CENTRE **I**NTERNATIONAL DE **F**ORMATION **E**UROPEENNE
INSTITUT **E**UROPEEN • **E**UROPEAN **I**NSTITUTE

SCHOOL OF GOVERNMENT

Digitalization as a tool to reduce corruption in the Public Administration

BY

Deanira Laze

A thesis submitted for the Joint Master degree in

Global Economic Governance & Public Affairs (GEGPA)

Academic year

2020 – 2021

July 2021

Supervisor: Ambassador Dr. Ugi Zvekic

Reviewer: Dr. Igor Tkalec

PLAGIARISM STATEMENT

I certify that this thesis is my own work, based on my personal study and/or research and that I have acknowledged all material and sources used in its preparation. I further certify that I have not copied or used any ideas or formulations from any book, article or thesis, in printed or electronic form, without specifically mentioning their origin, and that the complete citations are indicated in quotation marks.

I also certify that this assignment/report has not previously been submitted for assessment in any other unit, except where specific permission has been granted from all unit coordinators involved, and that I have not copied in part or whole or otherwise plagiarized the work of other students and/or persons.

In accordance with the law, failure to comply with these regulations makes me liable to prosecution by the disciplinary commission and the courts of the French Republic for university plagiarism.

Contents

Introduction

Research Question

Does the implementation of digital government strategies lead to increased transparency, efficiency and inclusiveness in public sector operations and a consequent reduction in the perceived level of corruption?

Aim

The aim of this paper is to analyze the efficiency and limitations of digitalization as a tool to reduce corruption and increase transparency, accountability and inclusiveness in the Public Administration.

The paper at hand constitutes an experimental thesis which aims to undertake a preliminary exploration of the role of digitalization as a tool in reducing corruption in the Public Administration through statistical means. The vastness and complexity of the topic makes this initiative ambitious yet daunting, laborious yet intriguing. However, the author does not wish to ignite false hopes in the reader, who may expect to find definitive solutions to the issue or an assessment of digitalization as an instrument in the fight against corruption. Rather, what the author hopes to do is to take the reader along on an exploratory journey which will allow both parties to unveil the multidimensional nature of these two concepts, corruption and digitalization, and step-by-step unravel their interaction in the pre-established setting. It is of great interest for the author to raise questions relative to what is known about these two widely-used concepts and to what extent pre-existing biases and preconceptions fit with the results of the analyses carried out in this preliminary study. Furthermore, this paper hopes to raise awareness against accepting certain proposals as a panacea without first analyzing the context of application, the limitations of the proposal and without putting into place an elaborate and visionary roadmap. Most importantly however, the author wishes to

lay the foundations for further studies which would explore the topic in more detail and perhaps provide more founded and reliable recommendations. For the purposes of this study, the author is aware of the limitations that the means and resources at hand impose and by no means claims to offer solutions or definitive recommendations. Therefore, a reductionist approach will be taken as a first step into a field which represents not only pressing obstacles towards the achievement of socio-economic wellbeing and sustainable development, but also great potential should the correct tools be utilized.

This paper will adopt a quantitative approach and will make use of three statistical analysis methods to detect conceptually meaningful patterns within the data; confirm or disprove the potential interaction between the variables corruption and digitalization and test their respective impact on a third latent variable which will be labelled as **overall performance of the Public Administration**; last but not least, it will attempt to explore potential correlations among the constituent indicators of digitalization and corruption for illustrative purposes. The statistical analysis methods which will be used are namely: Cluster Analysis, Principal Component Analysis and Simple Linear Regression. In order to preserve the exploratory nature of the study, the author proceeds to create an ad-hoc dataset, with data taken from repositories made available by reputable international institutions and organizations, to be inserted into the three analyses rather than adopt a pre-prepared one; therefore, the choice of variables is discretionary and aimed towards facilitating the analytical processes. The reductionist approach adopted for the purposes of this study does not allow the author to place great focus on semantics and thus broader definitions will be used when speaking of **Digitalization**, **Corruption** and **Public Administration**. More specifically, the author will make no distinction between petty and grand corruption, but will make reference to a single and unique variable. Furthermore, for this study the concept of digitalization will contain elements of digital transformation (which entails the implementation of digital innovation both in terms of physical infrastructure as well as the introduction of channels, tools and platforms aimed towards the delivery of services), e-government and digital government. Lastly, the setting defined as Public Administration will

contain elements of government and public sector services, but will leave out, if not unintentionally, the military and judicial system.

The results obtained through this study do not mean to provide reliable evidence which may confirm or disprove the role digitalization as a tool in reducing corruption in the Public Administration, nor pass judgement on currently implemented policies, but aim to provoke the thoughts of the reader and lay the groundwork for further research in the field.

Background and Literature Review

Digitalization

Definition: *"Digitization is the process of converting physical information into digital formats. It is commonly driven by technologies which focus on enhancing efficiency by automation of existing processes. On the other hand, Digitalization is the use of digital technologies to change an organization's* ***business model****, including creating new or improved ways of delivering services, and improving the quality of what is delivered"*[1].

This is the definition which the **UNDP** provides as part of its **Digital Strategy** aimed towards harnessing the power and potential of technological innovation and digital tools in order to deliver a more prosperous, bright and inclusive future for all. The **UNDP** is not alone in recognizing the immense potential and benefits that can be harnessed from digital technology. As a matter of fact, a staggering number of governments globally have implemented national **Digital Strategies** which adopt a holistic approach aimed towards combining digitization and digitalization giving way to a concept known as digital transformation. Furthermore, numerous supranational and intergovernmental institutions and organizations have highlighted the role of digital transformation in promoting economic growth and sustainable development. As such, the European Union through the **EU Next Generation** recovery package is committed to a **Digital Decade** which will help European Union member states emerge stronger, more prosperous and more united from the unprecedented crisis which struck mercilessly in 2020 leaving no sector or region unscathed. The approach that the **European Commission** has adopted, highlights the importance of employing a unified strategy and well-defined roadmap which accounts for a constantly evolving society which is ever more reliant on technological advances and digital innovation. It is an ambitious goal which sets out to overcome borders and socio-economic inequalities, while upholding the values and principles of the Union.

[1] United Nations Development Program, Digital Strategy, Available at : https://digitalstrategy.undp.org/

The technological progress and digital advancements, which have steadily consolidated themselves as of the dawn of the **Digital Revolution**, have imposed an ever growing presence of new technologies and, most noticeably, the Internet in almost all contexts and spheres of human activity. The evolution of the digital transformation presents numerous new opportunities and offers unique and tangible solutions to global challenges and issues. The current global circumstances have further emphasized the importance of these tools which have not only been instrumental in managing the COVID-19 health crisis but also ensuring a minimum of continuity in socio-economic activities. The imposed adoption of an increased use of new technologies has brought forward evidence that technological advancements offer benefits and advantages to all members of society; be it as individuals, communities, businesses, or organizations by expanding the boundaries of information availability, enhancing efficiency and agility, and reducing costs. These digital technologies offer novel means of addressing complex and unique issues, with the potential to reveal and provide long term and sustainable prosperity for various stakeholders and the planet.

As a modern society there is an ever more instilled perception that constant and rapid technological innovations drastically disrupt if not encapsulate virtually every element of everyday life. This increased level of awareness is due to a continuous exposure to media including the Internet. The ubiquitous “smart” devices which seem to have become an extension of human existence, along with the immense influence social media has on the society, have created the illusion of a highly technological reality. Nevertheless, this perception is affected by bias and distortion and as a result goes on to conceal a reality which is not as uniform in terms of technological and digital development. An increasingly digital world, in addition to unparalleled opportunities and solutions, poses also challenges and obstacles previously unencountered, which consequently have a significant impact on social and economic processes, as well as the environment. In order for all the previously mentioned initiatives to produce the positive effects which all stakeholders expect of them, it is necessary to place inclusivity at the center. The very same technologies that can offer

innovative solutions might further widen the 'digital divide' as was seen in the case of access to education during the imposed restrictions due to the pandemic. Digital tools were seen as a panacea without taking into account the existing **digital divide** due to both inadequate infrastructure and lack of access to digital tools and channels as a result of material deprivation, as well as digital illiteracy and insufficient digital skills. As a matter of fact, existing divergences between rural and urban areas and socio-economic backgrounds were emphasized during the last year, setting back by almost 20 years gains in education development[2].

The lack of a stable digital framework aimed towards facilitating the transition away from an outdated "pen and paper" approach and ready to face the latest demands proved to be detrimental to the process. Although more and more people living in developed countries, such as Italy or other member states of the EU, have access to digital devices and innovative technology, a **digital divide** persists. This phenomenon, which is most commonly associated with the generation gap, is actually much more complex as it comprises not only demographic factors but also socio – economic ones. This gap represents a social challenge resulting in inequality not only with regards to the unfair limits it imposes on certain demographic groups but also in terms of the inevitable socio-economic repercussions these limits bring about. This is on account of the fact that the digital divide, be it due to a lack of satisfactory infrastructure, education/digital (il)literacy, or cost, prevents certain individuals from taking advantage of various services offered while privileging others, leading them to have comparable advantage which in certain cases widens the income gap. To render the idea clear, it is sufficient to refer to the exclusion of citizens from receiving public services such as the issuing of documents because they do not have access to a computer or the extension of administrative procedures well beyond acceptable durations because the staff has not received appropriate training. Furthermore, it could be argued that in many countries worldwide, the digital divide is extremely noticeable also between the private and public

[2] United Nations, Department of Economics and Social Affairs, Statistics Division, Available at: https://unstats.un.org/sdgs/report/2021/goal-04/

sector where there seems to be a delay in adhering to innovative methods and implementation of technological advancements.

Therefore, one must be cautious when proposing a single instrument or even a policy as a potential panacea without taking into account the potential challenges, risks and drawbacks along the way. It is especially important for policymakers and leaders to adopt a long-term vision with an elaborate roadmap which abandons "silo-mentality" and "beggar thy neighbour" approach in favour of a team effort which aims towards a more uniform rate of global development and a relatively homogenous recovery across nations.

As previously stated, despite the interest of the author in the field of sustainable development and the potential of digital transformation (or rather digitalization for the purposes of this paper) in enabling the realization of the **2030 Sustainable Development Agenda**, the issue will no further be questioned as the scope is too wide to be tackled within the means and resources of this study. Therefore, the role of digitalization in the achievement of sustainable development will be taken as a given fact. The author will limit the focus of the study purely on the role of digitalization as a potential tool in the reduction of corruption in the Public Administration.

Digitalization in the Public Administration

In this section of the paper, the author aims to highlight the importance of innovation and digitalization in the Public Administration as propellers of transparency, accountability and deterrents of corruption and surreptitious behaviour. All these indicators are closely related to good governance and can be used to determine the quality of service delivery on behalf of the Public Administration. The achievement of sustainable socio-economic development be it national or global cannot be envisaged unless it is supported by the enactment of policies, concrete long-term strategies and roadmaps leading towards a digitalized Public Administration. That is, a Public Administration shielded from elements which may drain it of precious resources or taint the perception stakeholders such as citizens

and enterprises have of it. Digitalization would not only allow for the reduction of costs, increased efficiency, and easier access to services, but will most importantly shorten the distance between citizens, enterprises and the Public Administration encouraging transparent collaboration amongst these actors. Thus, one aim of digitalization is to offer a more user-centered approach where citizens and businesses become the focus as the receivers of services. Against this background, digital innovation can be considered a public investment with an immediate positive impact on the delivery of public services and, consequently its overall performance. There is a growing awareness among EU citizens that digital technologies play a significant role in their daily lives. According to a survey carried out in 2017 by the **European Commission**, two-thirds of respondents (EU citizens) felt that these technologies have a positive impact on society, the economy, and their own lives[3].

