indici	sdg_16_indexes	8.0	No	16
Index	Conteggio degli indici	sdg_17_indexes	8.0	No	17
DisabilityIn	Disability Equality Index	DisabilityIn_index	NaN	No	3, 10

Annex C4. EXTERNAL DATA: Controversial Activities

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
-------------	-------------	-------------	--------	-------------	-----------------

SIPRI	SIPRI--Stockholm International Peace Research ...	arms_sales	10.0	No	16
ChemSec SIN Producers	ChemSec SIN Producers	chem_producers	10.0	No	3, 6, 12, 14, 15
Good Jobs First Violation Tracker	Good Jobs First Violation Tracker (environment...	env_violations	10.0	No	4, 6, 7, 12, 13, 14, 15
Good Jobs First Violation Tracker	Good Jobs First Violation Tracker (consumer vi...	cons_violations	10.0	No	3, 12, 16
Good Jobs First Violation Tracker	Good Jobs First Violation Tracker (governance ...	gov_violations	10.0	No	16
Good Jobs First Violation Tracker	Good Jobs First Violation Tracker (labour viol...	lab_violations	10.0	No	1, 2, 3, 6, 8, 10, 12
FCA	FCPA Corporate Investigations (bribery)	FCA_rank_n	8.0	No	16
ECB	Ethical Consumer Palm oil list	ECB_palm_oil_rank_n	7.0	No	1, 8, 10, 11, 12, 15
PERI	Toxic 100 Water Polluters Index (2019 Report, ...	PERI_Water_rank_n	8.0	No	3, 12, 13, 15
PERI	Toxic 100 Greenhouse Gas Index (2019 Report, B...	PERI_Greenhouse_rank_n	8.0	No	3, 7, 12, 13
PERI	Toxic 100 Air Polluters Index (2019 Report, Ba...	PERI_Air_rank_n	8.0	No	6, 12, 13, 14
AAP	Tobacco companies	tobacco_n	8.0	No	3
Responsible Sourcing Network (RSN)	Conflict minerals	conflict_minerals_n	8.0	No	1, 8, 11, 12, 15, 16

Annex C5. EXTERNAL DATA: Reviews

Data source	Description	Metric Name	Weight	Val if miss	Associated SDGs
Glassdoor	Overall/Rating Trend	Glassdoor_overall	8.0	No	8
Glassdoor	Culture & Values/Rating Trend	Glassdoor_culture_and_values	8.0	No	16
Glassdoor	Worklife Balance/Rating Trend	Glassdoor_worklife_balance	8.0	No	8

Glassdoor	Senior Management/Rating Trend	Glassdoor_senior_management	8.0	No	16
Glassdoor	Comp & Benefits/Rating Trend	Glassdoor_comp_and_benefits	8.0	No	8
Glassdoor	Career Opportunites/Rating Trend	Glassdoor_career_opportunities	8.0	No	8
Glassdoor	Reccomend to a friend/RatingTrend	Glassdoor_recommend_to_a_friend_perc	8.0	No	8
Glassdoor	Ceo Approval /Rating Trend	Glassdoor_CEO_approval_perc	8.0	No	
Glassdoor	Positive Perspective	Glassdoor_positive_perspectives_perc	8.0	No	8
Glassdoor	Experience/ Interviews	Glassdoor_positive_interview_experience_perc	8.0	No	8
Glassdoor	Social responsibility Pledges/ Pledges	Glassdoor_Social_Responsibility_Pledge	8.0	No	8
Glassdoor	Pay Equality Pledges/ Pledge	Glassdoor_Pay_Equality_Pledge	8.0	No	8, 10
Glassdoor	Pledges to Thrive/ Pledge	Glassdoor_Pledge_to_Thrive	8.0	No	8
Glassdoor	Diversity Commitment/Pledge	Glassdoor_Diversity_Commitment	8.0	No	5, 8, 10
Glassdoor	Career Advancement Program/Pledge	Glassdoor_Career_Advancement_Program	8.0	No	8
Glassdoor	Registered Apprenticeships/Pledge	Glassdoor_Registered_Apprenticeships	8.0	No	8
Glassdoor	Returnship/Pledge	Glassdoor_Returnships	8.0	No	4, 8
Trustpilot	Rating generale	trustpilot_score	5.0	No	8
Indeed	Work Happiness Score	Indeed_Work_Happiness_Score	10.0	No	8
Indeed	Achievement	Indeed_Achievement	10.0	No	8
Indeed	Compensation	Indeed_Compensation	10.0	No	8
Indeed	Work/Life Balance	Indeed_Work_Life_Balance	10.0	No	8
Indeed	Pay and Benefits	Indeed_Compensation_Benefits	10.0	No	8
Indeed	Job Security and Advancement	Indeed_Job_Security_Advancement	10.0	No	8
Indeed	Management	Indeed_Management	10.0	No	16
Indeed	Inclusion	Indeed_Inclusion	10.0	NaN	8
Indeed	Learning	Indeed_Learning	10.0	NaN	4, 8

