Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale
A.a. 2023/2024
Sessione di Laurea Marzo 2024

Examining the impact of board diversity on firm profitability in Italian start-ups: a study of gender and racial representation

Relatori:
Anna D'Ambrosio
José María Sallán Leyes

Candidato:
Giovanni Urbano

Table of Contents

List of figures

List of abbreviations

BOD(s): Board(s) Of Directors

ROA: Return On Assets

ROE: Return On Equity

EPS: Earning Per Share

GLC (non-GLC): Government-Linked Companies

NA: Not Available

Abstract

Several studies have investigated whether there is a relationship between diversity on a company's board of directors and its performance. The diversity considered has been mainly gender and ethnicity, with a few studies also investigating that of education and age. The results obtained varied, with some studies finding no relationships and others finding positive or negative ones.

This study will focus on the Italian region of Piedmont, which is characterized by the presence of hi-tech and low-tech firms (such as wineries) and aims to investigate and verify the existence of a link between gender and ethnic diversity and firm performance in terms of ROA and ROE.

To do so, three data sets were analysed: one with startups only, one with manufacturing companies, and the last one combining the previous two. The analysis technique used is that of linear regression. Diversity was modelled as Blau indices and represents the independent variables. ROA and ROE, on the other hand, are the dependent variables. Company size, modelled as the logarithm of revenue, is instead used as the control variable. The models analyse the independent variables first individually and then jointly.

The results report the nonexistence of a relationship in almost all cases, evidenced by very high p-values and very low adjusted R squares. These findings are also reflected in the literature reviewed.

In contrast, the mixed sample reports a negative relationship between ethnic diversity and ROA and ROE, with very negative coefficients and highly significant p-values; however, the R square remains quite low. These results may be due to three main reasons: nationality mix increases transaction and communication costs, greater innovation by mixed boards leads to worse results in the short run and nationality mix does not add value to the company.

Resumen

Varios estudios han investigado si existe una relación entre la diversidad en el consejo de administración de una empresa y sus resultados. La diversidad considerada ha sido principalmente el género y la etnia, y unos pocos estudios han investigado también la de la educación y la edad. Los resultados obtenidos han sido variados, ya que algunos estudios no han encontrado ninguna relación y otros han encontrado relaciones positivas o negativas.

Este estudio se centrará en la región italiana del Piamonte, que se caracteriza por la presencia de empresas de alta y baja tecnología (como las bodegas) y pretende investigar y verificar la existencia de un vínculo entre la diversidad de género y étnica y el rendimiento de las empresas en términos de ROA y ROE.

Para ello, se analizaron tres conjuntos de datos: uno sólo con empresas de nueva creación, otro con empresas manufactureras y el último combinando los dos anteriores. La técnica de análisis utilizada es la de regresión lineal. La diversidad se modelizó como índices de Blau y representa las variables independientes. El ROA y el ROE, por su parte, son las variables dependientes. El tamaño de la empresa, modelizado como el logaritmo de los ingresos, se utiliza en cambio como variable de control. Los modelos analizan las variables independientes primero individualmente y luego conjuntamente.

Los resultados informan de la inexistencia de relación en casi todos los casos, evidenciada por valores p muy elevados y cuadrados R ajustados muy bajos. Estos resultados también se reflejan en la literatura revisada.

Por el contrario, la muestra mixta muestra una relación negativa entre la diversidad étnica y el ROA y el ROE, con coeficientes muy negativos y valores p muy significativos; sin embargo, el cuadrado R sigue siendo bastante bajo. Estos resultados pueden deberse a tres razones principales: la mezcla de nacionalidades aumenta los costes de transacción y comunicación, la mayor innovación por parte de los consejos mixtos conduce a peores resultados a corto plazo y la mezcla de nacionalidades no añade valor a la empresa.

1. Introduction

According to the Oxford English Dictionary diversity is the act or attribute of including or involving people from a variety of various social and ethnic origins, as well as persons of different genders, and as it will be seen, the data tell that diversity has a positive impact on the business.

There are different four macro types of diversity:

- Internal: any quality or characteristic that a person is born with, like sex, ethnicity, gender, sexual orientation and mental/physical ability;
- External: any trait, circumstance, or experience that shapes a person's identity and has been influenced by others, not inherited and can be intentionally modified. Such as socioeconomic status, education, marital status, religion and culture (linked to the society people grew up in and/or family's values);
- Organizational: differences in job function, work experience, seniority, department, or management level;
- Worldview: ideologies, stances on issues, cultural perspectives, travel experiences and everything that affects how people perceive the world and how they understand it.

There is evidence that diversity can influence companies' financial value in both short and long run (Carter et al, 2003). This relationship is due to some consequences of diversity: better understanding of the market and more creativity and innovation thanks to different attitudes and talents (Robinson and Dechant, 1997), better problem solving thanks to new perspectives. The positive impact of diversity on financial performance is also evident in growing economies like China and Russia, where better bank performance and lower loan spreads for the enterprise were found. Other quantitative example will be presented in the next section.

To be effective, diversity must be fully embraced by the firm leadership: the top management and the Board of Directors (BOD), that is the main influencer of the corporate governance. The latter is the framework of guidelines, procedures, and management techniques used to guide and oversee a business. BODs are typically composed of people who may be insiders or independent members. Insiders include major shareholders, founders, and executives with close ties to the company. Instead, independent directors do not have these close affiliations. Their selection is based on their extensive expertise in running other companies and is essential to avoid agency problems (i.e., all problems related to conflicts of interest). Independent directors are valued for their contribution to corporate governance since they reduce the centralization of power and foster the alignment of shareholders' interests with those of insiders. Beyond protecting investors' interests, BODs have other functions, such as monitoring and controlling the company throughout the auditing process; hire, evaluate, compensate and meet the CEO to design strategic plans and govern the organization; establish a policy-based governance system.

Companies are encountering both risk and unpredictability that make the performance and its driving factors hard to predict (Kuratko & Morris, 2003). Businesses need to have those constructive disputes that result in innovation and advancement if they want to stand a better chance of succeeding. Such conflicts are lost when there are homogeneity and groupthink (i.e., when members of a social group want to avoid disagreement and reach consensus without using enough means to refine, analyse, and critically evaluate ideas) in the BOD, whose diversity then becomes a contributing factor in business performance. Having a diverse board in a company improves its understanding of societal needs. This understanding can contribute to a positive image of the company, ultimately benefiting stakeholders and increasing the company's awareness of social responsibility toward the local community in which it operates (Nordberg, 2008).

As it will be seen in the next chapter, previous studies focus on American, Asian or Northern Europe cases, or just review the prior literature. For this reason, the following study will focus on the Italian region of Piedmont, which is characterized by the presence of hi-tech and low-tech companies (like wineries), and it will be aimed to investigate and verify the existence of a link between gender and ethnic diversity and firm performance. The results will bring out common patterns or highlight differences between the two kinds of firms and it will be able to show if for the Italian case previous theories are validated or disproved.

2. Literature review

2.1 Drivers of diversity: the similarity-attraction theory

A key factor of diversity in boards is the similarity-attraction theory, which can explain most cases where heterogeneity is not present or is rather rejected. This theory has inspired numerous studies on diversity and interpersonal relationships, as it explains how similar people interact more easily with each other, which can lead to more homogeneous teams and boards. The consequences of diversity in boards will be analysed in the next sections. In this section, however, the focus is on the causes of diversity and thus on the similarity-attraction theory. Many studies have built their hypotheses on the latter, so it is important to provide a brief explanation of it before moving on to case analysis.

According to Byrne (1971), when people perceive themselves as being similar to others, they experience positive feelings of attraction for them. Numerous factors that are frequently broken down into demographic (race, gender, ethnicity, socioeconomic background, and age) and psychological (personality, values, hobbies, religion, education, and career) categories are included in these similarities. The desire for social validation of one's perception of oneself and the outside world has a beneficial impact on its source, i.e., the similar person, who subsequently gains popularity. Alternately, because it questions one's perspective of oneself and the outside world, dissimilarity weakens epistemic requirements (i.e., the drive to develop understanding), which in turn leads to unpleasant sensations being associated with the dissimilar individual. Additionally, research have looked at the influence of perceived similarity, which many have found to be a more accurate predictor of attraction than real similarity (Condon and Crano, 1988; Montoya et al, 2008). However, in laboratory settings, contact with colleagues or perceptions of "bogus strangers" are where true similarity-attraction has the most of an impact.

According to Bandura (1991), the theory of social cognition, which holds that humans organize knowledge into useful cognitive classes or conceptual memory containers, explains how similarity and attraction have a positive association.
The social cognition theory, which holds that people assign membership to one social group or another based on prototypes, is the foundation of the self-categorization theory. People are then perceived as expressions of the relevant prototype rather than as distinct persons, which results in a depersonalization process.

According to Montoya and Horton (2014:60), attraction is the reaction to a specific person that is influenced by a person's cognitive assessments, emotional and/or positive instantaneous affective response, and/or behavioural response.
This viewpoint is clarified by Montoya, Kershaw, and Prosser (2018), who state that the cognitive component is not thought of as a component of attraction but rather as a method of anticipating the attraction reaction.

2.2 The costs of diversity

The ability of companies to attract, blend, and motivate a diverse workforce, such as women and minorities, results in competitive advantages for the company. These advantages come from retaining the highest quality human resources and avoiding the costs associated with turnover and absenteeism of workers dissatisfied with their careers and advancement prospects (Cox and Blake, 1991).

Nevertheless, diversity increases transaction costs, because of harder interaction and communication among the workforce coming from different backgrounds and groups. Furthermore, distrust, conflicts and dissatisfaction that may arise and the lack of economies of scale in the knowledge production contribute to the cost of diversity (Østergaard et al, 2011).

2.3 Diversity and economic performance

It is possible to approximate firms' knowledge with its employees' knowledge, and its diversity enhance new combinations and innovation (Cohen and Levinthal, 1990). Today's economy is a knowledge-based economy in which companies rely on their intangible assets more than the other assets (Teece et al., 1997), so the employee diversity becomes a pivotal factor for the firms' performance. Employee diversity is often measured through demographic factors: e.g., ethnicity as a proxy for cultural background and gender as the proportion of the two genders. The individual knowledge is influenced by the social networks, group membership and company's organisation (Walsh, 1995). For this reason, managers' interpretation of problems and their approach to solve them give the management diversity the greater predictive power for business performance (Finkelstein and Hambrick, 1990).

According to Carter, Simkins, and Simpson (2003), diversity can influence a company's financial value in both the long and short run. First, diversity fosters a better understanding of markets by bringing potential customers and suppliers together. Diversity also promotes creativity and innovation because attitudes, talent and cognitive attributes tend to distribute systematically according to demographic variables (Robinson and Dechant, 1997). In addition, diversity promotes more effective problem solving because it forces the team to explore new perspectives.