The ever-growing number of sectors resorting to digital channels due to the current pressing circumstances, makes digitalization an urgent need rather than a pipe dream.
The European Union is committed to facilitating and promoting digital transformation through structural and financial investments. This was illustrated in the recent proposal as for the **Digital Europe Programme** (for 2021-2027) – a historical initiative seeing as it would be the first funding programme designated solely to digital transformation in the EU[4].

In addition to the benefits stated above, digitalization encourages citizen participation in governance, thus rendering the processes through which decisions are made and implemented increasingly more transparent, enforcing accountability, and promoting efficiency. It offers a way to reduce the distance between those who make decisions and those who "suffer" the consequences. As far as this particular application of digitalization is concerned, one main benefit drawn from it is the overcoming of the rigid top-down governance approach, in addition to discouraging underhand transactions and agreements between public officials and other stakeholders. Citizens have not only the formal right to

[3] European Commission, Report, "Attitudes towards the impact of digitisation and automation on daily life", Available at: https://europa.eu/eurobarometer/surveys/detail/2160
[4] European Commission, Shaping Europe's Digital Future, The Digital Europe Program, Available at: https://digital-strategy.ec.europa.eu/en/activities/digital-programme

access information, but with the adoption of digital means have the tools to do so as well and actively participate in those activities carried out by the Public Administration which directly influence their legal sphere. An excellent example of the facilitation of the interaction between citizens and governance with the aim to increase transparency, as promoted by law and certainly supported in its achievement by digitalization, can be found in **Legislative Decree no. 97/2016 "Freedom of Information Act**"[5]. This act focuses on the prevention of corruption and increasing transparency relative to the activity carried out by the Public Administration. The paramount role of transparency in the prevention and countering of corruption has been stipulated since the **OECD Convention** signed in 1997 and entered into force in 1999[6], as well as **the Criminal Law Convention on Corruption of the Council of Europe**, signed on 27 January 1999 in Strasbourg[7]. These same notions are emphasized and reiterated by the **United Nations Convention against Corruption**[8] and by the **Council of Europe Convention on Access to Official Documents**, enacted on 18 June 2009[9]. The latter requires States Parties to implement policies aimed towards the prevention of corruption based on principles of transparency and accountability.

Legislative Decree no. 97/2016 has enriched the Italian legal system with a tool that has been present for some time in other judicial systems, the so called **"Freedom of Information Act"**. This act has amended the previous approach to administrative transparency by conferring the right to access information relative to Public Administration data and documents even on those who are not in possession of a specific or qualified interest without the need for further motivation as was required by **law no. 241/1990**[10] regulating

[5] Ministro per la Pubblica Amministrazione, Dlgs 25 maggio 2016, n. 97 - FOIA e Trasparenza, Available at: http://www.funzionepubblica.gov.it/articolo/ministro/12-02-2016/trasparenza
[6] OECD, OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions, Available at: https://www.oecd.org/corruption/oecdantibriberyconvention.htm
[7] Council of Europe, Criminal Law Convention on Corruption, Available at: https://rm.coe.int/168007f3f5
[8] United Nations Office on Drugs and Crime, United Nations Convention Against Corruption, Available at: https://www.unodc.org/documents/brussels/UN_Convention_Against_Corruption.pdf
[9] Council of Europe, Council of Europe Convention on Access to Official Documents, Available at: https://rm.coe.int/1680084826
[10] Gazzetta Ufficiale della Repubblica Italiana, Legge 7 agosto 1990, n. 241, Available at: https://www.gazzettaufficiale.it/eli/id/1990/08/18/090G0294/sg

administrative procedure. Moreover, it has added to the "proactive" type of transparency (proactive disclosure), which entails mandatory publication on websites as indicated by the law **(Legislative Decree 33/2013)**[11] - a "reactive" type, that is, reactive disclosure which comprises the so-called “right to know” of the collectivity. Reactive disclosure is obtained in response to inquiries brought forward by anyone who may be interested; thus, ensuring a “generalized” civic right to access information, much less restricted and conditioned compared to what was guaranteed by **law no. 241/1990**. This reform has brought about relevant improvements to transparency of public administrations, which was previously aimed exclusively at "control over the pursuit of institutional functions and the use of public resources". Instead, it is now also aimed at "participation".

The European Union is not the only region which has taken adamant steps towards a more digital future through the implementation of a unified **Digital Strategy** aimed towards promoting cross-border uniform development and striving towards a revitalized Europe. The LAC region has also shown remarkable progress over the years, with 73% of countries having a **Digital Agenda** in place[12]. However, while the European Union has historically attempted to present a united front and express relative unanimity in the adoption of development initiatives, the LAC region is known to be one of the most unequal globally, characterized by great gaps within and beyond national borders. Furthermore, the region has traditionally been plagued by high levels of mistrust in public institutions and dissatisfaction with public services. In such a context, the introduction and thorough implementation of digitalization would represent a unique opportunity to improve the function and the quality of public institutions performance and service delivery. Taking into account the institutional context and recent turmoil experienced in the area, the path towards digital governments would help public institutions become more trustworthy thanks to increased levels of transparency, efficient, user-centered and innovative. The contrast represented by the socio-

[11] Gazzetta Ufficiale della Repubblica Italiana,Decreto Legislativo 14 marzo 2013, n. 33, Available at: https://www.gazzettaufficiale.it/eli/id/2013/04/05/13G00076/sg

[12] The Dialogue, Latin America Advisor, “Can Latin America Find a Faster Path to Digital Government?”, Available at: https://www.thedialogue.org/analysis/can-latin-america-find-a-faster-path-to-digital-government/

economic and institutional contexts between the two regions, namely European Union and LAC, is of great interest to the study considering the potential bias generated by these elements. Preconceptions based on factors such as level of national income, political and institutional stability, cohesion within the region and level socio-economic development, could potentially lead the reader to expect certain patterns in the data. The purpose of the cluster analyses which will be attempted in this study is to firstly identify conceptually meaningful groups within the dataset, and consequently confirm or refute the preconceived notions stated above. Therefore, at this stage of the study the author refrains from criticizing or endorsing any hypotheses related to potential patterns or clusters which may arise, be they due to geographical location, level of national income or level of digitalization or corruption which are the two latent variables at the center of this study.

The drastic shift towards digitalization imposed by the pressing and unforeseen circumstances brought about by the global pandemic proved to be premature and arduous for many governments. This exposed the flaws and lack of structure in many service areas of the public sector on both ends: suppliers and recipients. There was a remarkable absence of coordination in the interaction between citizens, businesses, and the Public Administration due to a number of factors such as lack of infrastructure, inadequate control over information diffusion, and evident lack of digital skills. This made it clear that a true digital transformation cannot be achieved overnight, nor is it a singular effort; in order to achieve digital maturity, all actors must embrace a holistic approach and work in synchrony towards a common goal. In addition to the societal and organizational demands for a mass implementation of digitalization brought forward by the pandemic, there are several factors which present themselves persistently over time. Amongst the many, cost and budget pressures as well as citizen demands can be identified as the main drivers behind digital transformation. However, these drivers present themselves with different intensity in different national contexts, which once again highlights the need for meticulous coordination, well-planned strategies and targeted policies and funding over clearly defined timelines. The journey towards a digitally mature government is not without obstacles. In many cases digital transformation is often confused with the sole introduction of digital

technologies in institutions or businesses without taking the necessary measures in terms of organizational capabilities and skills to ensure that the full potential of these tools is harnessed. However, this could not be further from the truth as an organization's digital maturity is influenced, to a great degree, by its digital strategy. For instance, an organization which has been

provided with the latest computers cannot be considered digital if the employees do not have the skills to operate these tools. Therefore, in this case the attempt towards digitalization would be highly limited and funds would be considered misused. According to Mike Bracken, former chief digital and chief data officer for the UK government: *"Digital service design means designing the whole service, not just the digital bits. If you're redesigning a service, you need to think about the organization that runs it."*[13]

Corruption

This section of the paper aims to provide a brief introduction to the concept of corruption so the reader acquires some familiarity with the author's application of the notion to the study. Taking into account the complexity and vastness of the issue, the author by no means claims to offer an exhaustive treatment of the concept nor aims to take a thorough approach. Rather, the objective is to provide the reader with a glimpse of the magnitude and depth of the issue which spurred the author to take on such an ambitious analysis. The message which the author wishes to convey is that corruption is not an isolated problem neither in geographical nor sectoral terms, but rather a phenomenon which affects every echelon of the systems in which one operates and every aspect of daily socio-economic activities. It is against this backdrop that the author undertook the initiative to explore potential tools which may be used in the fight against corruption.

[13] Deloitte Insights, "The journey to government's digital transformation", Available at: https://www2.deloitte.com/za/en/insights/topics/digital-transformation/digital-transformation-in-government.html

In the words of Kofi A. Annan, Former **Secretary-General** of the United Nations: *"Corruption is an insidious plague that has a wide range of corrosive effects on societies. It undermines democracy and the rule of law, leads to violations of human rights, distorts markets, erodes the quality of life and allows organized crime, terrorism and other threats to human security to flourish."*[14]

According to **Transparency International**, the global coalition against corruption, this "evil" phenomenon is defined as the abuse of entrusted power for private gain which goes on to erode trust in institutions and authority figures, deteriorates democracy, hinders economic development and further aggravates inequality and poverty, propels social division and exclusion, and lastly intensifies the environmental crisis[15]. Corruption in itself is an extremely complex and agile variable which ranges not only in dimension, but also applicability and understanding; therefore, it ranges in its manifestations from bribery or granting favours to public servants in order to obtain or accelerate service delivery, to fraud and embezzlement of funds by public officials and executives for private gain, to sociopolitical transformations of the greatest magnitude[16]. In other words, the literature taken into consideration when attempting to provide an overarching definition for the notion corruption sometimes refers to the concepts of petty corruption, grand corruption and state capture. The term **petty corruption** is used to refer to isolated instances of corruption that do not involve the upper ranks of government leadership or economic power structures but is rather limited to lower ranking officials and public servants in the delivery of routine services. This type of corruption is most often contrasted with large-scale corruption or **grand corruption** which normally entails the abuse of high-level power and benefits few at the expense of many. Once corruption pervades leadership structures and sets within the strata of the system, it can lead to more institutionalized forms such as **state capture** (which is one of the elements constituting the composite variable "**control of corruption**" selected

[14] Ibidem 8
[15] Transparency International, What is Corruption, Available at: https://www.transparency.org/en/what-is-corruption
[16] UNODC, The Doha Declaration, Promoting a Culture of Lawfulness, Available at: https://www.unodc.org/e4j/en/anti-corruption/module-1/key-issues/corruption---baseline-definition.html

by the author and inserted in the ad-hoc dataset used throughout the statistical analyses) in which social elites (usually economic elites nowadays) co-opt and manipulate the government for their own purposes against those of the public[17]. However, **UNCAC (United Nations Convention Against Corruption)** does not distinguish among these categories of corruption, and for the purposes of this study no distinction will be made between them either. Rather **corruption** will be used as a single term, a latent variable which encompasses several indicators deemed fit and representative of this phenomenon particularly in the public sector. The indicators chosen as components of the latent variable corruption are explained more thoroughly in the **Methodology** section. In addition, the author will make no particular reference to corrupt behaviour in public procurement as that would lead to ulterior tangents which given the reductionist approach taken in this study cannot be fully explored.