Indeed	Company Culture	Indeed_Culture	10.0	No	8
PayScale	Overall Satisfaction	Payscale_overall_satisfaction	7.0	No	8
PayScale	Fair Pay	Payscale_fair_pay	7.0	No	5, 8, 10
PayScale	Learning and Development	Payscale_learning_and_development	7.0	No	4, 8
PayScale	Manager communication	Payscale_manager_communication	7.0	No	8, 16
PayScale	Manager Relationship	Payscale_manager_relationship	7.0	No	8, 16
PayScale	Pay Policy	Payscale_pay_policy	7.0	No	8, 16
PayScale	Pay Transparency	Payscale_pay_transparency	7.0	No	8, 16
PayScale	Company Outlook	Payscale_company_outlook	7.0	No	8
PayScale	Women Percentage	Payscale_women_pct	7.0	No	5, 8, 10
PayScale	Minimum Salary Gap	Payscale_min_salary_gap	7.0	No	5, 8, 10
PayScale	Max Salary Gap	Payscale_max_salary_gap	7.0	No	5, 8, 10

Ringraziamenti

Più che chiudersi con i ringraziamenti, questa tesi dovrebbe iniziare con un nuovo intero capitolo dedicato a tutte le persone che hanno contribuito alla scrittura di queste pagine e a chi, più in generale, ha arricchito questo percorso di studio e, per dirla proprio tutta, a coloro che sono al mio fianco da una quantità innumerevole di anni.

Per prima cosa non posso che ringraziare sinceramente chi mi ha dato l'opportunità e la materia prima per scrivere questa tesi: il meraviglioso team FinScience. Ringrazio prima di tutti Ilaria per avermi seguito ogni giorno da gennaio, o forse potrei dire da dicembre, con grande professionalità e umanità, dandomi consigli e spunti continui e preziosi. Ringrazio Shan e Alessio, sempre sul pezzo e puntualmente disponibili a rispondere qualsiasi tipo di domanda e pronti anche ad anticiparle con la loro acuta brillantezza. E ringrazio in generale tutte le ragazze e tutti i ragazzi di FinScience, da cui mi sono sentita accolta e straordinariamente arricchita.

Ringrazio il professor Veltri e il professor Giuliani per avermi supportato e consigliato nella stesura dell'elaborato con i loro preziosi feedback e li ringrazio, più in generale, per aver accettato di seguirmi in questa corsa contro il tempo. Sono grata all'Università di Trento per aver deciso di istituire un corso di laurea che fosse aperto anche a chi, come me, ne sapeva proprio poco di statistica e programmazione e per aver concretamente reso possibile l'integrazione tra conoscenze e ambiti che tradizionalmente vengono considerati distanti.

Ci sono poi le amiche e gli amici, un gran pezzo del mio cuore o forse, da ex studentessa di psicologia, preferirei dire della mia mente. Non sarei riuscita in questa corsa contro il tempo, e probabilmente non sarei proprio riuscita in niente, se non sapessi di avervi come costante nella mia vita. Nominerò alcuni di voi in ordine alfabetico, perché qualsiasi altro ordine non avrebbe senso. Ringrazio Ari, che con il suo sensibile ascolto delle emozioni sa leggere e scrivere tra le righe ciò che normalmente ad altri sfuggirebbe, e la cui presenza è una confortante indispensabile certezza. Ringrazio Benny, che sa riempire di intelligente leggerezza le mie giornate con la sua imprevedibilità continua e il suo spirito libero e leggiadro. Ringrazio Claudia, una bizzarra combinazione di cuore, intelletto, ambizioni e incertezze (e canzoni) perfettamente riuscita, e che è per me fonte continua di ispirazione. Ringrazio Domi, la meravigliosa pazza, che con la sua instancabile vitalità, le mille e una capacità e la voglia di fare (e disfare) colora sempre le mie giornate. Ringrazio Giaci, che ha sempre un progetto nascosto nel cassetto e il tempo ben scandito, ma sempre pronto a trovare lo spazio e il modo per ascoltare, incoraggiare, suggerire. Ringrazio Giova, per essere una certezza, anzi l'affidabile certezza del gruppo, in grado di incastrare sempre alla perfezione tutti i pezzi del puzzle con pianificazione, razionalità ma anche una quantità infinita di amorevoli sorrisi. Ringrazio Fede e Gianlu, ché, anche se ci vediamo ogni 5839 giorni, vi voglio un sacco di bene. Ringrazio tutti i colleghi e amici di corso, e in particolare il gruppo Sardegna, con la certezza di uno spritz futuro a riunirci.

I cannot but thank my other half, the one who completes myself and gives me a sense of belongingness to this world, the person that I can turn to when I'm overwhelmed by the present and scared by the future. I know, this will be very narcissistic, but that person is my Erasmus self, born in the moonlight behind a small bedroom window in Brighton and grown on the uphill-downhill streets of Edinburgh. We'll be back together soon. I also want to thank the amazing people that I unexpectedly met along the way from the first rainy day in England on. I'm thankful more than they could ever imagine, as they chose to share with me their stories, hopes and affection, no matter how short was the time that we could physically spend together. And yet their stories still inspire me to be a better person.

E ringrazio infine per ultimi, ma chiaramente per primi, coloro che sono stati la prima tessera del domino che ha dato inizio a una serie di reazioni a catena positive nella mia vita: mamma è papà. Grazie per essere stati in grado di costruire un senso di continuità, costanza e sicurezza in anni fatti di innumerevoli interruzioni, cambiamenti e incertezze, e per aver appoggiato e sostenuto ogni scelta, per essere la certezza della risposta immediata al primo squillo, dell'incoraggiamento e dell'ambizione, ma anche della calma e della pace dell'animo. Grazie a Leo per il suo personale vocabolario delle emozioni che sa declinare con intelligenza e sensibilità e per l'unicità del suo animo che arricchisce la nostra famiglia. E grazie a tutti miei familiari.

Grazie