Harrison and Klein (2007) list three meanings of diversity. The first is about horizontal diversity and separation, so the disagreements among the team members due to different values, attitude or beliefs. The second one is about variety, hence, the differences of knowledge, information and experiences among the team members due to different functional backgrounds. The third one is about vertical diversity and disparity, thus, the differences in status and resources among the team members, due to an asymmetrical distribution of wealth and power.

Thus, there are two lines of thinking regarding diversity. Scholars of the first one agree that diversity is a positive factor in the decision-making process (Jackson et al., 1995; Milliken and Martins, 1996), supported by findings that show a significant positive impact on operating and financial performance in transition economies, such as China and Russia (Herdhayinta, et al, 2021; Kim et al, 2020; Liu et al, 2014), better bank performance (Cardillo et al, 2021) and lower firm's loan spreads (Karavitis et al, 2021). Instead, scholars in the other group think that diversity can lead to disagreements, conflicts and turnover (Williams and O'Reilly, 1998), especially when people are split by attributes into subgroups (Lau and Murnighan 1998: 328). This idea is supported by findings from the US and the UK that highlight how diversity leads to increased portfolio risk (Berger et al, 2014) and reduced shareholder value (Adams and Ferreira, 2009; Evgeniou and Vermaelen, 2017), beyond a negative reaction of financial markets to female directors (Gregory et al, 2013).

According to a psychological study by Huang and Kisgen (2013) men are more overconfident than women, also in companies' management, instead women are more prudent (Ho et al, 2015) and these two traits could complement each other in gender diverse boards.
Zhou (2019) formulates two hypotheses: (i) women in boards reduces the possibility of financial distress and (ii) after their presence, at least one of the key financial factors is improved. An important note done by Zhou is that the advantage of women in the board lies in the balance between the two genders, so there should be a threshold to avoid the opposite effect. The results of the study are clear: a company with a diversified board of directors tends to reduce the possibility of financial distress by 0.7% and the risk of distress by about 25%, on average. These findings are consistent with control variables for financial characteristics and BOD's characteristics and also deal with eventual endogeneity problem. Thus, the results confirm the hypotheses since they are consistent with overconfidence theory. Women in a board change the view on company's decisions and control the risk of distress. They tend to increase solvency and liquidity and to tight corporate governance, making the firm stabler. In case of distress, women approach could broaden firm's life (Zhou, 2019). The concept of a threshold to balance the genders is confirmed by a German study (Joecks et al, 2013) and a Vietnamese study (Nguyen et al, 2015), which state that gender diversity brings benefits only if the percentage of women on the board is at least 20%.
There is much evidence that women and minorities are less present in boards and core committees like audit and finance (Zattoni et al, 2023). Furthermore, prior studies highlight that although women and minorities directors' number has increased, they still have lower probability to assume these roles despite higher skills (Field et al, 2020), and women earn less than their men level peer (Zelechowski and Bilimoria, 2004).

According to Pechersky (2016), gender diversity has a positive relationship with firms' performance in family-owned companies, where women lead businesses more frequently. Furthermore, companies with weaker corporate governance will have more advantages from gender diversity in the board of directors (Mateos de Cabo et al, 2012; Fondas and Sassalos, 2000), and this diversity may improve competitiveness thanks to different approaches (Gallego-Álvarez et al, 2010). Improved corporate performance through female board members finds positive evidence in several studies around the world: Hong Kong, South Korea, Malaysia, and Singapore (Low et al, 2015), Spain (Campbell and Mínguez-Vera, 2008; Lucas-Pérez et al, 2015), Norway (Ahern and Dittmar, 2012). In contrast, Hagendorff and

Keasey (2012) report that in the United States a board with a higher proportion of women has the same risk appetite as a male-dominated board.

According to Lückerath-Rovers and De Bos (2011), women's management style brings improvements to the firm's Return on Equity (ROE), compared to firms without women on the boards. In addition, Ismail, Abdullah, and Nachum (2013) state that women on boards improve firms' Return on Assets (ROA). Instead, a study by Shrader, Blackburn and Iles (1997), based on 200 Fortune 500 companies, analysed the relationship between the percentage of women on boards of directors and the financial value of companies, such as ROA and ROE, finding a significant negative relationship. In contrast, similar research conducted by Zahra and Stanton (1988) on ROE and Earning Per Share (EPS) shows no trend or relationship.

Focusing on innovation, Østergaard, Timmermans and Kristinsson (2011) analyse new products and services of companies in a Danish database. The research shows how gender diversity is strongly related to innovation: in fact, the odds ratio shows that a team with different sexes is 68% more likely to introduce a new product or service than a team with only one sex. Instead, the ethnicity diversity has no significant influence on the innovation. The final suggestion of the study is that probably diversity is more important in interactive jobs than in repetitive ones. The positive relationship between gender diversity and innovation efficiency was found to be more pronounced in male-dominated industries (Cumming and Leung, 2021), including in terms of patents (Griffin et al, 2021).
Also, Miller and del Carmen Triana (2009) define innovation as the opportunity for the firm to create new products or services, which makes R&D expenses a proxy of innovation. In the same paper they also analyse reputation, defined as a social comparison between companies to rank status and prestige of each company (Deephouse and Carter, 2005) and represented as a Fortune score from 0 (poor) to 10 (excellent). In the study, diversity (gender and ethnic) is taken into account both as proportion and Blau's index, because, according to the behavioural theory, it can lead to innovation. Other variables, such as firm size, firm age, industry, product diversification and others, are used as control variables. Results show a positive relation between gender diversity and innovation (like Østergaard et al, 2011), but they do not report any influence of gender on reputation. This probably happens because women in leadership positions are not seen a diversity signal as strong as minorities. Actually, reputation is based on visibility and influenced by the amount of information disclosed (Brammer and Millington, 2005; Ferrier, 1997; Fombrun and Shanley, 1990). Another possible explanation of the lack of relation between gender diversity and firm reputation is that, in a global economy, ethnic diversity is seen more effective rather than gender diversity, but this not means that the latter does not foster firms (Miller and del Carmen Triana, 2009).

In their study, Carter, Simkins, and Simpson (2003) control Fortune 1000 firms by size, industry, and other parameters, looking for a relationship between the percentage of women or minorities and Tobin's Q. They report a significant positive relationship between the percentage of women and minorities and Tobin's Q and also find that the percentage of women/minority directors increases with firm size and decreases when the number of insider directors increases. Finally, the results show that companies with women on the board are more committed to having more minority directors and vice versa.

An ethnically diverse board of directors can strengthen firm value because diversity forces managers to present their ideas as clearly as possible, improving decision making, policies, procedures, and business networking (Yusoff, 2010).
It is possible to find positive results about the relation between ethnic diversity and operating performance (Erhardt et al, 2003). In fact, nationality diversity has a significant positive influence on company's performance, ROA and ROE (Kanakriyah, 2021). This result is consistent with the theory of diversity as new points of view, new solution and variety of skills.
Besides, results show a positive relation between ethnic diversity and both reputation and innovation (Miller and del Carmen Triana, 2009), that confirms the expectations linked to the behavioural theory.

Bathula (2008) finds that education diversity, in terms of PhD and non-PhD, has a negative influence on company's performance and PhD members do not add value to this performance. A similar study in an emerging economy like Mauritius, on firms listed on Mauritius Stock Exchange, from Mahadeo, Soobaroyen and Hanuman (2012) confirms Bathula's findings. In Malaysia, a study about the relationship between education background and firm's performance (ROA and ROE) reports no significant influences for both government-linked companies (GLCs) and non-GLCs (Adnan et al, 2016). The study focuses on GLCs because of their pivotal role in the Malaysian economy and firms are controlled by size and industry.
Considering age diversity, Mahadeo, Soobaroyen and Hanuman (2012) find mixed results. On the one hand age heterogeneity could have positive effects in terms of innovation and creativity. On the other hand, age homogeneity will create stronger connections in terms of values and backgrounds, improving communication.

A given variable may have different effects in different models across studies, according to the number and the content of the other variables included. This makes the whole regression model considerably differ as well (Backhaus et al., 2006), so it is hard to find true direct relationships.
As noted above, the potential pros of heterogeneous teams are remarkable: wider information, contacts and networks than homogeneous groups and also more knowledge, skills and experience thanks to the different backgrounds. Thus, diversity is positive for teams' performance especially when groups deal with complex tasks, e.g., new product/service development (O'Reilly et al., 1998; Polzer et al., 2002).
However, the findings about age and gender fit perfectly with the similarity-attraction theory. In fact, people of the same age or gender are more willing to exchange positive sentiments and build constructive communication (Haas, 2010). Diversity in these attributes can create conflicts in teams, that, if not overcome, affect the group performance.

Therefore, diversity has pros and cons: it can create socio-emotional conflicts and cliques because of social categorization (Dahlin et al., 2005), but if the hiring step is done properly, it can exploit all the single knowledges thanks to the perfect fit of skills, values, backgrounds and experience. Thus, in general, diversity should enhance innovation, but extreme diversity may lead to conflicts (Østergaard et al, 2011).

The correlation between diversity and performance is found in many studies, but in as many studies it is not found, maybe due to the fact that they investigate only diversity as pivotal characteristic within groups, ignoring other attributes that may create conflicts and affect team performance (Haas, 2010).

3. Hypotheses and models

The analysis will be based on two hypotheses suggested by the literature review. In fact, the previous studies have contrasting results about the relationship between gender/ethnicity and firm performance: some of them found positive correlations, some reported mixed or negative correlations, and some found no link at all. This literature findings led to the objective of this study: to test whether there is a relationship between these variables and, if so, what kind of relationship is. Thus, the research will be based on one hypothesis:

$diversity\ has\ no\ effects\ on\ Piedmont's\ firms\ performance.$

For this assumption to be in line with the objective, the type of diversity and the performance index must be specified:

1) $Gender\ diversity\ has\ no\ effects\ on\ ROA\ in\ Piedmont's\ firms$
2) $Gender\ diversity\ has\ no\ effects\ on\ ROE\ in\ Piedmont's\ firms$
3) $Ethnic\ diversity\ has\ no\ effects\ on\ ROA\ in\ Piedmont's\ firms$
4) $Ethnic\ diversity\ has\ no\ effects\ on\ ROE\ in\ Piedmont's\ firms$

The examination will be conducted utilizing three datasets: one dedicated to innovative startups, another focused on craft companies, and a third comprising the combination of the aforementioned two.