UNODC (United Nations Office on Drugs and Crime) defines corruption as a "polyvalent concept" which occurs in a variety of contexts and engages and involves numerous actors; what is worse, corruption is a phenomenon which morphs and adapts to circumstances, infiltrating and contaminating systems and mechanisms in response even to regulatory and legislative changes. Corruption, however, does not always bring about the collapse of a system. In numerous cases, corruption is perceived as a "suboptimal way" or rather the price to be paid for getting things done when less surreptitious means are perceived as being unavailable, flawed, or unattainable. This perception is especially characteristic of national contexts where flawed institutional frameworks, political instability, and weak legislative structures dominate the picture. Thus, the phenomenon of corruption emerges almost inevitably and proceeds to fuel a detrimental vicious cycle. In cases where corruption does not lead to an immediate fracture in the system, it infiltrates deeply within the cells of an organization of any dimension and perpetuates a persistent pattern of unscrupulous and unethical behaviour that is repeated and in many cases reinforced over the years as it adapts to changing circumstances. This brief description of an extremely complicated notion provides a glimpse of the immense battle which organizations and governments globally

[17] Ibidem

face. As a matter of fact, the notion itself is so complex that the **United Nations Convention Against Corruption** goes on to define specific acts of corrupt conduct and urges States parties to recognize and criminalize these acts within their jurisdictions.

The cost of this phenomenon is not limited only to one sphere. Just as it is capable of infiltrating and adapting to a variety of contexts, corruption is also capable of draining precious resources from areas such as healthcare and education, diminishing the rule of law, undermining good governance and democracy and last but not least, establishing a pathological dependency of stakeholders towards this method of obtaining what is rightfully theirs. Paradoxically, in certain national contexts, petty corruption is so deeply instilled in the culture that it is no longer even initiated by the figure of authority but rather seen as an expected additional expense for not having been denied a particular service. Therefore, this goes on to indicate that power and influence is so polarized and tightly held in the hands of authority figures, be they simple public servants or executive public officials, that citizens feel intrinsically obliged to show their "gratitude" when receiving a service which perhaps by law they are entitled to and cannot be deprived of. This peculiar occurrence can range from slipping a banknote under the counter to a public servant providing a certificate to bringing gifts to the doctor who receives you without an appointment. Once can see how such a system is unsustainable and, among other costs, entrenches an enormous gap in society as those who cannot afford such means would be deprived of access to basic services. Such an unsustainable reality characterized by diminished state capacity cannot subsist for long and thus leads to increasing illiberal populist movements, rising organized crime and terrorism and potentially civil unrest and turmoil as has been seen in several LAC and Eastern European countries in recent years.

As has been stated numerous times throughout this paper, the concept of corruption is not one which can be easily framed, defined or quantified. Therefore, any attempts to measure this phenomenon cannot be considered free of error or subjectivity. While measuring corruption can be considered essential in combatting the phenomenon, it is also a challenging and arduous feat. Any attempts and instruments implemented in the

measurement of corruption are intended to reveal and shed light upon the multifaceted nature of this notion, the magnitude of the issue and impact it has on the economy, society and politico-institutional framework, not only nationally but at a global scale. Furthermore, the various indicators and tools used in the definition and quantification of corruption are necessary in the development of anti-corruption responses. Due to the polyvalent nature of this phenomenon, there cannot be one single policy or tool used as a panacea. On the contrary, measurements of corruption must be used wisely and strategically in order to detect trends, identify the particular types of corruption a system may be burdened by and illustrate the scale of the problem prior to offering a one-size-fits-all solution. Moreover, global leaders and policymakers must take into account the context of application as well as national, regional and historical circumstances when adopting initiatives. "Policy borrowing" or short-sighted strategies would not only be ineffective and inefficient, but would risk being counter effective and contribute to the perpetuation of the root problem. The fight against corruption does not allow for "silo-mentality", but requires inter-institutional and inter-governmental cooperation, collaboration and transparency. Through such an approach, policymakers, analysts and scholars can gauge the dimensions of the problem and work towards the development of tools to reduce corruption effectively. As discussed previously, there are, as there should be, different methods for measuring corruption with each one presenting benefits and drawbacks. Analogously there are also numerous elements which are proposed as potential tools in the fight against corruption. The focus of this paper will be strictly limited to the potential of one such tool, namely digitalization, in combatting corruption in the Public Administration.

Literature Review

Numerous studies and reports have attempted to highlight the role of digitalization in increasing effectiveness and efficiency in service delivery from the public authorities to stakeholders such as citizens and enterprises as well as improve the overall performance of the Public Administration. Many of these works have gone as far as to propose a more holistic approach to the implementation of digitalization especially due to the pressing

circumstances presented in the last year. The year 2020 has marked a determining point in history and particularly the era which has been defined as “digital” due to the restrictions imposed by the COVID-19 health emergency. The intensified adoption of digital infrastructure, tools and channels became paramount in providing continuity in socio-economic activities and keeping stakeholders engaged and informed. Against this backdrop, the **United Nations Department of Economics and Social Affairs** has published the **2020 Survey on E-Government “Digital Government in the Decade of Action for Sustainable Development”** which focuses on the importance of digitalization in the Public Administration in order to provide quality, sustainable and equitable services to all. This report further emphasizes the pivotal role of digitalization and digital transformation applied to the public sector in the realization of the **2030 Agenda** and enabling of Sustainable Development. Thus, it urges leaders and policy makers to consider e-government as a mission and priority in national strategies while keeping a global and long-term vision which promotes collaboration and uniform development across borders.

The literature consulted for the purposes of this study is not used as a foundation or reference for the paper due to the experimental nature of the study. Rather the author relies on quantitative data to carry out the research work. Nevertheless, the conclusions provided by other academic papers consulted do not provide definitive answers to the research question.

Methodology, Definitions and Limitations

The main rationale behind this study is to infer a potential nexus between the level of Digitalization in the Public Administration and the level of corruption in the same sector. It aims to sustain the hypothesis that the implementation of digital tools, both in terms of infrastructure as well as service delivery platforms, may lead to a reduced level of perceived corruption thanks to increased transparency in interactions with citizens and enterprises.

This study is inspired by the literature taken into consideration thus far, which mainly focuses on the potential of digitalization as a tool and, and adopts a quantitative approach aimed towards exploring and illustrating the impact and prospects of this instrument. However, considering the limited nature of this paper, the scope of study and field of research are too ample to be thoroughly analyzed. Although this study does not deny the potential of digitalization in enabling sustainable development, the limitations of this study do not allow for such a broad topic to be tackled or definitive conclusions to be drawn. The focus of the author will be restricted to a hypothetical nexus between digitalization in the Public Administration and level of corruption in the Public Administration, rather than the numerous and ample benefits "ubiquitous" Digital Transformation could bring about. Therefore, any references made to digitalization and digital transformation enabling the realization of the **Sustainable Development 2030 Agenda**, and/or contributing to sustainable socio-economic development, and/or facilitating the fight against climate change are to be taken as axioms and not hypotheses to be proven through quantitative means. Thus, all hypotheses, considerations and potential conclusions drawn, must take into account the reductionist approach of the study imposed by the numerous limits encountered. As a result, the primary purpose of this study, although it attempts to adopt a quantitative approach, is not to provide a final value, coefficient or formula, but rather to explore through two statistical models the role and potential of digitalization in relation to reducing corruption in the public sector.

This study will make use of three statistical tools in order to explore the research question. Namely, a **Cluster Analysis** will be used initially as an unsupervised method expected to generate hypotheses; should there be conceptually meaningful patterns detected

in the data, a second statistical model known as **Principal Component Analysis** will be used in order to test the hypotheses generated. The first two techniques will then be followed by a **Simple Regression Analysis** for illustrative purposes and will not aim to provide a coefficient. For the purposes of this study, the author will not use a pre-set database but rather will collect values from secondary data repositories such as **World Bank Group**, **OECD**, **UN Public Administration**, **UN DESI** and **Eurostat** in order to create an ad-hoc database. Furthermore, the countries taken into consideration are members of the European Union and Latin American and Caribbean countries. The reason behind this is the intention to explore the potential results deriving from recent policies adopted both at a national and Union level as regards the respective Digital Agendas. It is well known that the European Union has made a commitment towards a more prosperous future for all, and on the basis of this commitment has highlighted the importance of the **Digital Strategy** in the enactment of **Next Generation EU**. At the same time, in recent years several governments in the LAC region have begun to formulate national "digital" agendas or strategies to accelerate and improve digitisation of government processes as well as develop digital public service delivery with remarkable progress in some cases. Consistent efforts have been made to provide online services to their citizens and businesses meant to respond to an ever-evolving socio-economic context and meet the demands of a more digital society and have shown notable improvements over time. Therefore, the choice of units of observation, in this case European Union member states and LAC countries, is not coincidental but rather is aimed towards potentially confirming or refuting pre-existing biases such as expected uniformity or clustering due to common Union policies, national income level or perceived economic development.

Due to the complex nature of both **digitalization** and **corruption**, for the purposes of this study, they must be considered as latent variables which cannot be quantified through a single unit of measurement. Therefore, several variables, which will be defined in the following paragraphs of this methodology section, will be taken into account during the creation of the database. This section will also proceed to provide the limitations which the variables present and the restrictions imposed on the study itself.

Statistical Analysis Softwares

The statistical analyses will be carried out through the use of R and Rstudio (Version 1.3.1073) and GraphPad Prism (version 7.00)

Techniques used

The succinct explanations of each technique provided within this section does not aim to take on a didactic nature and thus assumes that the reader has a certain degree of familiarity with each technique.

1) **Cluster Analysis:** A **cluster analysis** is an exploratory analysis aimed towards identifying structures and grouping, in a conceptually meaningful way, the observations contained within a dataset. Seeing as a cluster analysis is exploratory, it does not distinguish between dependent and independent variables. These homogenous groups within the data, which are known as clusters may also be thought of "*as regions of high density separated from other such regions by regions of low density*"[18]. The results obtained from the cluster analysis will be interpreted based on the author's understanding of the data in order to establish whether the output produced by the analysis provides meaningful patterns. The method of cluster analysis which will be used for the purposes of this study is known as **hierarchical clustering** which does not require the analyst to pre-specify the number of clusters to be generated differently from the k-means approach. More specifically, this study will make use of **Agglomerative Clustering,** which is otherwise known as AGNES **(Agglomerative Nesting)**. This type of hierarchical clustering works in a bottom up manner, meaning that each object or observation is initially considered as a single-element cluster or leaf, and at each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster which is known as a node. The process continues until all elements are members of one big cluster known as a root. It is useful to note that prior to running the algorithm the original dataset will be **scaled** in order to avoid any discrepancies which may derive from different units of

[18] Hartigan, J. A. (1975) Clustering Algorithms. Wiley Series in Probability and Mathematical Statistics

measurement. The output can be visualized through a tree-like diagram known as a dendrogram. The aim is to increase the similarity among the members within a cluster and increase the dissimilarity between distinct clusters, for this reason, the linkage method which will be used in this analysis is known as the **Ward Method**; it minimizes the variation within a cluster and maximizes the variation between distinct clusters. Lastly the number of clusters to be retained can be obtained either through the dendrogram where the dissimilarity between the clusters is expressed by the height of the fusion on the vertical axis (the higher the height of the fusion, the more dissimilar the clusters are); otherwise, it can be determined through the use of the **average silhouette width**. This method is visualized through a silhouette plot. The average silhouette width increases as the number of clusters increases; therefore, it is reasonable to choose a number of clusters which maximizes the average silhouette width or stop at a number of clusters when the silhouette width no longer increases significantly.

2) **Principal Component Analysis:** For the purposes of this study, a **Principal Component Analysis** will be used to confirm of disprove the output obtained through the **Cluster Analysis**. A Principal Component Analysis is a statistical method which aims to synthesize an extensive dataset by losing as little information as possible. The purpose of this tool is to reduce dimensions of a large set of variables to a smaller set which still contains most of the information of the original larger set. It should be noted that although the **principal components** represent or replace one or more of the original variables, they are not just a one-to-one transformation; therefore, inverse transformations are not possible. Once the original data matrix has been standardized, a new data matrix called **Z** is obtained**,** where each column mean is 0 and each column variance is 0. Each column in the new **Z** matrix is a vector variable. The main idea behind a **Principal Component Analysis** model is to derive a **linear function y** for each of the vector variables. The **variance** of this function is maximized. The number of the PCs is selected taking into account their cumulative contribution to the **variance** of the **latent variable**. A threshold of 75% of variance is set. Thus, the author chooses enough **principal components** so that their

cumulative proportion represents at least 75% of the **variance** of the **latent variable**. In order to summarize the number of variables at hand, the **correlation matrix** needs to be considered and analyzed. The higher the correlation between the variables, the higher the possibility of obtaining a more contained number of variables, which will be the **principal components** of the model. The **correlation matrix** is obtained by standardizing the data in order to avoid bias due to the heterogeneity of variable measurement units. Once the **correlation matrix** has been derived, it is possible to see whether the variables that have been chosen are correlated amongst them and whether this correlation can reflect the descriptive summarizing variables the author aims to obtain. These are otherwise known as **principal components**. The **loadings matrix** contains the **coefficients** which will allow the author to obtain the PCs in relation to the original variables. The objective is to reduce the number of variables as much as possible without losing valuable information.

The **Principal Components** should be:

1) Not correlated amongst them (orthogonal)

2) Listed in a decreasing order relative to their variance

The **correlation function** allows the reader to understand whether the variables are correlated amongst them and whether they could be an expression or representation of a characteristic which is not directly discernible. The higher the correlation amongst the variables the higher the probability of obtaining a more contained number of variables which will give a more synthetic view. The latent variable, the variation of which the author will aim to explain through principal components, will be "**performance of public administration**". Presumably, based on the original dataset to be synthetized, the most influential principal components will be "corruption" and "digitalization". Respectively, it is expected that corruption will influence the latent variable negatively whereas digitalization will influence it positively.

3) **Regression Analysis:** The purpose of the **Simple Linear Regression Analysis** is to confirm the positive impact of **digitalization** on the performance of **Public Administrations** through a reduction of **corruption**. Digitalization, which entails

innovation, an increase of transparency, efficiency, and citizen/enterprise participation and interaction is expected to have a positive impact, illustrated through a negative association with corruption. This positive effect of digitalization and innovation will be analyzed through linear regression models which take into account the relationship between **digitalization** and **corruption**. Variables which make up the "principal component" digitalization will be tested against variables which constitute the "principal component" corruption. A negative trend is to be expected in each simple linear regression, which can be illustrated through a **line of best fit** and quantified through the **R^2 coefficient**. Nevertheless, this simple linear approach is limited since it is difficult to predict the exact values of the dependent variable i.e. "Public sector corrupt exchanges" by taking into consideration only the unitary changes of each independent variable selected. Therefore, the author recognizes that this explorative quantitative analysis is useful for illustrative purposes only, and does not provide reliable and quantifiable causal links. Thus, it should be further developed.

Units of Observation: 65 units of observation; European Union member states and LAC countries.

Variables: 22 variables, obtained from secondary data repositories. In order to avoid potential misinterpretations of the variables, the definitions provided will be referred directly from the source of data. As such, should the reader be interested in a brief explanation of each variable, the definitions for each variable can be found below, as quoted directly from the source.

Data Sources:

1) **World Bank Group**
2) **IDEA**
3) **World Economic Forum**
4) **Transparency International**

5) **Eurostat**
6) **Knoema**
7) **UN Public Administration**
8) **UN DESA**
9) **OECD**

Definitions

1) **GDP (Gross Domestic Product):** *measured at current $US in Millions. Where available, the most recent value (2020) was used. "GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used"*[19].
2) **Corruption Perception Index (0-100): "***The CPI scores and ranks countries/territories based on how corrupt a country's public sector is perceived to be by experts and business executives. This index uses a scale of 0-100, where 0 is highly corrupt and 100 is very clean"*[20].
3) **Public Sector Corrupt Exchanges (0-1): "***This scaled variable (where 0 represents the lowest score and 1 the highest) measures how routinely public sector employees grant favours in exchange for bribes or other material inducements. This variable takes into account a typical person employed in the public sector, excluding the*

[19] World Bank national accounts data, and OECD National Accounts data files. Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=EU-ZJ&view=map
[20] Transparency International, Corruption Perception Index. Available at: https://www.transparency.org/en/cpi/2020/index/nzl

military and averages out any major discrepancies which may derive from different branches of the public sector"[21].

4) **Public Sector Theft (0-1): "***This scaled variable (where 0 represents the lowest score and 1 the highest) measures how often public sector employees steal, embezzle or misappropriate public funds or other state resources for personal or family use. This variable takes into account a typical person employed in the public sector, excluding the military and averages out any major discrepancies which may derive from different branches of the public sector"*[22].
5) **Bribery Incidence (% of Firms experiencing at least one bribe payment request): "***Bribery incidence is the percentage of firms experiencing at least one bribe payment request across 6 public transactions dealing with utilities access, permits, licenses, and taxes"*[23].
6) **Informal payments to public officials (%): "***Informal payments to public officials are the percentage of firms expected to make informal payments to public officials to "get things done" with regard to customs, taxes, licenses, regulations, services, and the like"*[24].
7) **Executive embezzlement and theft (0-1): "***This scaled variable (where 0 represents the lowest score and 1 the highest) measures how often members of the executive (the head of state, the head of government and cabinet ministers), or their agents, steal, embezzle or misappropriate public funds or other state resources for personal or family use"*[25]*. In this case, these individuals are considered a part of the public administration in broad terms, excluding however courts, parliaments and the military.*
8) **Executive bribery and corrupt exchanges (0-1):** *This scaled variable (where 0 represents the lowest score and 1 the highest) measures how routinely members of*

[21] International Institute for Democracy and Electoral Assistance, The Global State of Democracy Indices Codebook Version 3 (2019). Available at: https://www.idea.int/gsod-indices/sites/default/files/idea-gsodi-2019-codebook-v3.pdf
[22] Ibidem
[23] World Bank, Enterprise Surveys, Available at: https://data.worldbank.org/indicator/IC.FRM.BRIB.ZS
[24] World Bank, Enterprise Surveys, Available at: https://data.worldbank.org/indicator/IC.FRM.CORR.ZS
[25] Ibidem 4

the executive (the head of state, the head of government and cabinet ministers) or their agents grant favours in exchange for bribes or other material inducements"[26]*. In this case, these individuals are considered a part of the public administration in broad terms, excluding however courts, parliaments and the military.*

9) **Absence of corruption (0-1): "***The absence of corruption variable denotes the extent to which the executive and the public administration, more broadly, do not abuse their office for personal gain. Four V-Dem indicators explicitly refer to corruption in the government broadly understood, i.e., the executive and public administration more generally (but excluding courts and parliaments). In order to obtain this variable, these and another expert-coded but broader indicator on government corruption from the ICRG data set is made use of. The five indicators have been aggregated into the absence of corruption subattribute using IRT*[27]*.This variable is scaled from 0 indicating the lowest score to 1 indicating the highest score".*

10) **Political Corruption Index:** *"This variable, which is part of the larger group of indicators measuring "Quality of Government" indicates the perceived level of corruption in the political system. It is based on the subjective assessments of carefully selected country experts. The questionnaire included 71 substantive questions and is scaled from 0 which indicates a very clean system to 1 which indicates a highly corrupt system"* [28]*.*

11) **Impact Of ICTs On Access To Basic Services (1-7):** "*This scaled variable measures in each country, to what extent ICTs enable access for all citizens to basic services (e.g., health, education, financial services, etc.) (1 = not at all; 7 = to a great extent)"* [29]*. In cases where the value for the last published version of the report is not available, the second most recent value available is taken.*

---

[26] Ibidem 4
[27] Ibidem 4
[28] World Bank, Quality of Government, Available at: https://govdata360.worldbank.org/indicators/h7b6f8df1?indicator=44172&viz=line_chart&years=1946,2019
[29] World Economic Forum, Insight Report, The Global Information Technology Report, Innovating in the Digital Economy, Available at: https://www.weforum.org/reports/the-global-information-technology-report-2016/

12) **Availability of latest technologies (1-7):** "*This scaled variable measures to what extent the latest technologies are available in each country (1 = not at all; 7 = to a great extent)*"[30].*In cases where the value for the last published version of the report is not available, the second most recent value available is taken.*
13) **Internet access in schools (1-7):** *"This scaled variable measures the extent to which Internet access is widespread in schools in each country,? (1 = nonexistent; 7 = extremely widespread)*"[31]. *In cases where the value for the last published version of the report is not available, the second most recent value available is taken.*
14) **ICT use & gov't efficiency (1-7):** "*This scaled variable measures the extent to which the use of ICTs by the government improve the quality of government services to the population (1 = not at all; 7 = to a great extent)"*[32]. *In cases where the value for the last published version of the report is not available, the second most recent value available is taken.*
15) **Fixed broadband subscriptions (per 100 people): "***Fixed broadband subscriptions refers to fixed subscriptions to high-speed access to the public Internet (a TCP/IP connection), at downstream speeds equal to, or greater than, 256 kbit/s. This includes cable modem, DSL, fiber-to-the-home/building, other fixed (wired)-broadband subscriptions, satellite broadband and terrestrial fixed wireless broadband. This total is measured irrespective of the method of payment. It excludes subscriptions that have access to data communications (including the Internet) via mobile-cellular networks. It should include fixed WiMAX and any other fixed wireless technologies. It includes both residential subscriptions and subscriptions for organizations"*[33].
16) **Individuals using the Internet (% of population):** *"Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can*

[30] Ibidem
[31] Ibidem 12
[32] Ibidem 12
[33] International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database, Available at: https://data.worldbank.org/indicator/IT.NET.BBND.P2

be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc"[34].

17) **E-Participation Index (0-1):** "*The E-Participation Index assesses, on a 0-to-1 (best) scale, the quality, relevance, and usefulness of government websites in providing online information and participatory tools and services to their citizens"*[35].

18) **Online Service Index (0-1):** *"This scaled variable (where 0 indicates the lowest value and 1 indicates the highest) measures the evolution of e-government services (smart services) in terms of availability, quality, connectivity and diversity of channels and the use by the public of these services"*[36].

19) **Telecommunication Infrastructure Index (0-1):** "*The Telecommunication Infrastructure Index is an arithmetic average composite of five indicators: (i) estimated Internet users per 100 inhabitants (ii) number of main fixed telephone lines per 100 inhabitants (iii) number of mobile subscribers per 100 inhabitants (iv) number of wireless broadband subscriptions per 100 inhabitants; and (v) number of fixed broadband subscriptions per 100 inhabitants. The International Telecommunication Union is the primary source of data in each case."*[37] *This variable is scaled from 0 which represents the lowest value to 1 which represents the highest value.*

20) **E-Government Index (0-1):** "*The E-Government Development Index presents the state of E-Government Development of the United Nations Member States. Along with an assessment of the website development patterns in a country, the E-Government Development index incorporates the access characteristics, such as the infrastructure and educational levels, to reflect how a country is using information technologies to promote access and inclusion of its people. The EGDI is a composite measure of three important dimensions of e-government, namely: provision of online*

[34] International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database, Available at: https://data.worldbank.org/indicator/IT.NET.USER.ZS
[35] Ibidem 12
[36] United Nations, Department of Economics and Public Affairs, Division for Public Institutions and Digital Government, Available at: https://publicadministration.un.org/egovkb/en-us/Data-Center
[37] Ibidem

services, telecommunication connectivity and human capacity."[38] *This variable is scaled from 0 which represents the lowest value to 1 which represents the highest value.*

21) **Control of Corruption (-2.5 - 2.5): "***Control of Corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 which represents the lowest score to 2.5 which represent the highest score, meaning free of corruption"*[39].