First, the study will search for effects of only one variable per time, controlled by size:

$$(1.a)\ ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$
$$(1.b)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$
$$(2.a)\ ROA = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$$
$$(2.b)\ ROE = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$$

Then, the mixed effect of the diversities will be investigated:

$$(3)\ ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$$
$$(4)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$$

4. Methods

The following paragraphs will go deeper into the data and the analysis process, explaining how data were prepared and the study was conducted.

4.1 Description of the sample

All the data come from Bureau Van Dijk and more specifically its Italian database AIDA (Analisi Informatizzata delle Aziende Italiane - Computerized Analysis of Italian Companies). The sample is composed of two macro categories: innovative start-ups (367 firms) and craft companies (2980 firms). For each company there are 12 variables:

- name: the business name of the firm;
- id: the unique tax code;
- k_sales_2022: the amount of sales in 2022, expressed in k€;
- k_sales_2021: the amount of sales in 2022, expressed in k€;
- roa_2022: the Return On Assets in 2022, expressed in percentage and calculated as net profit/assets;
- roa_2021: the Return On Assets in 2021, expressed in percentage and calculated as net profit/assets;
- roe_2022: the Return On Equity in 2022, expressed in percentage and calculated as net profit/equity;
- roe_2021: the Return On Equity in 2021, expressed in percentage and calculated as net profit/equity;
- n_directors: the number of directors in the board;
- gender: the gender of a director;
- age: the age of a director;
- nationality: the nationality of a director.

An excerpt of the data from both tables is reported below.

name	id	k_sales_2022	k_sales_2021	roa_2022	roa_2021	roe_2022	roe_2021	n_directo	gender	nationality	age
MINERVA ENERGIA S.R.L.	2573970031	9171	3754	3.71	3.7	13.44	21.98	3	M	Italy	42
MINERVA ENERGIA S.R.L.	2573970031	9171	3754	3.71	3.7	13.44	21.98	3	M	Italy	45
MINERVA ENERGIA S.R.L.	2573970031	9171	3754	3.71	3.7	13.44	21.98	3	M	Italy	54
TICOPTER S.R.L.	3809380045	6560	3792	8.73	4.1	11.7	34.84	3	M	Italy	32
TICOPTER S.R.L.	3809380045	6560	3792	8.73	4.1	11.7	34.84	3	F	Italy	33
TICOPTER S.R.L.	3809380045	6560	3792	8.73	4.1	11.7	34.84	3	F	Italy	27
XFARM TECHNOLOGIES ITALIA S.R.L.	2594980068	4020	1984	-24.4	-0.43	n.s.	1.66	4	M	Italy	48
XFARM TECHNOLOGIES ITALIA S.R.L.	2594980068	4020	1984	-24.4	-0.43	n.s.	1.66	4	M	Italy	35
BICINCITTA ITALIA S.R.L.	1,196E+10	3604	3035	1.31	3.44	1.77	12.21	7	M	Italy	53
BICINCITTA ITALIA S.R.L.	1,196E+10	3604	3035	1.31	3.44	1.77	12.21	7	M	Italy	53
BICINCITTA ITALIA S.R.L.	1,196E+10	3604	3035	1.31	3.44	1.77	12.21	7	M	Italy	53
BICINCITTA ITALIA S.R.L.	1,196E+10	3604	3035	1.31	3.44	1.77	12.21	7	M	Italy	53
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	44
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	56
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	39
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	39
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	56
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	56
BENEFICY S.R.L. START-UP COSTITUITA	1,4381E+10	3134	612	22.2	-79.2	94.62	n.s.	6	M	Italy	40
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	M	Italy	52
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	M	Italy	52
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	M	Italy	50
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	M	Italy	50
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	F	Italy	39
BIT S.R.L.	1,2177E+10	2807	1761	-12.1	-11.58	n.s.	-47.13	4	M	Italy	43
RE MAT S.R.L.	1,1867E+10	2153	864	-20.22	-15.52	n.s.	n.s.	9	M	Italy	58
RE MAT S.R.L.	1,1867E+10	2153	864	-20.22	-15.52	n.s.	n.s.	9	M	Italy	31
RE MAT S.R.L.	1,1867E+10	2153	864	-20.22	-15.52	n.s.	n.s.	9	M	Italy	33
RE MAT S.R.L.	1,1867E+10	2153	864	-20.22	-15.52	n.s.	n.s.	9	M	Italy	33

Figure 1: First 30 rows of startups' table

name	id	k_sales_2022	k_sales_2021	roa_2022	roa_2021	roe_2022	roe_2021	n_directors	gender	nationality	age
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	46
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	46
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	46
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	46
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	70
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	F	Italy	43
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	F	Italy	43
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	68
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	49
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	49
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	53
TEKNOSERVICE S.R.L.	8854760017	164607	131629	1.41	1.85	2.48	4.65	9	M	Italy	50
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	81
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	81
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	54
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	54
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	50
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	54
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	63
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	F	Italy	62
MAS PACK PACKAGING S.P.A.	1185560057	25393	20128	8.3	4.93	19.31	11.23	8	M	Italy	51
F.LLI DENINA S.R.L.	2904860042	22262	23283	2.81	1.8	23.94	4.56	3	M	Italy	62
F.LLI DENINA S.R.L.	2904860042	22262	23283	2.81	1.8	23.94	4.56	3	M	Italy	62
TRE ELLE GROUP S.R.L.	12390450018	22174	6152	30.4	8.55	90.63	99.21	5	M	Italy	51
TRE ELLE GROUP S.R.L.	12390450018	22174	6152	30.4	8.55	90.63	99.21	5	M	Italy	22
TRE ELLE GROUP S.R.L.	12390450018	22174	6152	30.4	8.55	90.63	99.21	5	M	Italy	56
RINALDI E PETTINAROLI S.R.L.	2228010035	21821	18845	3.8	0.84	12.55	1.44	3	M	Italy	54
RINALDI E PETTINAROLI S.R.L.	2228010035	21821	18845	3.8	0.84	12.55	1.44	3	M	Italy	51
RINALDI E PETTINAROLI S.R.L.	2228010035	21821	18845	3.8	0.84	12.55	1.44	3	M	Italy	87

Figure 2: First 30 rows of craft companies' table

4.2 Measurement of variables

The first step was to clean the data from the NAs in the pivotal variables: gender and nationality. After analysing the number of NAs of other variables, the decision was to keep only data from 2021 (ROA, ROE and sales), since there were less occurrences of missing values, and this helps to get a better model. All the missing values of dependent and independent variables were removed, and the result of the cleaning is the reduction of the sample to 171 startups and 2432 craft companies.

The second step was to calculate the Blau's index (Blau, 1977) for each firm, in order to assess the diversity of gender and nationality in each board. To do this, I created a function that in each iteration saves all the data of one firm in a local loop variable and, when all the rows of the firm were considerated, calculates the Blau's index for gender and nationality using a function by Solanas et al (2010). All the indexes are saved into a data frame linked to the id of the firm. Below it is possible to see an excerpt of 20 rows of the number of observations of diversity for each firm:

id	n
10158040963	4
11011520969	1
11570340015	4
11591030017	6
11623680011	5
11624500010	4
11628590017	3
11631140016	5
11642080011	2
11686290013	4
11691690017	5
11692450015	3
11694460012	3
11716270019	8
11740050015	6
11740750010	5
11744440014	8
11753710018	1
11795570016	4
11810610011	5

Figure 3: Number of observations of diversities for each startup

id	n
10017820019	1
1002290011	7
10029320016	2
10038300017	3
10044440013	5
10050280014	3
10058700013	3
1007620030	2
10079090014	1
10081600016	1
10084560019	5
10087110010	1
10095790019	2
10113800014	2
10118150019	3
10127280013	10
10132720011	2
10134660017	2
1014490013	3
10148070013	4

Figure 4: Number of observations of diversities for each craft company

Now, it will be shown the distribution of the Blau's indexes. First, the Blau's index for the gender diversity in startups:

startup$blau_gender	n	percent
0.0000000	120	0.701754386
0.1420118	1	0.005847953
0.2187500	1	0.005847953
0.2448980	1	0.005847953
0.2777778	4	0.023391813
0.3200000	7	0.040935673
0.3750000	11	0.064327485
0.3966942	1	0.005847953
0.4444444	15	0.087719298
0.5000000	10	0.058479532

Figure 5: Blau's index of gender distribution - startup

As it is evident, the diversity of gender in startups is quite low, in fact the 70% of indexes is 0. In average, the Blau's index of gender for startups is 0.1178, with a standard deviation of 0.1870.

Second, the Blau's index for the nationality diversity in startups:

startup$blau_nationality	n	percent
0.0000000	129	0.75438596
0.1420118	40	0.23391813
0.2603550	2	0.01169591

Figure 6: Blau's index of nationality distribution - startup

As for the gender, also nationality diversity is low (~75% of zeros), but in this case it is possible to note how about the 23% of startup is quite diversified in directors' nationality. The mean is 0.0363 and the standard deviation is 0.065.

Third, the Blau's index for the gender diversity in craft companies:

craft$blau_gender	n	percent
0.0000000	1473	0.6056743421
0.1527778	1	0.0004111842
0.1652893	2	0.0008223684
0.1800000	4	0.0016447368
0.1975309	2	0.0008223684
0.2187500	4	0.0016447368
0.2448980	19	0.0078125000
0.2603550	1	0.0004111842
0.2777778	21	0.0086348684
0.2975207	2	0.0008223684
0.3046875	1	0.0004111842
0.3200000	54	0.0222039474
0.3456790	7	0.0028782895
0.3750000	106	0.0435855263
0.4081633	28	0.0115131579
0.4200000	2	0.0008223684
0.4444444	300	0.1233552632
0.4628099	1	0.0004111842
0.4687500	5	0.0020559211
0.4800000	100	0.0411184211
0.4897959	14	0.0057565789
0.4938272	5	0.0020559211
0.4958678	2	0.0008223684
0.5000000	278	0.1143092105

Figure 7: Blau's index of gender distribution - craft companies

In this case, due to the higher number of firms in the sample, the index has many values, but the main remains the zero (60.5% of cases), with a mean of 0.1724 and a standard deviation of 0.2182.

Fourth, the Blau's index for the nationality diversity in craft companies:

craft$blau_nationality	n	percent
0.0000000	2420	0.9950657895
0.1171875	2	0.0008223684
0.1420118	1	0.0004111842
0.1527778	3	0.0012335526
0.2187500	3	0.0012335526
0.2777778	1	0.0004111842
0.3203125	1	0.0004111842
0.4260355	1	0.0004111842

Figure 8: Blau's index of nationality distribution - craft companies

In this case, the distribution is almost punctual since the 99% of firms has no diversity in their directors' nationality.