22) **Individuals using the internet to interact with public authorities (% of population): "***Within the last 12 months before the survey for private purposes. Derived variable on use of eGovernment services. Individuals used at least one of the following services: for obtaining information from public authorities websites, for downloading official forms, for submitting completed forms"*[40].

Limitations of study

The author of this paper is cognizant of the limitations that this research work is subject to, and thus makes no claim to provide any thorough quantitative results nor draw any definitive conclusions at this stage of the study despite its data – driven nature. Rather, the goal of this paper is to make use of exploratory statistical tools to scratch the surface of an ample research topic, which is of great interest in the operating field of the author. Through these preliminary statistical analyses the author aims to pique the interest of the reader by generating potential hypotheses and attempting to explore through more concrete

[38]United Nations, Department of Economics and Public Affairs, Division for Public Institutions and Digital Government, Available at: https://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government-Development-Index

[39] World Bank Group, Worldwide Governance Indicators, Available at: https://databank.worldbank.org/source/worldwide-governance-indicators/preview/on

[40] Eurostat, Data Browser, Available at: https://ec.europa.eu/eurostat/databrowser/view/tin00012/default/table?lang=en

and unconventional means declarations and concepts which are recently at the center of attention, also as a result of the unforeseen circumstances imposed by the COVID-19 global health emergency. Nevertheless, as previously stated, due to the reductionist approach taken in this study, the results obtained cannot be considered reliable neither to confirm nor disprove the claims made in relation to a potential nexus between digitalization and corruption in the Public Administration. This paper aims however, to shed light upon the difficulty of drawing conclusions relative to this topic due to the extremely complex nature of both variables to be analyzed, and aims to sow the seeds for further in-depth studies in the field.

The main limitations encountered regard the scope of study and the variables chosen for the creation of an ad-hoc database to carry out the various statistical analyses.

1) **Time constraints and the width of scope** – for the purposes of this study the author cannot fully explore the scope of the issue due to the limited amount of time at hand and the ampleness of the research topic. Considering the dynamic nature of both variables, corruption and digitalization, mid-long term observations would be necessary in order to accumulate sufficient empirical data. This might be interesting to follow up in further studies.
2) **The existing literature** – the literature consulted poses limits in terms of time frame and variables considered. Furthermore, the sources taken into consideration do not provide sufficient empirical evidence to provide a foundation for a reliable statistical model.
3) **Units of observation** – the units of observation not always contain sufficient data for each variable chosen. Therefore, the database may contain missing values as a result. This needs to be taken into consideration when the data is inserted into each statistical model. However, in order to account for missing values a **K-nearest neighbours** algorithm was used, which is the simplest algorithm used in **machine learning**. It allows the user to recognize a pattern of object classification based on the features and characteristics of the objects nearest to the missing one.

4) **Variables chosen** – both latent variables (corruption and digitalization) are of an extremely complex and composite nature, which means that they cannot be quantified through a single, objective unit of measurement but rather need to be expressed either through a series of other indicators or an "artifice" variable such as CPI. Against this background, it is inevitable that the choice of variables included in the database is discretionary and presents only a sample of the indicators which may be used to describe the two latent variables. On the subject of corruption, the indicators taken into consideration are derived through both direct and indirect measures of corruption. Measures of corruption may pose restrictions due to stigma bias and reporting bias. Furthermore, there may be a gap between facts and perception. This study will not focus on the methodology through which each variable was obtained, nor the reasons behind each measure of corruption (e.g. main reasons behind bribes), but will take the variables as a given. Lastly, this study will not distinguish between direct and indirect measures, nor exogenous factors which may have influenced indicators based on perception. Analogous limitations may apply to the latent variable digitalization, where there are objective elements such as lack of access to digital tools due to inadequate infrastructure or more subjective elements related to lack of skills and competencies which make citizen interaction more difficult.
5) **The unit of measurement and type of variables** – the variables taken into account are quantitative; nevertheless they are not all of the same unit of measurement and in order to avoid discrepancies during the analyses, the database will be scaled. Furthermore, it is important to consider that for the purposes of this study no distinction will be made among composite indicators, raw scores, indices or previously scaled variables.

Data Analysis and Discussion

In **Figure 1** the reader can find the dataset created specifically for the purposes of this study. It contains 65 units of observation among which European Union member states and LAC countries. The dataset also contains 22 variables selected at the discretion of the author as deemed representative of the two latent variables (corruption and digitalization) and fit for the aim of the analyses.