The last step was to approximate the size of the firms with the logarithm of their sales in 2021. Then, it is interesting to highlight that the size of the startups is included in

the range 0-8.241 with a mean of 4.212 and a median of 4.263. Instead, craft companies' size value goes from -Inf to 11.788, with a median of 6.930. In order to allow the calculations, the -Inf was substituted by -100, and in this case the mean of craft companies' size is 6.515. These numbers are coherent with the expectations, since startups generally have less revenue because they are innovative companies that are physiologically loss-making in their early years. In addition, their innovation often causes them to propose a supply ahead of or in excess of current market demand.
In contrast, artisanal companies are generally more established and enter more mature markets, being able to sell more easily. Conversely, because of the maturity of the markets in which they operate, there is more competition, and they are therefore subject to greater variability in revenues.

At the end of these operations, the final data frames were ready: they are composed by 6 variables: id, roa_2021, roe_2021, size, blau_gender and blau_nationality. These are the pivotal variables that will be used in the following analysis, all the others were unnecessary for the goal of the study. In addition to the two data frames (startups and craft), a third one was created merging the others, in order to analyse the data regardless the type of business. The latter data frame has 2601 firms, that means that two firms were both craft companies and startups.

An excerpt of the data from both tables after all these steps is reported below.

id	roa_2021	roe_2021	size	blau_gender	blau_nationality
10158040963	18.09	26.81	4.8598124	0.3750000	0.0000000
11011520969	-5.46	-7.19	1.7917595	0.0000000	0.0000000
11570340015	20.32	78.35	4.5217886	0.0000000	0.0000000
11591030017	-3.57	-1.77	3.3322045	0.0000000	0.0000000
11623680011	54.84	78.82	5.2203558	0.0000000	0.0000000
11624500010	2.06	1.64	5.5947114	0.0000000	0.0000000
11628590017	9.44	11.52	2.4849066	0.0000000	0.0000000
11631140016	-0.25	-2.62	4.4067192	0.3200000	0.0000000
11642080011	3.25	5.84	5.6524892	0.5000000	0.1420118
11686290013	-12.02	-27.19	6.6745614	0.0000000	0.0000000
11691690017	-74.11	-139.49	6.2499752	0.3200000	0.1420118
11692450015	-15.38	-24.29	5.9661467	0.4444444	0.0000000
11694460012	-10.07	-26.77	5.0106353	0.4444444	0.0000000
11716270019	1.77	2.41	6.7776466	0.2187500	0.0000000
11740050015	0.27	2.07	5.1357984	0.0000000	0.1420118

Figure 9: First 15 rows of startups' table after set-up operations

id	roa_2021	roe_2021	size	blau_gender	blau_nationality
10017820019	9.01	12.71	6.892642	0.0000000	0
1002290011	0.83	5.12	6.678342	0.4081633	0
10029320016	17.95	21.36	7.571988	0.0000000	0
10038300017	10.11	11.24	6.982863	0.4444444	0
10044440013	0.64	0.21	6.659294	0.4800000	0
10050280014	11.24	11.70	7.270313	0.0000000	0
10058700013	13.40	20.21	6.878326	0.0000000	0
1007620030	23.82	21.54	6.204558	0.0000000	0
10079090014	30.36	35.89	5.010635	0.0000000	0
10081600016	1.90	8.81	2.772589	0.0000000	0
10084560019	9.20	35.75	7.351158	0.0000000	0
10087110010	17.48	23.09	5.533389	0.0000000	0
10095790019	7.19	8.18	6.628041	0.0000000	0
10113800014	3.14	2.69	6.659294	0.0000000	0
10118150019	44.85	37.23	7.533694	0.4444444	0

Figure 10: First 15 rows of craft companies' table after set-up operations

4.3 Analytical techniques used

A linear regression model was chosen to analyze the relationship between gender and nationality diversity and firm performance, in terms of ROA and ROE. More specifically, the regression provides that ROA and ROE are dependent variables, instead gender and nationality Blau's indexes are the independent variables. These two variables are used first in a standalone way, and then in pair. The third independent variable, used as a control variable and always present, is the size of the company.

The analysis is composed of three macro categories: startups, craft companies and the merged data frames. Each area has in turn two models: one for the ROA and the other for the ROE. For each area, the relative sample was divided into two sets: train and test.

Before building the model, the correlation matrix was checked and as expected, there is no collinearity between the considered independent variables. To evaluate the results the p-value's threshold is set to 0.05.

5. Results

The first step was to subset the samples into train and test sets. The criterion to do this was a split ratio of 0.75 taking random ids in order to have heterogenous results with respect to the size of the companies.
All the models were also tested with the square of the Blau's indexes, getting worse results: less significant p-values and adjusted R^2.

5.1 Startups' sample

To check collinearity, it is useful to show the correlation matrix:

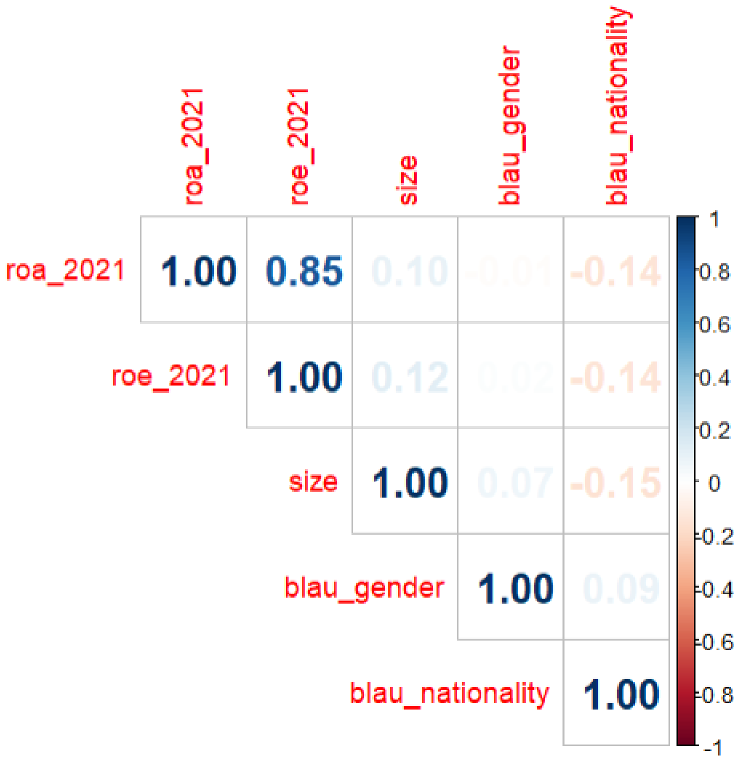

Figure 11: Startups' train set correlation matrix

As expected, there are no collinearity problems since ROA and ROE will be not used to predict each other. It also important to notice that gender diversity has no correlations with ROA and ROE. With regard to nationality diversity, the correlations negative, but very weak.

5.1.1 ROA analysis

The linear regression model for Return On Assets did not give significant results in all three cases considered: the two models with standalone diversity and the model with both.
Looking at the first model:

$$ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$

The result was the following one:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.5815     4.4414  -0.356    0.722
blau_gender   -1.6089     8.4190  -0.191    0.849
size           1.0940     0.9873   1.108    0.270

Residual standard error: 18.43 on 128 degrees of freedom
Multiple R-squared:  0.009603,  Adjusted R-squared:  -0.005872
```

Figure 12: Result of the startups' model for ROA with gender diversity

As it possible to see, the p-value is high, so the gender diversity is a non-significant to predict the ROA. In this case, the model is not useful, and this is confirmed by the negative adjusted R^2 and the mean square error of 588.266.

Similar results are obtained by the second model ($ROA = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$) whose result follows:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.5876     4.6198   0.127    0.899
blau_nationality -35.2391    23.9888  -1.469    0.144
size               0.8612     0.9883   0.871    0.385

Residual standard error: 18.28 on 128 degrees of freedom
Multiple R-squared:  0.02574,   Adjusted R-squared:  0.01052
```

Figure 13: Result of the startups' model for ROA with ethnicity diversity

Also in this case, the diversity is not significant (p-value = 0.144), and the model has a bad adjusted R^2, even if it is better than the previous. The variance is equal to 586.438.

Finally, the third model, $ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$, confirms that startups' ROA is not related to gender or nationality diversity according to the examined sample:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.6154     4.6742   0.132    0.895
blau_gender       -0.4037     8.4239  -0.048    0.962
blau_nationality -35.1248    24.2007  -1.451    0.149
size               0.8651     0.9957   0.869    0.387

Residual standard error: 18.35 on 127 degrees of freedom
Multiple R-squared:  0.02576,   Adjusted R-squared:  0.002749
```

Figure 14: Result of the startups' model for ROA with both diversities

In this case, the combined effect of both diversities makes their p-values slightly increase. The adjusted R^2 and the mean square error (585.806) are more similar to the second model rather than the first one.

5.1.2 ROE analysis

The results of the Return On Equity prediction models were worse than those of ROA.

Since the results are very similar, it follows a joined analysis of all the three models:

$$a)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$
$$b)\ ROE = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$$
$$c)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$$

Whose results are:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -4.499      9.376  -0.480    0.632
blau_gender    2.309     17.772   0.130    0.897
size           2.833      2.084   1.359    0.177

Residual standard error: 38.9 on 128 degrees of freedom
Multiple R-squared:  0.01461,	Adjusted R-squared:  -0.0007909
```

Figure 15: Result of the startups' model for ROE with gender diversity

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.5645     9.7520   0.058    0.954
blau_nationality -74.1655    50.6380  -1.465    0.145
size               2.3885     2.0863   1.145    0.254

Residual standard error: 38.58 on 128 degrees of freedom
Multiple R-squared:  0.03072,	Adjusted R-squared:  0.01557
```

Figure 16: Result of the startups' model for ROE with ethnicity diversity

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.2261     9.8640   0.023    0.982
blau_gender        4.9015    17.7768   0.276    0.783
blau_nationality -75.5537    51.0706  -1.479    0.142
size               2.3403     2.1011   1.114    0.267

Residual standard error: 38.72 on 127 degrees of freedom
Multiple R-squared:  0.0313,	Adjusted R-squared:  0.008417
```

Figure 17: Result of the startups' model for ROE with both diversities

In all cases the independent variables do not explain the dependent variable (very low adjusted R^2, even negative in the case a) and the p-values are high (again, gender has a much higher p-value). Considering the variances, they are more than three times the previous ones, respectively for the models a, b, c: 1741.178, 1702.751 and 1717.706.