Country	GDP (curren	Corruptior	Public Sect	Public Sect	Bribery Inc	Informal p	Executive (	Executive (	Absence o	Political Cc	Impact Of	Availabilit	Internet a	ICT use &	Fixed broa	Individuals	E-Participa	Online Ser	Telecomm	E-Governn	Control of	Individuals
Austria	428965.40	76.00	0.72	0.86	.	.	0.82	0.83	0.76	0.05	5.95	6.09	5.29	4.79	28.13	88.00	0.63	0.95	0.82	0.89	1.54	72.00
Belgium	515332.50	76.00	0.79	0.79	4.30	0.90	0.84	0.92	0.80	0.03	5.71	6.19	5.62	4.37	39.78	90.00	0.63	0.66	0.80	0.80	1.55	62.90
Bulgaria	69105.10	44.00	0.36	0.33	10.00	13.40	0.56	0.34	0.41	0.66	4.26	4.64	4.74	3.98	28.78	68.00	0.25	0.77	0.78	0.80	-0.15	27.00
Croatia	55966.58	47.00	0.45	0.53	10.10	3.80	0.65	0.58	0.53	0.32	4.34	4.97	4.62	3.58	27.96	79.00	0.33	0.75	0.73	0.77	0.12	41.00
Cyprus	23804.34	57.00	0.56	0.69	0.20	10.90	0.72	0.57	0.61	0.18	4.63	5.17	5.00	3.92	37.79	86.00	0.31	0.87	0.91	0.87	0.60	53.00
Czech Repu	243530.38	54.00	0.56	0.56	3.50	19.90	0.51	0.50	0.54	0.25	4.92	5.63	5.38	3.77	34.98	81.00	0.25	0.72	0.81	0.81	0.50	57.00
Denmark	355184.02	88.00	0.90	0.88	.	.	0.90	0.92	0.88	0.01	5.72	5.97	5.93	5.07	43.95	98.00	0.55	0.97	1.00	0.98	2.10	91.89
Estonia	31029.97	75.00	0.81	0.79	0.00	3.40	0.85	0.83	0.79	0.04	5.88	5.77	6.10	5.84	32.53	90.00	0.76	0.99	0.92	0.95	1.54	80.00
Finland	271233.88	85.00	0.74	0.87	.	.	0.87	0.84	0.79	0.03	5.66	6.60	5.96	4.96	32.48	90.00	0.71	0.97	0.91	0.95	2.15	88.00
France	2603004.40	69.00	0.74	0.81	.	.	0.70	0.76	0.73	0.09	5.44	6.05	4.88	4.77	45.69	83.00	0.96	0.88	0.87	0.87	1.30	75.00
Germany	3806060.14	80.00	0.78	0.86	.	.	0.86	0.76	0.78	0.03	5.76	6.22	4.96	4.77	41.99	88.00	0.71	0.74	0.89	0.85	1.90	66.00
Greece	189410.11	50.00	0.48	0.60	5.50	5.20	0.71	0.56	0.56	0.30	3.96	4.99	3.93	3.50	39.62	76.00	0.80	0.71	0.81	0.80	-0.01	53.00
Hungary	155012.93	44.00	0.54	0.71	4.70	20.20	0.37	0.38	0.53	0.49	4.34	5.15	4.85	3.90	31.72	80.00	0.45	0.75	0.73	0.77	0.00	60.00
Ireland	418621.82	72.00	0.74	0.80	.	8.30	0.81	0.85	0.76	0.06	4.95	6.11	5.32	4.74	29.95	85.00	0.65	0.77	0.81	0.84	1.45	62.00
Italy	1886445.27	53.00	0.61	0.62	11.80	8.40	0.77	0.77	0.64	0.20	3.89	5.10	3.90	4.43	28.85	74.00	0.78	0.83	0.79	0.82	0.24	29.00
Latvia	33505.19	57.00	0.73	0.59	2.00	4.40	0.81	0.89	0.71	0.15	5.05	5.77	5.69	4.38	26.69	86.00	0.71	0.58	0.84	0.78	0.48	76.00
Lithuania	55887.27	60.00	0.62	0.76	0.20	6.30	0.91	0.73	0.68	0.14	5.34	5.76	5.99	4.85	28.69	82.00	0.65	0.85	0.82	0.87	0.68	58.00
Luxembour	73263.98	80.00	0.67	0.87	0.00	0.50	0.93	0.89	0.83	0.08	5.95	6.20	5.63	5.53	37.37	97.00	0.55	0.76	0.91	0.83	2.10	63.00
Malta	14647.38	53.00	.	.	2.40	0.00	.	.	.	0.26	5.27	5.35	5.63	4.82	45.99	86.00	0.47	0.81	0.93	0.85	0.24	55.00
Netherlanc	912242.34	82.00	0.85	0.85	.	.	0.92	0.84	0.86	0.02	6.17	6.30	6.13	4.96	43.63	93.00	1.00	0.91	0.93	0.92	1.99	86.00
Poland	594164.69	56.00	0.62	0.71	2.80	10.40	0.75	0.72	0.67	0.09	3.94	4.64	4.78	3.48	20.54	85.00	0.49	0.86	0.80	0.85	0.60	42.00
Portugal	231255.59	61.00	0.64	0.71	0.30	2.80	0.86	0.86	0.72	0.16	5.57	6.10	5.36	5.37	38.80	75.00	0.65	0.84	0.79	0.82	0.76	45.00
Romania	248715.55	44.00	0.52	0.45	5.40	15.60	0.51	0.53	0.49	0.56	3.76	4.65	4.78	3.37	27.25	74.00	0.47	0.72	0.76	0.76	-0.13	13.00
Slovakia	104574.15	49.00	0.58	0.57	2.40	25.30	0.59	0.57	0.56	0.27	4.33	5.47	5.32	3.81	29.05	83.00	0.63	0.83	0.80	0.78	0.33	62.00
Slovenia	52880.47	60.00	0.69	0.69	1.80	4.40	0.75	0.71	0.68	0.26	4.99	5.47	5.68	3.99	30.21	83.00	0.39	0.85	0.79	0.85	0.91	67.00
Spain	52880.47	62.00	0.77	0.93	.	4.40	0.94	0.82	0.86	0.04	5.02	5.52	4.32	4.21	33.41	91.00	0.78	0.89	0.85	0.88	0.65	63.00
Sweden	1281199.09	85.00	0.85	0.80	1.90	.	0.86	0.96	0.83	0.02	6.05	6.48	6.26	5.25	40.24	94.00	0.61	0.90	0.96	0.94	2.12	86.00
Antigua an	1415.06	.	.	.	6.90	4.80	.	.	.	.	.	.	.	.	9.43	73.00	.	0.45	0.62	0.61	0.28	.
Argentina	383066.98	42.00	0.41	0.60	9.30	6.80	0.58	0.43	0.52	0.43	3.52	3.73	4.13	2.98	19.64	74.00	0.55	0.85	0.73	0.83	-0.07	.
Aruba	3202.19	.	.	.	.	.	.	.	.	.	.	.	.	.	18.21	97.00	.	.	.	.	1.21	.
The Baham	11250.00	63.00	.	.	21.00	19.10	.	.	.	.	.	.	.	.	21.13	85.00	.	0.68	0.67	0.70	1.18	.
Barbados	4365.50	64.00	0.76	0.65	1.20	14.70	0.80	0.78	0.73	0.07	4.47	5.81	5.04	3.82	37.21	82.00	0.10	0.58	0.75	0.73	1.26	.
Belize	1763.70	.	.	.	6.20	3.00	.	.	.	.	.	.	.	.	7.58	47.00	.	0.26	0.41	0.45	-0.19	.
Bolivia	36688.86	31.00	0.46	0.44	9.10	19.20	0.52	0.42	0.47	0.62	3.76	3.80	3.50	3.35	6.49	44.00	0.41	0.58	0.52	0.61	-0.74	.
Brazil	1444733.26	38.00	0.48	0.60	11.70	12.50	0.48	0.42	0.51	0.51	3.50	4.47	3.63	3.38	15.60	70.00	0.71	0.87	0.65	0.77	-0.33	52.79
Cayman Isl	5935.77	.	.	.	.	.	.	.	.	.	.	.	.	.	49.28	81.00	.	.	.	.	0.47	.
Chile	252940.02	67.00	0.68	0.81	1.30	0.70	0.88	0.80	0.75	0.06	4.99	5.59	4.76	4.46	18.09	82.00	0.94	0.85	0.76	0.83	1.08	47.35
Colombia	271346.90	39.00	0.40	0.50	6.60	10.80	0.64	0.59	0.50	0.43	4.21	4.49	4.11	4.18	13.81	65.00	0.88	0.76	0.61	0.72	-0.23	39.23
Costa Rica	61520.67	57.00	0.59	0.65	8.70	3.70	0.61	0.62	0.60	0.18	4.63	4.88	4.68	3.79	17.92	81.00	0.82	0.68	0.75	0.76	0.72	53.68
Cuba	103131.00	47.00	0.30	0.14	.	.	0.33	0.72	0.33	0.40	.	.	.	.	1.61	62.00	.	0.26	0.25	0.44	0.04	.
Curacao	3101.79	.	.	.	.	.	.	.	.	.	.	.	.	.	32.61	68.00	.	.	.	.	.	.
Dominica	469.87	55.00	.	.	5.80	0.00	.	.	.	.	.	.	.	.	16.08	70.00	.	0.45	0.69	0.60	0.53	.
Dominican	78844.70	28.00	0.04	0.08	12.30	15.30	0.21	0.38	0.17	0.88	4.07	4.90	3.49	3.79	8.86	75.00	0.33	0.76	0.53	0.68	-0.76	.
Ecuador	98808.01	39.00	0.45	0.57	5.90	4.40	0.54	0.47	0.51	0.44	4.35	4.52	4.12	4.14	12.04	54.00	0.49	0.81	0.51	0.70	-0.49	.
El Salvador	24638.72	36.00	0.42	0.31	4.20	5.50	0.46	0.49	0.44	0.62	4.01	4.32	3.62	3.44	7.67	34.00	0.61	0.58	0.51	0.57	-0.55	.
Grenada	1089.20	53.00	.	.	6.60	7.10	.	.	.	.	.	.	.	.	22.84	59.00	.	0.34	0.54	0.58	0.34	.
Guatemala	77604.63	25.00	0.30	0.31	2.80	8.50	0.31	0.31	0.33	0.81	4.32	5.33	3.62	3.45	3.14	41.00	0.20	0.51	0.48	0.52	-0.89	.
Guyana	5471.26	41.00	.	.	14.90	18.40	.	.	.	0.29	3.88	4.46	4.11	3.67	8.37	37.00	0.33	0.46	0.36	0.49	-0.09	.
Haiti	13418.00	18.00	0.22	0.17	.	.	0.24	0.19	0.22	0.91	2.49	3.44	2.68	2.27	0.28	32.00	0.18	0.19	0.24	0.27	-1.34	.
Honduras	23827.84	24.00	0.23	0.21	8.70	10.90	0.29	0.31	0.27	0.85	4.22	4.92	3.94	3.85	4.01	32.00	0.33	0.46	0.32	0.45	-0.80	.
Jamaica	13812.42	44.00	0.67	0.69	19.30	17.90	0.74	0.63	0.64	0.12	4.03	5.19	4.15	3.73	10.78	55.00	0.20	0.39	0.52	0.54	-0.06	.
Mexico	1076163.32	31.00	0.45	0.47	17.60	11.60	0.51	0.46	0.48	0.61	3.98	4.95	3.88	3.88	15.17	70.00	0.61	0.82	0.59	0.73	-0.82	62.94
Nicaragua	12621.48	22.00	0.29	0.28	6.50	10.00	0.24	0.25	0.30	0.87	3.16	4.00	2.69	2.83	3.32	28.00	0.10	0.55	0.38	0.51	-1.12	.
Panama	52938.10	35.00	0.57	0.45	7.10	30.50	0.54	0.48	0.50	0.55	4.75	5.52	4.72	4.46	12.33	64.00	0.49	0.62	0.65	0.67	-0.58	.
Paraguay	35304.24	28.00	0.33	0.42	13.80	8.20	0.34	0.39	0.40	0.76	3.37	4.01	2.92	3.04	5.36	69.00	0.25	0.71	0.54	0.65	-0.83	.
Peru	202014.36	38.00	0.36	0.55	17.60	13.60	0.65	0.59	0.50	0.48	3.96	4.47	3.71	3.38	7.35	60.00	0.71	0.75	0.58	0.71	-0.45	.
Puerto Ric	103138.30	.	.	.	.	.	.	.	.	.	4.62	6.08	4.79	4.15	20.79	71.00	0.71	.	.	.	0.01	.
St. Kitts an	927.45	.	.	.	2.20	6.10	.	.	.	.	.	.	.	.	16.65	81.00	.	0.39	0.71	0.64	0.39	.
St. Lucia	1703.36	56.00	.	.	11.60	0.40	.	.	.	.	.	.	.	.	17.74	51.00	.	0.38	0.53	0.54	0.52	.
St. Vincent	809.68	59.00	.	.	4.90	2.80	.	.	.	.	.	.	.	.	20.34	21.00	.	0.47	0.49	0.56	0.78	.
Suriname	3807.92	38.00	.	.	12.70	24.80	.	.	.	0.49	2.97	4.19	2.75	2.98	13.82	49.00	0.14	0.29	0.55	0.52	-0.39	.
Trinidad ar	21529.82	40.00	0.58	0.75	10.50	13.40	0.79	0.71	0.66	0.11	3.79	4.95	4.46	3.55	24.33	77.00	0.31	0.61	0.68	0.68	-0.19	.
Turks and (	924.58	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.00	.	.	.	.	.	.
Uruguay	53628.83	71.00	0.65	0.79	2.40	2.00	0.79	0.88	0.74	0.05	4.76	4.78	5.68	4.10	29.25	77.00	0.98	0.84	0.86	0.85	1.25	.
Venezuela	47260.00	15.00	0.02	0.02	10.30	23.70	0.07	0.06	0.04	0.87	3.31	3.33	3.48	2.56	8.98	64.00	0.57	0.32	0.48	0.53	-1.51	.

Figure 1

Cluster Analysis

The original dataset was inserted into the statistical analysis software RStudio, in order to carry out a **Hierarchical Cluster Analysis** more specifically, an **Agglomerative Nesting** method was attempted with the aim of detecting potentially meaningful patterns within the data and consequently forming cohesive groups to illustrate the identified trends. These groups known as clusters would then need to be interpreted accordingly based on the authors knowledge and insight relative to the variables with the intention of generating hypotheses to be confirmed or disproved. However, the outcome obtained by the first attempt was not satisfactory due to an elevated number of missing values for certain variables. In some cases, the number of missing values reached 33, which negatively and significantly impacted the outcome produced by the function algorithm since the missing data values could not be omitted or substituted by the median of the variable of interest. Therefore, in order to avoid biases and further obstacles, the author proceeded to reduce the database both in terms of units of observation (countries) by omitting the ones which lacked the highest number of values per variable, as well as variables that contained a shortage of values, or added little to no additional value to the dataset as were contained within other composite variables.

Figure 2 shows the reduced dataset which was used to carry out the **Cluster Analysis** which will be interpreted in this section of the study. It should be noted that the reduced dataset no longer contains missing values because a **K-nearest neighbours** algorithm was used to account for them (see Methodology for further details).

Country	GDP..current.U	Corruption.Per	Public.Sector.C	Public.Sector.TI	Bribery.Inciden	Informal.payme	Absence.of.con	Impact.Of.ICTs.	Availability.of.I	ICT.use...gov.t.e	Fixed.broadban	Individuals.usin	E.Participation.	Online.Service.I	Telecommunicz	E.Government.I	Control.of.Corruption.
Austria	428965.4	76	0.72	0.86	0.2	3.4	0.76	5.95	6.09	4.79	28.13	88	0.63	0.95	0.82	0.89	1.54
Belgium	515332.5	76	0.79	0.79	4.3	0.9	0.8	5.71	6.19	4.37	39.78	90	0.63	0.66	0.8	0.8	1.55
Bulgaria	69105.1	44	0.36	0.33	10	13.4	0.41	4.26	4.64	3.98	28.78	68	0.25	0.77	0.78	0.8	-0.15
Croatia	55966.58	47	0.45	0.53	10.1	3.8	0.53	4.34	4.97	3.58	27.96	79	0.33	0.75	0.73	0.77	0.12
Cyprus	23804.34	57	0.56	0.69	0.2	10.9	0.61	4.63	5.17	3.92	37.79	86	0.31	0.87	0.91	0.87	0.6
Czech Repub	243530.38	54	0.56	0.56	3.5	19.9	0.54	4.92	5.63	3.77	34.98	81	0.25	0.72	0.81	0.81	0.5
Denmark	355184.02	88	0.9	0.88	1.2	3.4	0.88	5.72	5.97	5.07	43.95	98	0.55	0.97	1	0.98	2.1
Estonia	31029.97	75	0.81	0.79	0	3.4	0.79	5.88	5.77	5.84	32.53	90	0.76	0.99	0.92	0.95	1.54
Finland	271233.88	85	0.74	0.87	0.2	3.4	0.79	5.66	6.6	4.96	32.48	90	0.71	0.97	0.91	0.95	2.15
France	2603004.4	69	0.74	0.81	1.3	2.8	0.73	5.44	6.05	4.77	45.69	83	0.96	0.88	0.87	0.87	1.3
Germany	3806060.14	80	0.78	0.86	0.3	2.8	0.78	5.76	6.22	4.77	41.99	88	0.71	0.74	0.89	0.85	1.9
Greece	189410.11	50	0.48	0.6	5.5	5.2	0.56	3.96	4.99	3.5	39.62	76	0.8	0.71	0.81	0.8	-0.01
Hungary	155012.93	44	0.54	0.71	4.7	20.2	0.53	4.34	5.15	3.9	31.72	80	0.45	0.75	0.73	0.77	0
Ireland	418621.82	72	0.74	0.8	1.3	8.3	0.76	4.95	6.11	4.74	29.95	85	0.65	0.77	0.81	0.84	1.45
Italy	1886445.27	53	0.61	0.62	11.8	8.4	0.64	3.89	5.1	4.43	28.85	74	0.78	0.83	0.79	0.82	0.24
Latvia	33505.19	57	0.73	0.59	2	4.4	0.71	5.05	5.77	4.38	26.69	86	0.71	0.58	0.84	0.78	0.48
Lithuania	55887.27	60	0.62	0.76	0.2	6.3	0.68	5.34	5.76	4.85	28.69	82	0.65	0.85	0.82	0.87	0.68
Luxembourg	73263.98	80	0.67	0.87	0	0.5	0.83	5.95	6.2	5.53	37.37	97	0.55	0.76	0.91	0.83	2.1
Malta	14647.38	53	0.64	0.69	2.4	0	0.68	5.27	5.35	4.82	45.99	86	0.47	0.81	0.93	0.85	0.24
Netherlands	912242.34	82	0.85	0.85	1.3	3.4	0.86	6.17	6.3	4.96	43.63	93	1	0.91	0.93	0.92	1.99
Poland	594164.69	56	0.62	0.71	2.8	10.4	0.67	3.94	4.64	3.48	20.54	85	0.49	0.86	0.8	0.85	0.6
Portugal	231255.59	61	0.64	0.71	0.3	2.8	0.72	5.57	6.1	5.37	38.8	75	0.65	0.84	0.79	0.82	0.76
Romania	248715.55	44	0.52	0.45	5.4	15.6	0.49	3.76	4.65	3.37	27.25	74	0.47	0.72	0.76	0.76	-0.13
Slovakia	104574.15	49	0.58	0.57	2.4	25.3	0.56	4.33	5.47	3.81	29.05	83	0.63	0.83	0.8	0.78	0.33
Slovenia	52880.47	60	0.69	0.69	1.8	4.4	0.68	4.99	5.47	3.99	30.21	83	0.39	0.85	0.79	0.85	0.91
Spain	52880.47	62	0.77	0.93	1.8	4.4	0.86	5.02	5.52	4.21	33.41	91	0.78	0.89	0.85	0.88	0.65
Sweden	1281199.09	85	0.85	0.8	1.9	0.9	0.83	6.05	6.48	5.25	40.24	94	0.61	0.9	0.96	0.94	2.12
Antigua and	1415.06	53	0.64	0.65	6.9	4.8	0.71	4.63	4.88	3.79	9.43	73	0.33	0.45	0.62	0.61	0.28
Argentina	383066.98	42	0.41	0.6	9.3	6.8	0.52	3.52	3.73	2.98	19.64	74	0.55	0.85	0.73	0.83	-0.07
The Bahama:	11250	63	0.68	0.65	21	19.1	0.71	4.95	5.59	3.99	21.13	85	0.39	0.68	0.67	0.7	1.18
Barbados	4365.5	64	0.76	0.65	1.2	14.7	0.73	4.47	5.81	3.82	37.21	82	0.1	0.58	0.75	0.73	1.26
Belize	1763.7	53	0.64	0.57	6.2	5	0.5	4.03	4.49	3.67	7.58	47	0.33	0.26	0.41	0.45	-0.19
Bolivia	36688.86	31	0.46	0.44	9.1	19.2	0.47	3.76	3.8	3.35	6.49	44	0.41	0.58	0.52	0.61	-0.74
Brazil	1444733.26	38	0.48	0.6	11.7	12.5	0.51	3.5	4.47	3.38	15.6	70	0.71	0.87	0.65	0.77	-0.33
Chile	252940.02	67	0.68	0.81	1.3	0.7	0.75	4.99	5.59	4.46	18.09	82	0.94	0.85	0.76	0.83	1.08
Colombia	271346.9	39	0.4	0.5	6.6	10.8	0.5	4.21	4.49	4.18	13.81	65	0.88	0.76	0.61	0.72	-0.23
Costa Rica	61520.67	57	0.59	0.65	8.7	3.7	0.6	4.63	4.88	3.79	17.92	81	0.82	0.68	0.75	0.76	0.72
Cuba	103131	47	0.3	0.14	6.9	7.1	0.33	4.21	4.49	3.67	1.61	62	0.33	0.26	0.25	0.44	0.04
Dominican R	78844.7	28	0.04	0.08	12.3	15.3	0.17	4.07	4.9	3.79	8.86	75	0.33	0.76	0.53	0.68	-0.76
Ecuador	98808.01	39	0.45	0.57	5.9	4.4	0.51	4.35	4.52	4.14	12.04	54	0.49	0.81	0.51	0.7	-0.49
El Salvador	24638.72	36	0.42	0.31	4.2	5.5	0.44	4.01	4.32	3.44	7.67	34	0.61	0.58	0.51	0.57	-0.55
Grenada	1089.2	53	0.64	0.65	6.6	7.1	0.68	4.63	4.88	3.79	22.84	59	0.33	0.34	0.54	0.58	0.34
Guatemala	77604.63	25	0.3	0.31	2.8	8.5	0.33	4.32	5.33	3.45	3.14	41	0.2	0.51	0.48	0.52	-0.89
Guyana	5471.26	41	0.46	0.44	14.9	18.4	0.47	3.88	4.46	3.67	8.37	37	0.33	0.46	0.36	0.49	-0.09
Haiti	13418	18	0.22	0.17	8.7	10.9	0.22	2.49	3.44	2.27	0.28	32	0.18	0.19	0.24	0.27	-1.34
Honduras	23827.84	24	0.23	0.21	8.7	10.9	0.27	4.22	4.92	3.85	4.01	32	0.33	0.46	0.32	0.45	-0.8
Jamaica	13812.42	44	0.67	0.69	19.3	17.9	0.64	4.03	5.19	3.73	10.78	55	0.2	0.39	0.52	0.54	-0.06
Mexico	1076163.32	31	0.45	0.47	17.6	11.6	0.48	3.98	4.95	3.88	15.17	70	0.61	0.82	0.59	0.73	-0.82
Nicaragua	12621.48	22	0.29	0.28	6.5	10	0.3	3.16	4	2.83	3.32	28	0.1	0.55	0.38	0.51	-1.12
Panama	52938.1	35	0.57	0.45	7.1	30.5	0.5	4.75	5.52	4.46	12.33	64	0.49	0.62	0.65	0.67	-0.58
Paraguay	35304.24	28	0.33	0.42	13.8	8.2	0.4	3.37	4.01	3.04	5.36	69	0.25	0.71	0.54	0.65	-0.83
Peru	202014.36	38	0.36	0.55	17.6	13.6	0.5	3.96	4.47	3.38	7.35	60	0.71	0.75	0.58	0.71	-0.45
Suriname	3807.92	38	0.48	0.6	12.7	24.8	0.47	2.97	4.19	2.98	13.82	49	0.14	0.29	0.55	0.52	-0.39
Trinidad and	21529.82	40	0.58	0.75	10.5	13.4	0.66	3.79	4.95	3.55	24.33	77	0.31	0.61	0.68	0.68	-0.19
Uruguay	53628.83	71	0.65	0.79	2.4	2	0.74	4.76	4.78	4.1	29.25	77	0.98	0.84	0.86	0.85	1.25
Venezuela	47260	15	0.02	0.02	10.3	23.7	0.04	3.31	3.33	2.56	8.98	64	0.57	0.32	0.48	0.53	-1.51

Figure 2

Prior to proceeding with the **Hierarchical Cluster Analysis** the data was scaled in order to account for differences in unit of measurement.

The aim of a cluster analysis is to identify cohesive grouped structures within a sample dataset, minimizing the distance between the members within a cluster and maximizing the dissimilarity between two distinct clusters. There are several methods which can be used to identify the linkage between the observation in order to classify them within similar or dissimilar clusters. For the purposes of this study, the **Ward method** was used.

Figure 3 shows the **Dendrogram** which is the illustrative representation of an **Agglomerative Nesting Hierarchical Cluster Analysis**. Considering that a hierarchical method is being used the number of clusters to be retained is not predetermined, but is derived through the graph. Furthermore, it should be noted that in an agglomerative method each observation is first represented as a single cluster (**leaf**) and as the aggregation process is repeated, the two clusters that are the most similar are combined into a new bigger cluster which is otherwise known as a node. The process is repeated until all elements are members of one big cluster known as a root.

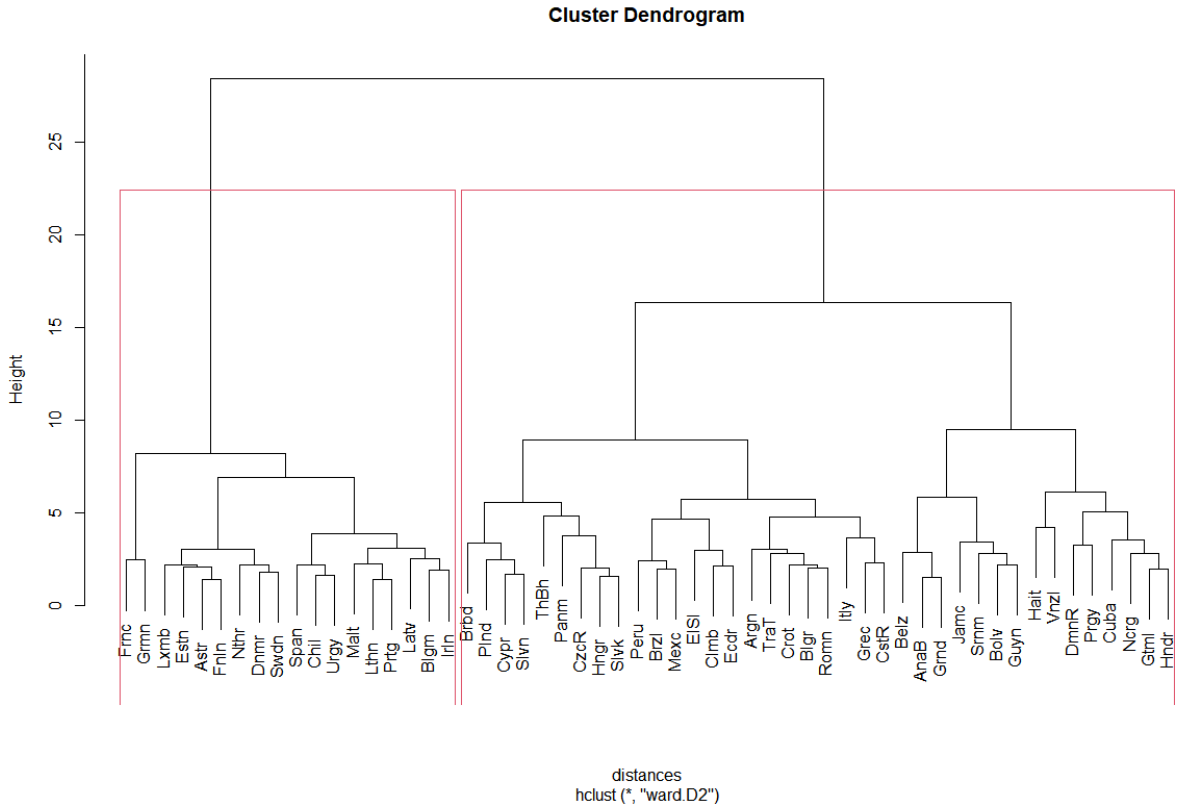

Figure 3

The dendrogram is represented along a vertical axis which indicates the height i.e. the distance between the distinct clusters. The greater the difference in height, the more dissimilarity; hence, elements of a similar height are grouped within the same cluster. The data used to carry out the cluster analysis seems to have been clustered into two distinct groups through the method know as **ward cut** which maximizes the dissimilarity between the different clusters. (see red rectangular markings in Figure 3). The number of clusters to

be retained can also be obtained through a plot indicating the average silhouette width. The average silhouette width increases as the number of clusters increases; therefore, it is reasonable to choose a number of clusters which maximizes the average silhouette width or stop at a number of clusters when the silhouette width no longer increases significantly. As can be visualized through the average silhouette width plot below (Figure 4), the optimal number of clusters is two. This indicates that the units of observation belong to two main groups as shown in Table 1 below; within the first cluster there are 18 countries, only two of which are LAC countries; namely, Chile and Uruguay. It is interesting to note that both these countries have made enormous efforts and progress towards digitalization in the Public Administration. Instead the second cluster contains 38 observations including a mix of EU member states and LAC countries.