5.2 Craft companies' sample

Again, to check collinearity and it is useful to show the correlation matrix:

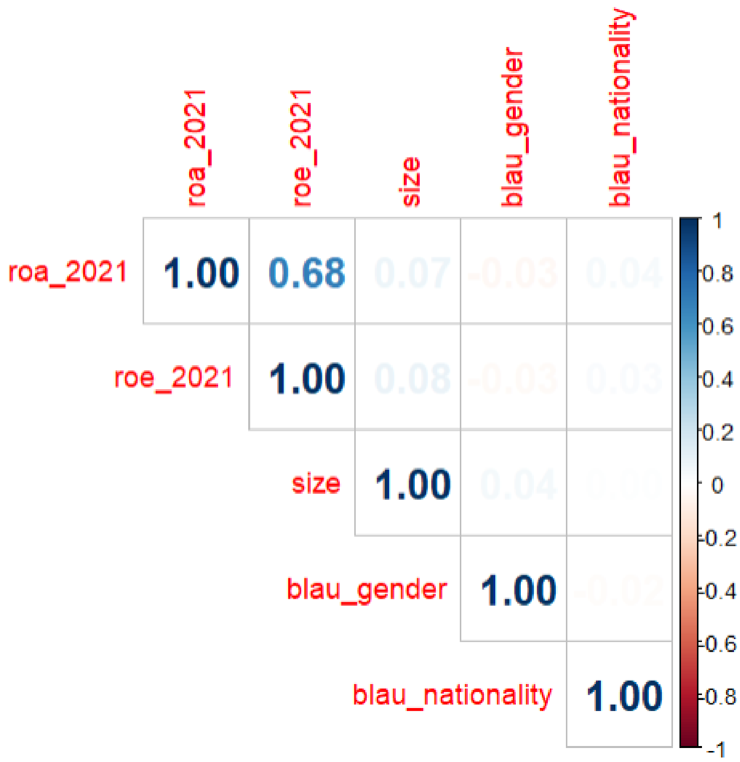

Figure 18: Craft companies' train set correlation matrix

As before, there are no collinearity problems since ROA and ROE will be not used to predict each other. In this case, both gender and ethnicity diversities have no correlations with ROA and ROE.

5.2.1 ROA analysis

The values of the correlations suggest that the models will not be so significant. In fact, the following analysis will be again a joined analysis of the three models, that have similar results.

The models were:

$$a)\ ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$
$$b)\ ROA = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$$
$$c)\ ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$$

And the respective results:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.06191    0.44758  22.481  < 2e-16 ***
blau_gender -2.06163    1.31753  -1.565  0.11781
size         0.13182    0.04273   3.085  0.00207 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.2 on 1820 degrees of freedom
Multiple R-squared:  0.006315,	Adjusted R-squared:  0.005223
```

Figure 19: Result of craft companies' model for ROA with gender diversity

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        9.7020     0.3938  24.637  < 2e-16 ***
blau_nationality  26.7159    17.7483   1.505  0.13243    
size               0.1287     0.0427   3.014  0.00261 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.2 on 1820 degrees of freedom
Multiple R-squared:  0.006215,	Adjusted R-squared:  0.005123
```

Figure 20: Result of the craft companies' model for ROA with ethnicity diversity

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      10.03109    0.44791  22.395  < 2e-16 ***
blau_gender      -2.02868    1.31729  -1.540  0.12372    
blau_nationality 26.25407   17.74416   1.480  0.13916    
size              0.13156    0.04272   3.079  0.00211 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.19 on 1819 degrees of freedom
Multiple R-squared:  0.00751,	Adjusted R-squared:  0.005873
```

Figure 21: Result of the craft companies' model for ROA with both diversities

In all cases the independent variables do not explain the dependent variable (very low adjusted R^2) and the p-values are high (in this case gender has a slight lower p-value). Considering the variances, they are much lower with respect to the startups' sample, respectively for the models a, b, c: 136.916, 136.948 and 137.225.

5.2.2 ROE analysis

As for the ROA, also for the ROE, the models are not so significant.
The models were equal to the ones showed before:

$$a)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$$

$$b)\ ROE = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$$

$$c)\ ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$$

And the results were:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  18.8861     1.0532  17.933  < 2e-16 ***
blau_gender  -3.9049     3.1002  -1.260 0.207996    
size          0.3679     0.1006   3.658 0.000261 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.7 on 1820 degrees of freedom
Multiple R-squared:  0.007957,	Adjusted R-squared:  0.006867
```

Figure 22: Result of craft companies' model for ROE with gender diversity

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        18.2087     0.9267  19.650  < 2e-16 ***
blau_nationality   46.1278    41.7646   1.104 0.269534    
size                0.3620     0.1005   3.603 0.000323 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.7 on 1820 degrees of freedom
Multiple R-squared:  0.007757,	Adjusted R-squared:  0.006667
```

Figure 23: Result of craft companies' model for ROE with ethnicity diversity

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        18.8330     1.0543  17.864  < 2e-16 ***
blau_gender        -3.8481     3.1005  -1.241 0.214725    
blau_nationality   45.2518    41.7644   1.084 0.278730    
size                0.3674     0.1006   3.654 0.000266 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.7 on 1819 degrees of freedom
Multiple R-squared:  0.008597,	Adjusted R-squared:  0.006961
```

Figure 24: Result of craft companies' model for ROE with both diversities

In all cases the independent variables do not explain the dependent variable (very low adjusted R^2) and the p-values are high (in this case gender has a slight lower p-value). Considering the variances, they are about the half of the ones of startups’ sample for the ROE, but they are more than five times the ones of the ROA for the same sample, respectively for the models a, b, c: 787.595, 786.813 and 788.757.

5.3 Mixed companies’ sample

This sample is the one that merges together startups and craft companies. Its correlation matrix confirms that there are no collinearity problems:

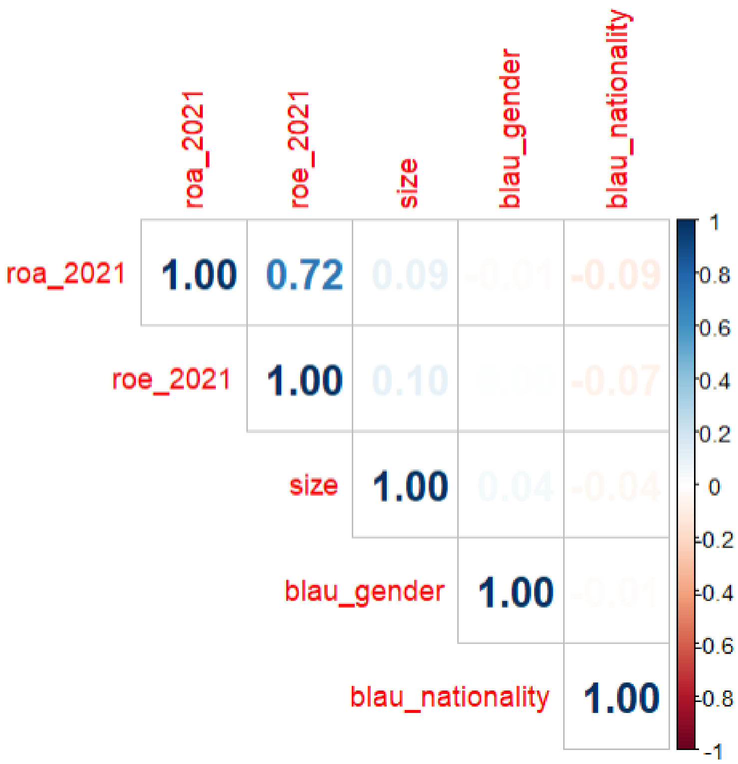

Figure 25: Merged companies' train set correlation matrix

Obviously, the correlations are still very low since the train set is a mix of startups and craft companies. Nevertheless, this sample had interesting results probably due to the mix and the higher number of observations.

5.3.1 ROA analysis

The model that considered only the gender diversity, $ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$, had a bad result:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.95110    0.44803  19.979  < 2e-16 ***
blau_gender -0.88771    1.32222  -0.671    0.502
size         0.18538    0.04396   4.217 2.59e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.62 on 1947 degrees of freedom
Multiple R-squared:  0.00917,    Adjusted R-squared:  0.008152
```

Figure 26: Result of merged companies' model for ROA with gender diversity

As in all the previous models, the p-value is high and the adjusted R^2. The mean square error is equal to 183.114.

The second model, $ROA = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$, instead, reported a significant result:

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        9.00617    0.39680  22.697  < 2e-16 ***
blau_nationality -45.24850   11.51122  -3.931 8.76e-05 ***
size               0.17729    0.04379   4.049 5.35e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.57 on 1947 degrees of freedom
Multiple R-squared:  0.01674,	Adjusted R-squared:  0.01573
```

Figure 27: Result of merged companies' model for ROA with ethnicity diversity

In fact, for the first time, the diversity (ethnic in this case) has a very significant p-value ($8.76*10^{-5}$) and one of the highest adjusted R^2 registered. The R^2 between the prediction and the actual values is 0.0018 and the variance is slightly higher than the previous and it is 183.952.

The third model, with both diversities as regressors ($ROA = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$) confirms these results:

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        9.15765    0.44944  20.376  < 2e-16 ***
blau_gender       -0.94593    1.31740  -0.718    0.473    
blau_nationality -45.34127   11.51338  -3.938 8.50e-05 ***
size               0.17865    0.04383   4.075 4.78e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.58 on 1946 degrees of freedom
Multiple R-squared:  0.017,	Adjusted R-squared:  0.01549
```

Figure 28: Result of merged companies' model for ROA with both diversities

Also in this case, the gender diversity is not significant, but the nationality one is. The adjusted R^2 is still about 0.0155. The R^2 between the prediction and the actual values is 0.0022 with a variance of 183.802.

5.3.2 ROE analysis

The results for the ROE's models are similar to the ones for ROA. The first model ($ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot size + \varepsilon$) is not significant:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  16.5122     1.0398  15.880  < 2e-16 ***
blau_gender  -0.6175     3.0686  -0.201    0.841    
size          0.4701     0.1020   4.607 4.35e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.3 on 1947 degrees of freedom
Multiple R-squared:  0.01078,	Adjusted R-squared:  0.009768
```

Figure 29: Result of merged companies' model for ROE with gender diversity

Again, the p-value is high and the adjusted R^2 is very low. The variance is equal to 902.264.

In the second model ($ROE = \beta_0 + \beta_1 \cdot diversity_{ethnicity} + \beta_2 \cdot size + \varepsilon$), as for the ROA, the nationality diversity resulted as a significant variable to predict the ROE (p-value = 0.0041), even with an adjusted R^2 of about 0.014. The R^2 between the prediction and the actual values is 0.0062 and the mean square error is 901.562.

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       16.7481     0.9225  18.155  < 2e-16 ***
blau_nationality -76.9066    26.7616  -2.874   0.0041 **
size               0.4576     0.1018   4.495 7.36e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.23 on 1947 degrees of freedom
Multiple R-squared:  0.01494,	Adjusted R-squared:  0.01393
```

Figure 30: Result of merged companies' model for ROE with ethnicity diversity

The last model, $ROE = \beta_0 + \beta_1 \cdot diversity_{gender} + \beta_2 \cdot diversity_{ethnicity} + \beta_3 \cdot size + \varepsilon$, confirms the effect of ethnic diversity (p-value = 0.0041) on ROE:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       16.8628     1.0450  16.137  < 2e-16 ***
blau_gender       -0.7163     3.0631  -0.234  0.81512
blau_nationality -76.9769    26.7698  -2.876  0.00408 **
size               0.4586     0.1019   4.500  7.2e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.24 on 1946 degrees of freedom
Multiple R-squared:  0.01497,	Adjusted R-squared:  0.01345
```

Figure 31: Result of merged companies' model for ROE with both diversities

The gender diversity is once again not significant (p-value = 0.8151). The model has an adjusted R2 of 0.01345 that means that it explains about the 1.3% of the sample. The R^2 between the prediction and the actual values is 0.0065 with a variance of 901.288.

6. Discussion and conclusions

This study aimed to investigate whether there is a relationship between gender or ethnic diversity on the BOD and firm performance in Piedmonts' firms. A number of papers were analysed that yielded mixed results: no relationship, a positive relationship, and a negative relationship. All these results led to the hypothesis that gender and ethnic diversity on the BOD has no effect on the ROA and ROE of Piedmonts' firms.

Diversity was measured as Blau's index to be consistent with most previous work. At first glance, it was evident that companies had a board of directors that was not very diverse in terms of both gender and ethnicity.

The results of the linear regression models are rather mixed.

The models referring to the samples of startups only or craft companies only report the same results: there is no relationship between gender/ethnic diversity on the BOD and firm performance in terms of ROA and ROE. These results are also confirmed for gender diversity when the two samples are combined.
The nonexistence of a relationship is not a surprise: Joecks et al (2013) and Nguyen et al (2015) found that gender diversity on the board of directors only brings benefits when a share of at least 20 percent is achieved. In addition, Zahra and Stanton's (1988) study of ROE and Earning Per Share (EPS) showed no trend or relationship. Miller and del Carmen Triana (2009) also found no link between gender diversity and corporate reputation.

In contrast, models based on the mixed sample to investigate nationality diversity yielded an interesting result. In fact, these results show a rather strong negative relationship between nationality diversity and ROA/ROE.
In fact, the coefficient of ethnic diversity in the ROA model is about -45 and that in the ROE model is about -77, both with a significant p-value (0.0085% and 4%, respectively), but a not-too-significant adjusted R^2 should also be considered.
The first question that arises is why these results did not emerge in the separate samples. The answer is that the phenomenon probably becomes observable after a certain number of observations, achieved by merging the data sets.
The second question is why the relationship is negative. In this case, the answers may be different:

- The first hypothesis is related to the costs of diversity in terms of increased transaction and communication costs, but not only that. In fact, Østergaard et al. (2011) stated that distrust, conflict and dissatisfaction can arise and that the lack of economies of scale in knowledge production contributes to the cost of diversity;
- The second is that companies with ethnically diverse boards innovate more than others, as suggested by Cohen and Levinthal (1990) and Robinson and Dechant (1997) and reported by Miller and del Carmen Triana (2009). Moreover, as Baláž et al (2023) found in the Slovakian case, innovation takes

longer to consolidate a competitive advantage and thus financial performance, with an average payback period of 8 years (Williamson et al, 2020);

- The third hypothesis is based on a study by Brixy et al (2020), who introduced the idea of "unusualness" by stating that not all ethnic combinations foster innovation, but only those with nonredundant exchanges. For this reason, it is important to analyse the history of the country where the firm is based, because immigration can influence a nation's knowledge base and thus the skills of its ruling class;

Hence, the reasons of these results can be many and many different, including the number of observations. In light of these considerations, conducting similar analyses on larger data sets could offer valuable insights.

In addition, the study takes into account ethnic diversity just as nationality diversity, but cultural distinctions include various dimensions, such as language, religious beliefs, and attitudes toward trust. Future surveys should explore these nuances further.

7. References

login.bvdinfo.com/R0/AidaNeo

www.betterup.com/blog/what-diversity-really-means-and-why-its-crucial-in-the-workplace#why-is-diversity-so-important

www.extension.iastate.edu/agdm/wholefarm/html/c5-71.html

www.investopedia.com/terms/c/corporategovernance.asp

www.oed.com

Abbasi, Z., Billsberry, J., & Todres, M. (2023). Empirical studies of the "similarity leads to attraction" hypothesis in workplace interactions: a systematic review. *Management Review Quarterly*, 1-49.

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of financial economics*, 94(2), 291-309.

Adnan, M. F., Sabli, N., Rashid, M. Z. A., Hashim, A., Paino, H., & Abdullah, A. (2016). The impact of educational level of board of directors on firms' performance. In *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014) Business and Social Sciences* (pp. 37-48). Springer Singapore.

Ahern, K. R., & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *The quarterly journal of economics*, 127(1), 137-197.

Backhaus, K. (2012). Backhaus, K./Erichson, B./Plinke, W./Weiber, R.(2006): Multivariate Analyseme. *Marken-Authentizität: Konstrukt, Determinanten und Wirkungen aus Sicht der identitätsbasierten Markenführung*, 34, 201.

Baláž, V., Jeck, T. & Balog, M. Firm performance over innovation cycle: evidence from a small European economy. *J Innov Entrep* 12, 40 (2023).

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, 50(2), 248-287.

Bathula, H. (2008). *Board characteristics and firm performance: Evidence from New Zealand* (Doctoral dissertation, Auckland University of Technology).

Berger, A. N., Kick, T., & Schaeck, K. (2014). Executive board composition and bank risk taking. *Journal of corporate finance*, 28, 48-65.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7, pp. 677-683). New York: Free Press.

Boshkoska, M. (2015). The agency problem: Measures for its overcoming. *International Journal of Business and Management*, 10(1), 204.

Brammer, S., & Millington, A. (2005). Corporate reputation and philanthropy: An empirical analysis. *Journal of business ethics*, 61, 29-44.

Brixy, U., Brunow, S., & D'Ambrosio, A. (2020). The unlikely encounter: Is ethnic diversity in start-ups associated with innovation?. *Research Policy*, 49(4), 103950.

Byrne, D. (1971). The Attraction Paradigm Academic Press. New York, NY, USA.

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance. *Journal of business ethics*, 83, 435-451.

Cardillo, G., Onali, E., & Torluccio, G. (2021). Does gender diversity on banks' boards matter? Evidence from public bailouts. *Journal of Corporate Finance*, 71, 101560.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. *Financial review*, 38(1), 33-53.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Condon, J. W., & Crano, W. D. (1988). Inferred evaluation and the relation between attitude similarity and interpersonal attraction. *Journal of personality and social psychology*, 54(5), 789.

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Perspectives*, 5(3), 45-56.

Cumming, D., & Leung, T. Y. (2021). Board diversity and corporate innovation: Regional demographics and industry context. *Corporate Governance: An International Review*, 29(3), 277-296.

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. (2005). Team diversity and information use. *Academy of management journal*, 48(6), 1107-1123.

Deephouse, D. L., & Carter, S. M. (2005). An examination of differences between organizational legitimacy and organizational reputation. *Journal of management Studies*, 42(2), 329-360.

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate governance: An international review*, 11(2), 102-111.

Evgeniou, T., & Vermaelen, T. (2017). Share buybacks and gender diversity. *Journal of Corporate Finance*, 45, 669-686.

Ferrier, W. J. (1997). Part V: Other Consequences of Corporate Reputation: 'Tough talk' and market leaders: The role of overt signaling and reputation-building behaviors in sustaining industry dominance. *Corporate Reputation Review*, 1, 98-103.

Field, L. C., Souther, M. E., & Yore, A. S. (2020). At the table but can not break through the glass ceiling: Board leadership positions elude diverse directors. *Journal of Financial Economics*, 137(3), 787-814.

Finkelstein, S., & Hambrick, D. C. (1990). Top-management-team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative science quarterly*, 484-503.

Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of management Journal*, 33(2), 233-258.

Fondas, N., & Sassalos, S. (2000). A different voice in the boardroom: How the presence of women directors affects board influence over management. *Global focus*, 12(2), 13-22.

Gallego-Álvarez, I., García-Sánchez, I. M., & Rodríguez-Dominguez, L. (2010). The influence of gender diversity on corporate performance: La influencia de la diversidad de género en el rendimiento empresarial. *Revista de Contabilidad-Spanish Accounting Review*, 13(1), 53-88.

Gregory, A., Jeanes, E., Tharyan, R., & Tonks, I. (2013). Does the stock market gender stereotype corporate boards? Evidence from the market's reaction to directors' trades. *British Journal of Management*, 24(2), 174-190.

Griffin, D., Li, K., & Xu, T. (2021). Board gender diversity and corporate innovation: International evidence. *Journal of Financial and Quantitative Analysis*, 56(1), 123-154.

Haas, H. (2010). How can we explain mixed effects of diversity on team performance? A review with emphasis on context. *Equality, Diversity and Inclusion: An International Journal*, 29(5), 458-490.

Hagendorff, J., & Keasey, K. (2012). The value of board diversity in banking: evidence from the market for corporate control. *The European Journal of Finance*, 18(1), 41-58.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of management review*, 32(4), 1199-1228.

Herdhayinta, H., Lau, J., & Shen, C. H. H. (2021). Family female directors versus non-family female directors: Effects on firm value and dividend payouts in an extreme institutional environment. *British Journal of Management*, 32(4), 969-987.

Ho, S. S., Li, A. Y., Tam, K., & Zhang, F. (2015). CEO gender, ethical leadership, and accounting conservatism. *Journal of Business Ethics*, 127, 351-370.

Huang, J., & Kisgen, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives?. *Journal of financial Economics*, 108(3), 822-839.

Ismail, K., Abdullah, S. N., & Nachum, L. (2013). Women on boards of directors of Malaysia firms: impact on market and accounting performance. *Academy of Management Proceedings*, 1(1).

Jackson, S. E., May, K. E., Whitney, K., Guzzo, R. A., & Salas, E. (1995). Understanding the dynamics of diversity in decision-making teams. *Team effectiveness and decision making in organizations*, 204, 261.

Joecks, J., Pull, K., & Vetter, K. (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a “critical mass?”. *Journal of business ethics*, 118, 61-72.

Kanakriyah, R. (2021). The impact of board of directors' characteristics on firm performance: a case study in Jordan. *The Journal of Asian Finance, Economics and Business*, 8(3), 341-350.

Karavitis, P., Kokas, S., & Tsoukas, S. (2021). Gender board diversity and the cost of bank loans. *Journal of Corporate Finance*, 71, 101804.

Kim, O., Kuang, Y. F., & Qin, B. (2020). Female representation on boards and CEO performance-induced turnover: Evidence from Russia. *Corporate Governance: An International Review*, 28(3), 235-260.

Kuratko, D. F., & Morris, M. H. (2003). Corporate entrepreneurship: The dynamic strategy for 21st century organizations. In *Issues in Entrepeneurship* (pp. 21-46). Emerald Group Publishing Limited.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of management review*, 23(2), 325-340.

Liu, Y., Wei, Z., & Xie, F. (2014). Do women directors improve firm performance in China?. *Journal of corporate finance*, 28, 169-184.

Low, D. C., Roberts, H., & Whiting, R. H. (2015). Board gender diversity and firm performance: Empirical evidence from Hong Kong, South Korea, Malaysia and Singapore. *Pacific-Basin Finance Journal*, 35, 381-401.

Lucas-Pérez, M. E., Mínguez-Vera, A., Baixauli-Soler, J. S., Martín-Ugedo, J. F., & Sánchez-Marín, G. (2015). Women on the board and managers' pay: Evidence from Spain. *Journal of Business Ethics*, 129, 265-280.

Lückerath-Rovers, M., & De Bos, A. (2011). Code of conduct for non-executive and supervisory directors. *Journal of business ethics*, 100, 465-481.

Mahadeo, J. D., Soobaroyen, T., & Hanuman, V. O. (2012). Board composition and financial performance: Uncovering the effects of diversity in an emerging economy. *Journal of business ethics*, 105, 375-388.

Mateos de Cabo, R., Gimeno, R., & Nieto, M. J. (2012). Gender diversity on European banks' boards of directors. *Journal of business ethics*, 109, 145-162.

Miller, T., & del Carmen Triana, M. (2009). Demographic diversity in the boardroom: Mediators of the board diversity–firm performance relationship. *Journal of Management studies*, 46(5), 755-786.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of management review*, 21(2), 402-433.

Montoya, R. M., & Horton, R. S. (2014). A two-dimensional model for the study of interpersonal attraction. *Personality and Social Psychology Review*, 18(1), 59-86.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25(6), 889-922.

Montoya, R. M., Kershaw, C., & Prosser, J. L. (2018). A meta-analytic investigation of the relation between interpersonal attraction and enacted behavior. *Psychological bulletin*, 144(7), 673.

Nguyen, T., Locke, S., & Reddy, K. (2015). Does boardroom gender diversity matter? Evidence from a transitional economy. *International Review of Economics & Finance*, 37, 184-202.

Nordberg, D. (2009). Some are more equal: The politics of shareholder activism. *Available at SSRN 1150130*.

O'Reilly III, C. A., Williams, K. Y., & Barsade, S. (1998). Group demography and innovation: Does diversity help?.

Østergaard, C. R., Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation. *Research policy*, 40(3), 500-509.

Pechersky, A. (2016). Diversity in board of directors: Review of diversity as a factor to enhance board performance. *Studia Commercialia Bratislavensia*, 9(33), 88.

Polzer, J. T., Milton, L. P., & Swarm Jr, W. B. (2002). Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly*, 47(2), 296-324.

Robinson, G., & Dechant, K. (1997). Building a business case for diversity. *Academy of Management Perspectives*, 11(3), 21-31.

Shrader, C. B., Blackburn, V. B., & Iles, P. (1997). Women in management and firm financial performance: An exploratory study. *Journal of managerial issues*, 355-372.

Solanas, A., Selvam, R. M., Navarro, J., & Leiva, D. (2010). On the measurement of diversity in organizations. *Unpublished manuscript*.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, 18(7), 509-533.

Walsh, J. P. (1995). Managerial and organizational cognition: Notes from a trip down memory lane. *Organization science*, 6(3), 280-321.

Williams, K. Y., & O'Reilly III, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in organizational behavior*, 20, 77-140.

Williamson, S. K., Babcock, A., & He, A. (2020). Funding the Future: Investing in Long-Horizon Innovation. *FCLTGlobal*, August.

Yusoff, W. F. W. (2010). *Characteristics of boards of directors and board effectiveness: a study of Malaysian public listed companies* (Doctoral dissertation, Victoria University).

Zahra, S. A., & Stanton, W. W. (1988). The implications of board of directors composition for corporate strategy and performance. *International journal of management*, 5(2), 229-236.

Zattoni, A., Leventis, S., Van Ees, H., & De Masi, S. (2022). Board diversity's antecedents and consequences: A review and research agenda. *The Leadership Quarterly*, 101659.

Zelechowski, D. D., & Bilimoria, D. (2004). Characteristics of women and men corporate inside directors in the US. *Corporate governance: an international review*, 12(3), 337-342.

Zhou, G. (2019). Financial distress prevention in China: Does gender of board of directors matter?. *Journal of Applied Finance and Banking*, 9(6), 127-153.

Acknowledgements

I express heartfelt gratitude to my two mentors, Professor Anna D'Ambrosio and Professor José María Sallán Leyes, for their invaluable guidance throughout the process of conducting and documenting this work over the past months. Your expertise and mentorship have played a pivotal role in the successful completion of this study.

Appendix A: R code of the analysis

````
```{r libraries}
library(tidyverse)
library(corrplot)
library(janitor)
library(caTools)
```

```{r blau.index_function}
blau.index <- function (X,categories)
{
  if (!is.character(X)) cat('ERROR: String vector should be specified in X.\n')
  else {
     blau.index <- 1-sum(prop.table(table(X))**2)
     n <- length(X)
     k <- categories
     a <- n - k * floor(n / k)
     blau.max <- (n**2*(k-1)+a*(a-k))/(k*n**2)
     blau.norm <- blau.index/blau.max
     res <- list(call = match.call(),categories = categories,
     blau.index = blau.index, blau.max = blau.max, blau.norm = blau.norm)
     class(res) <- "blau"
     res
  }
}
```
````


Blau, P. M. (1977). Inequality and heterogeneity. New York: Free Press.

Solanas, A., Selvam, R. M., Navarro, J., & Leiva, D. (2010). On the measurement of diversity in organizations. Unpublished manuscript.

```{r blau.index_calculation}
calculate.blau <- function(vector)
{
 g <- c()                                    #vector for the genders of directors of a firm
 n <- c()                                    #vector for the nationalities of directors of a firm
 b <- data.frame(id=unique(vector$id), gender=-1, nationality=-1)   #final df to save the Blau's indexes for each firm
 k <- 1
 priorId <- vector$id[1]
 for(i in 1:length(vector$id)){              #for each id
  if(vector$id[i] == priorId){               #if the current firm id is equal to the previous one
   if(vector$id[i] == vector$id[length(vector$id)]){   #if it is the last of the vector
    g[k] <- vector$gender[i]                 #calculate Blau's indexes for the firm
    indexG <- blau.index(g, 2)
    b$gender[b$id==priorId] <- indexG$blau.index

    n[k] <- vector$nationality[i]
    catN <- length(unique(n))
    indexN <- blau.index(n, catN)
    b$nationality[b$id==priorId] <- indexN$blau.index
   }
   else{                                     #save data of the director
    g[k] <- vector$gender[i]
    n[k] <- vector$nationality[i]
    k <- k+1
   }
  }
  else{                                      #if it is different from the previous, it belongs to a new firm
   indexG <- blau.index(g, 2)                #so, calculate Blau's indexes for the previous firm
   b$gender[b$id==priorId] <- indexG$blau.index      #and save them
```

```
    catN <- length(unique(n))
    indexN <- blau.index(n, catN)
    b$nationality[b$id==priorId] <- indexN$blau.index

    g <- c()                          #reset the working variables
    k <- 1
    priorId <- vector$id[i]               #now the current id becomes the prior one
    g[k] <- vector$gender[i]              #save data of the director of the new firm
    n[k] <- vector$nationality[i]
    k <- k+1
    if(vector$id[i] == vector$id[length(vector$id)]){ #if it is the last of the vector
     g[k] <- vector$gender[i]               #Blau's indexes of it = Blau's indexes of the firm
     indexG <- blau.index(g, 2)
     b$gender[b$id==priorId] <- indexG$blau.index

     n[k] <- vector$nationality[i]
     catN <- length(unique(n))
     indexN <- blau.index(n, catN)
     b$nationality[b$id==priorId] <- indexN$blau.index
    }
   }
  }
  return(b)
}
```

```{r functionRsqROA, message=FALSE, warning=FALSE}
rsquared <- function(actual, model, test){
 cor(actual, predict(model, test))^2
}
```

````
```

```{r setting_data}
startup <- read_csv2("startup.csv", locale=locale(encoding="latin1")) %>% clean_names()
startup$id <- as.character(startup$id)
startup$k_sales_2022 <- as.numeric(startup$k_sales_2022)
startup$k_sales_2021 <- as.numeric(startup$k_sales_2021)
startup$roa_2022 <- as.double(startup$roa_2022)
startup$roa_2021 <- as.double(startup$roa_2021)
startup$roe_2022 <- as.double(startup$roe_2022)
startup$roe_2021 <- as.double(startup$roe_2021)

startup <- startup %>% filter(!is.na(k_sales_2021) & !is.na(nationality) & !is.na(gender)) %>%
#only keep the data with all the useful variables
  filter(!is.na(roa_2021) & !is.na(roe_2021))

nObsStartup <- startup %>% group_by(id) %>% count()

startupBlau <- calculate.blau(startup)

startup <- startup %>%
  select(id, roa_2021, roe_2021) %>%
  cbind(size=log(startup$k_sales_2021)) %>%
  merge(startupBlau, by="id") %>%
  rename(blau_gender=gender, blau_nationality=nationality) %>%
  unique()

startup <- startup %>%
  cbind(sqr_gen = (startup$blau_gender)^2) %>%
  cbind(sqr_nat = (startup$blau_nationality)^2) %>%
````

```
  unique()

startup$size <- ifelse(startup$size==-Inf, -100, startup$size)

tabyl(startup$blau_gender)
mean(startup$blau_gender)
tabyl(startup$blau_nationality)
mean(startup$blau_nationality)

craft <- read_csv2("craft.csv", locale=locale(encoding="latin1")) %>% clean_names()
craft$id <- as.character(craft$id)
craft$k_sales_2022 <- as.numeric(craft$k_sales_2022)
craft$k_sales_2021 <- as.numeric(craft$k_sales_2021)
craft$roa_2022 <- as.double(craft$roa_2022)
craft$roa_2021 <- as.double(craft$roa_2021)
craft$roe_2022 <- as.double(craft$roe_2022)
craft$roe_2021 <- as.double(craft$roe_2021)

craft <- craft %>% filter(!is.na(k_sales_2021) & !is.na(nationality) & !is.na(gender)) %>%
  filter(!is.na(roa_2021) & !is.na(roe_2021))

nObsCraft <- craft %>% group_by(id) %>% count()

craftBlau <- calculate.blau(craft)

craft <- craft %>%
  select(id, roa_2021, roe_2021) %>%
  cbind(size=log(craft$k_sales_2021)) %>%
  merge(craftBlau, by="id") %>%
  rename(blau_gender=gender, blau_nationality=nationality)
```

```
craft <- craft %>%
  cbind(sqr_gen = (craft$blau_gender)^2) %>%
  cbind(sqr_nat = (craft$blau_nationality)^2) %>%
  unique()

craft$size <- ifelse(craft$size==-Inf, -100, craft$size)

tabyl(craft$blau_gender)
mean(craft$blau_gender)
tabyl(craft$blau_nationality)
mean(craft$blau_nationality)

companies <- rbind(startup, craft) %>% unique()
```

```
```{r model_companies_roa_gender}
set.seed(123)
split <- sample.split(companies$id, SplitRatio = 0.75)
train <- subset(companies, split == TRUE)
test <- subset(companies, split == FALSE)

corrplot(cor(select(train, c(-id))), method = 'number', type = 'upper', number.cex=1.5)

model1.1 <- lm(roa_2021 ~ blau_gender + size, data=train)
summary(model1.1)

model1.1s <- lm(roa_2021 ~ sqr_gen + size, data=train)
summary(model1.1s)

mean((test$roa_2021 - predict(model1.1, test)) ^ 2)
```
```

```
rsquared(test$roa_2021, model1.1, test)
```

```{r model_companies_roa_ethnicity}
model1.2 <- lm(roa_2021 ~ blau_nationality + size, data=train)
summary(model1.2)

model1.2s <- lm(roa_2021 ~ sqr_nat + size, data=train)
summary(model1.2s)

mean((test$roa_2021 - predict(model1.2, test)) ^ 2)

rsquared(test$roa_2021, model1.2, test)
```

```{r model_companies_roa}
model1.3 <- lm(roa_2021 ~ blau_gender + blau_nationality + size, data=train)
summary(model1.3)

model1.3s <- lm(roa_2021 ~ sqr_gen + sqr_nat + size, data=train)
summary(model1.3s)

mean((test$roa_2021 - predict(model1.3, test)) ^ 2)

rsquared(test$roa_2021, model1.3, test)
```

```{r model_companies_roe_gender}
model2.1 <- lm(roe_2021 ~ blau_gender + size, data=train)
summary(model2.1)
```

~~~
model2.1s <- lm(roe_2021 ~ sqr_gen + size, data=train)

summary(model2.1s)


mean((test$roe_2021 - predict(model2.1, test)) ^ 2)


rsquared(test$roe_2021, model2.1, test)

```


```{r model_companies_roe_nationality}

model2.2 <- lm(roe_2021 ~ blau_nationality + size, data=train)

summary(model2.2)


model2.2s <- lm(roe_2021 ~ sqr_nat + size, data=train)

summary(model2.2s)


mean((test$roe_2021 - predict(model2.2, test)) ^ 2)


rsquared(test$roe_2021, model2.2, test)

```


```{r model_companies_roe}

model2.3 <- lm(roe_2021 ~ blau_gender + blau_nationality + size, data=train)

summary(model2.3)


model2.3s <- lm(roe_2021 ~ sqr_gen + sqr_nat + size, data=train)

summary(model2.3s)


mean((test$roe_2021 - predict(model2.3, test)) ^ 2)


rsquared(test$roe_2021, model2.3, test)

```
~~~

```{r model_startup_roa_gender}
set.seed(456)
splitS <- sample.split(startup$id, SplitRatio = 0.75)
trainS <- subset(startup, split == TRUE) %>% unique() %>% filter(!is.na(id))
testS <- subset(startup, split == FALSE) %>% unique() %>% filter(!is.na(id))

corrplot(cor(select(trainS, c(-id))), method = 'number', type = 'upper', number.cex=1.5)

modelS1.1 <- lm(roa_2021 ~ blau_gender + size, data=trainS)
summary(modelS1.1)

modelS1.1s <- lm(roa_2021 ~ sqr_gen + size, data=trainS)
summary(modelS1.1s)

mean((testS$roa_2021 - predict(modelS1.1, testS)) ^ 2)

rsquared(testS$roa_2021, modelS1.1, testS)
```

```{r model_startup_roa_nationality}
modelS1.2 <- lm(roa_2021 ~ blau_nationality + size, data=trainS)
summary(modelS1.2)

modelS1.2s <- lm(roa_2021 ~ sqr_nat + size, data=trainS)
summary(modelS1.2s)

mean((testS$roa_2021 - predict(modelS1.2, testS)) ^ 2)

rsquared(testS$roa_2021, modelS1.2, testS)
```

```{r model_startup_roa}
modelS1.3 <- lm(roa_2021 ~ blau_gender + blau_nationality + size, data=trainS)
summary(modelS1.3)

modelS1.3s <- lm(roa_2021 ~ sqr_gen + sqr_nat + size, data=trainS)
summary(modelS1.3s)

mean((testS$roa_2021 - predict(modelS1.3, testS)) ^ 2)

rsquared(testS$roa_2021, modelS1.3, testS)
```

```{r model_startup_roe_gender}
modelS2.1 <- lm(roe_2021 ~ blau_gender + size, data=trainS)
summary(modelS2.1)

modelS2.1s <- lm(roe_2021 ~ sqr_gen + size, data=trainS)
summary(modelS2.1s)

mean((testS$roe_2021 - predict(modelS2.1, testS)) ^ 2)

rsquared(testS$roe_2021, modelS2.1, testS)
```

```{r model_startup_roe_nationality}
modelS2.2 <- lm(roe_2021 ~ blau_nationality + size, data=trainS)
summary(modelS2.2)

modelS2.2s <- lm(roe_2021 ~ sqr_nat + size, data=trainS)
summary(modelS2.2s)
```

````
mean((testS$roe_2021 - predict(modelS2.2, testS)) ^ 2)

rsquared(testS$roe_2021, modelS2.2, testS)
```

```{r model_startup_roe}
modelS2.3 <- lm(roe_2021 ~ blau_gender + blau_nationality + size, data=trainS)
summary(modelS2.3)

modelS2.3s <- lm(roe_2021 ~ sqr_gen + sqr_nat + size, data=trainS)
summary(modelS2.3s)

mean((testS$roe_2021 - predict(modelS2.3, testS)) ^ 2)

rsquared(testS$roe_2021, modelS2.3, testS)
```

```{r model_craft_roa_gender}
set.seed(789)
splitC <- sample.split(craft$id, SplitRatio = 0.75)
trainC <- subset(craft, split == TRUE) %>% unique() %>% filter(!is.na(id))
testC <- subset(craft, split == FALSE) %>% unique() %>% filter(!is.na(id))

corrplot(cor(select(trainC, c(-id))), method = 'number', type = 'upper', number.cex=1.5)

modelC1.1 <- lm(roa_2021 ~ blau_gender + size, data=trainC)
summary(modelC1.1)

modelC1.1s <- lm(roa_2021 ~ sqr_gen + size, data=trainC)
summary(modelC1.1s)
````

```
mean((testC$roa_2021 - predict(modelC1.1, testC)) ^ 2)

rsquared(testC$roa_2021, modelC1.1, testC)
```

```{r model_craft_roa_nationality}
modelC1.2 <- lm(roa_2021 ~ blau_nationality + size, data=trainC)
summary(modelC1.2)

modelC1.2s <- lm(roa_2021 ~ sqr_nat + size, data=trainC)
summary(modelC1.2s)

mean((testC$roa_2021 - predict(modelC1.2, testC)) ^ 2)

rsquared(testC$roa_2021, modelC1.2, testC)
```

```{r model_craft_roa}
modelC1.3 <- lm(roa_2021 ~ blau_gender + blau_nationality + size, data=trainC)
summary(modelC1.3)

modelC1.3s <- lm(roa_2021 ~ sqr_gen + sqr_nat + size, data=trainC)
summary(modelC1.3s)

mean((testC$roa_2021 - predict(modelC1.3, testC)) ^ 2)

rsquared(testC$roa_2021, modelC1.3, testC)
```

```{r model_craft_roe_gender}
```

````
modelC2.1 <- lm(roe_2021 ~ blau_gender + size, data=trainC)
summary(modelC2.1)

modelC2.1s <- lm(roe_2021 ~ sqr_gen + size, data=trainC)
summary(modelC2.1s)

mean((testC$roe_2021 - predict(modelC2.1, testC)) ^ 2)

rsquared(testC$roe_2021, modelC2.1, testC)
```

```{r model_craft_roe_nationality}
modelC2.2 <- lm(roe_2021 ~ blau_nationality + size, data=trainC)
summary(modelC2.2)

modelC2.2s <- lm(roe_2021 ~ sqr_nat + size, data=trainC)
summary(modelC2.2s)

mean((testC$roe_2021 - predict(modelC2.2, testC)) ^ 2)

rsquared(testC$roe_2021, modelC2.2, testC)
```

```{r model_craft_roe}
modelC2.3 <- lm(roe_2021 ~ blau_gender + blau_nationality + size, data=trainC)
summary(modelC2.3)

modelC2.3s <- lm(roe_2021 ~ sqr_gen + sqr_nat + size, data=trainC)
summary(modelC2.3s)

mean((testC$roe_2021 - predict(modelC2.3, testC)) ^ 2)
````

```
rsquared(testC$roe_2021, modelC2.3, testC)
```

REFERENCES

[1] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL \<<https://www.R-project.org/>\>.\

[2] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686 (URL: \<<https://doi.org/10.21105/joss.01686>\>).\

[3] Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0. \<<https://CRAN.R-project.org/package=janitor>\>\

[4] Jarek Tuszynski. CaTools library. R package version 1.18.2 \<<https://cran.r-project.org/package=caTools>\>\

[5] Taiyun Wei [cre, aut], Viliam Simko [aut], Michael Levy [ctb], Yihui Xie [ctb], Yan Jin [ctb], Jeff Zemla [ctb], Moritz Freidank [ctb], Jun Cai [ctb], Tomas Protivinsky [ctb]. Corrplot library. R package version 0.92 \<<https://cran.r-project.org/package=corrplot>\>\