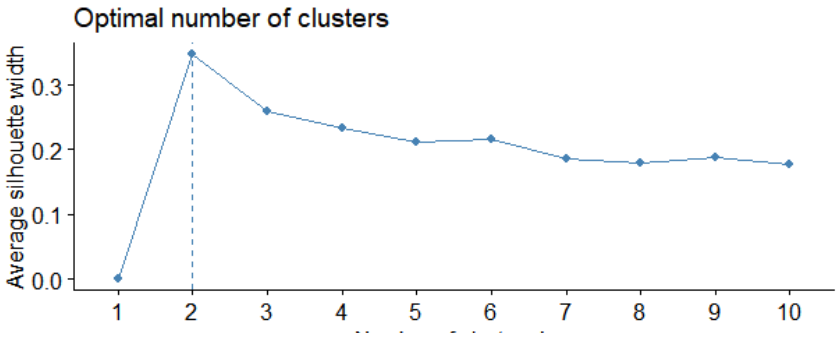

Figure 4

Original_name	Abbreviated_name	Cluster
Austria	Astr	1
Belgium	Blgm	1
Bulgaria	Blgr	2
Croatia	Crot	2
Cyprus	Cypr	2
Czech Republic	CzcR	2
Denmark	Dnmr	1
Estonia	Estn	1
Finland	Fnln	1
France	Frnc	1
Germany	Grmn	1
Greece	Grec	2
Hungary	Hngr	2
Ireland	Irln	1
Italy	Itly	2
Latvia	Latv	1
Lithuania	Lthn	1
Luxembourg	Lxmb	1
Malta	Malt	1
Netherlands	Nthr	1
Poland	Plnd	2
Portugal	Prtg	1
Romania	Romn	2
Slovakia	Slvk	2
Slovenia	Slvn	2
Spain	Span	1
Sweden	Swdn	1
Antigua and Barbuda	AnaB	2
Argentina	Argn	2
The Bahamas	ThBh	2
Barbados	Brbd	2
Belize	Belz	2
Bolivia	Bolv	2
Brazil	Brzl	2
Chile	Chil	1
Colombia	Clmb	2
Costa Rica	CstR	2
Cuba	Cuba	2
Dominican Republic	DmnR	2
Ecuador	Ecdr	2
El Salvador	ElSl	2
Grenada	Grnd	2
Guatemala	Gtml	2
Guyana	Guyn	2
Haiti	Hait	2
Honduras	Hndr	2
Jamaica	Jamc	2
Mexico	Mexc	2
Nicaragua	Ncrg	2
Panama	Panm	2
Paraguay	Prgy	2
Peru	Peru	2
Suriname	Srnm	2
Trinidad and Tobago	TraT	2
Uruguay	Urgy	1
Venezuela	Vnzl	2

Table 1

Expected Findings

Taking into consideration the dataset with the units of observation and variables included, the author and reader might expect the objects to be clustered according to the following criteria: i) Geographical region→ considering socio-economic and institutional elements

ii) Level of national income→ although the use of GDP as a clustering criteria is often criticized for its ambiguity, it is in many cases a strong influential factor.

iii) Level of corruption→ The level of corruption might influence the clustering of observations either in a positive or negative way

iv) Degree of digitalization→ The degree of digitalization might influence the clustering of observations either in a positive or negative way.

Potential hypothesis: Digitalization plays a role in the reduction of corruption in the Public Administration.

The author expects the variables digitalization and corruption to interact in the clustering of observations, meaning that countries with a low level of corruption and a high level of digitalization be clustered together; these countries would be the **best-in-class** performers regardless of their geographical location. On the other hand, the author expects countries with a high level of corruption and a low level of digitalization to be grouped together; these countries would instead be known as **worst-in-class** performers. In order to further analyze the expected findings, the clusters will be compared amongst them selecting the most representative indicators for each latent variable i.e. digitalization and corruption; fully aware of the fact that these two composite variables can neither be represented nor measured through a single parameter. The purpose of this simple test is to illustrate in an explorative and preliminary way the basis upon which the units of observation are clustered. The author will carry out the same test also for level of national income in order to gauge the influence of this controversial criterium in the grouping of the observations. No test is needed as far as regional influence is concerned, as basic geographical knowledge will be used.

In order to test for the normality of the data included in the aforementioned variables a **D'Agostino-Pearson** normality test was carried out. For non normally distributed data a ranks **Mann-Whitney** test will be used. Instead for normally distributed data a **two-tailed unpaired t test with Welch's correction** will be used. The variables and parameters considered for testing are as follows:

Level of national income→ GDP (not normally distributed)

Corruption → Absence of corruption

Digitalization → E-Government Index

The results of the test are shown in Figure 5 below.

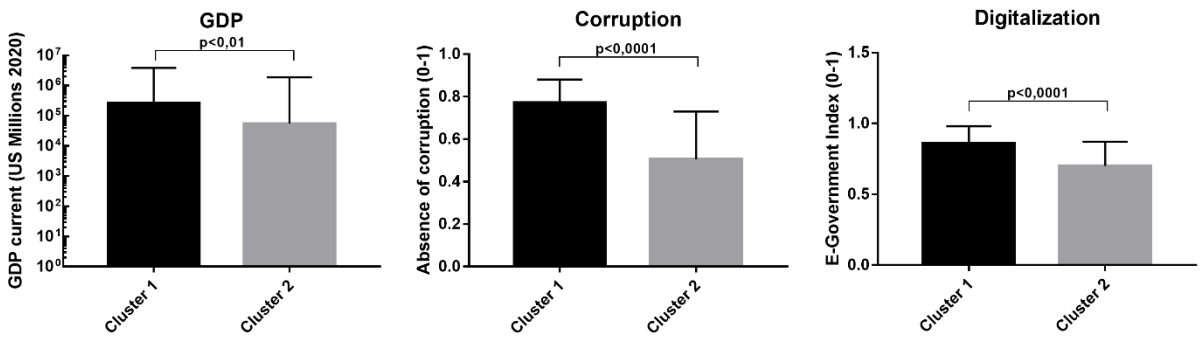

Figure 5

As expected, the three variables taken into consideration differ in a statistically significant manner between the two clusters. As far as GDP is concerned, it remains and ambiguous and confounding factor as it may influence both the level of digitalization as well as corruption in each unit of observation. Nevertheless, according to the test carried out, there is a statistical significance in the difference between the two clusters when GDP is considered. On the other hand, for both parameters used to represent digitalization and corruption respectively, the tests conducted highlight an even more prominent statistical significance when considering the difference between the two clusters. These results

obtained guide the author closer towards the adoption of a hypothesis which analyzes the interaction between digitalization and corruption in the two clusters, distinguishing between best-in-class and worst-in-class.

Furthermore, the dissimilarity between the two groups can be illustrated through a cluster plot produced during the cluster analysis. Figure 6 below depicts the degree of separation between the two clusters taking into account two different components. As shown by the cluster plot below, these two components alone explain 73.94% of the point variability. In order to identify the two components and to potentially confirm the hypothesis of digitalization playing a role in the reduction of corruption in the Public Administration, a **Principal Component Analysis** will be carried out.

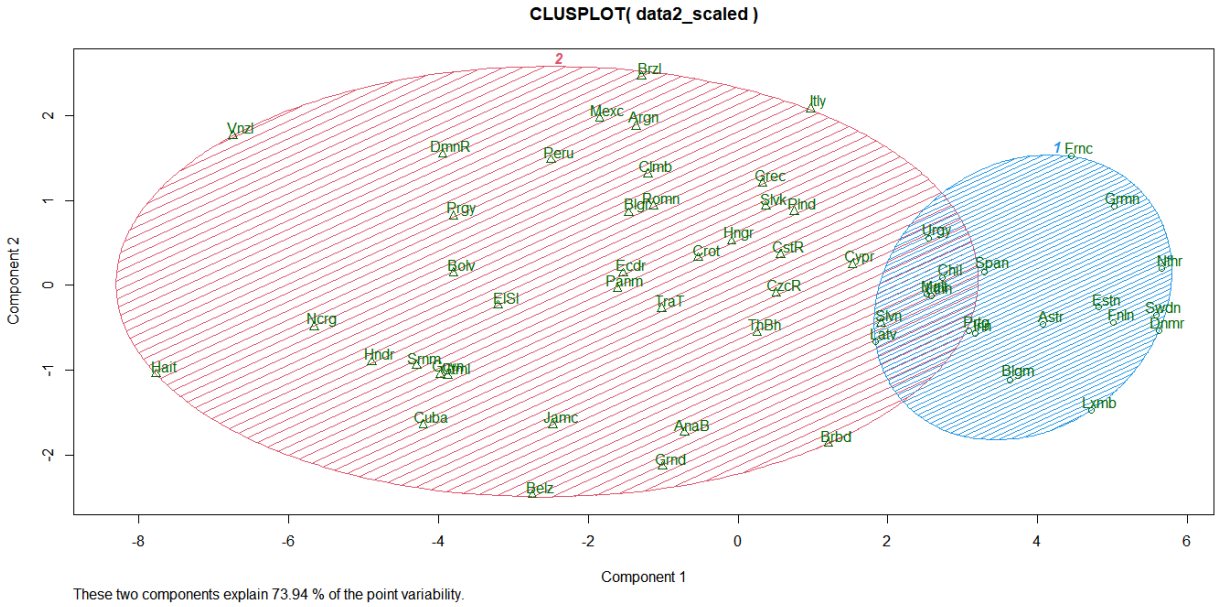

Figure 6

Principal Component Analysis

Latent variable: Performance of Public Administration

Expected Principal Components: Digitalization and Corruption

Expected Findings

The author expects the **Principal Component Analysis** to confirm the hypothesis generated by the **AGNES Hierarchical Cluster Analysis**; The level of digitalization is expected to influence the Performance of Public Administration positively, whereas the level of corruption is expected to influence the Performance of Public Administration negatively. Taking this interaction into consideration, digitalization could then be potentially considered as a tool in the reduction of corruption in the Public Administration in order to increase transparency, accountability and efficiency; all of which are parameters indicating high performance and quality service delivery. Nevertheless, the limitations of the data collected as well as the limitations of the model itself do not allow the study to provide an exhaustive analysis and definite conclusions through the model. The results and conclusions which will be drawn will serve rather for explorative and illustrative purposes as well as to provide greater insight of the patterns which were detected through the Cluster Analysis.

The database used for the **Principal Component Analysis** will be the same as the one used to carry out the **Cluster Analysis**. The missing values were accounted for through a **K-nearest neighbours** algorithm. The data was also scaled in order to eliminate the effects of diverging measurement units.

The purpose of a Principal Component Analysis is to shrink dimensions of a large set of variables to a smaller and more manageable set which still contains most of the information of the original larger set. As was stated in the Methodology section, it is important to note that although the principal components represent or replace one or more of the original variables, they are not just a one-to-one transformation; therefore, inverse transformations are not possible.

Once the data has been standardized, the author moves on to the analysis of the correlation matrix. The correlation function makes it possible to understand whether the variables are correlated amongst them and whether they could be an expression or representation of a characteristic which is not directly discernible. Once the correlation matrix has been obtained, the author proceeds to effectively check whether the variables that have been chosen are correlated amongst them and whether this correlation can reflect the descriptive summarizing variables the author aims to obtain, i.e. corruption and digitalization. According to the correlation matrix a sample of which can be seen in Figure 7 below, there is a high degree of correlation between the variables chosen (to facilitate the visualization of the high degree of correlation, some of them have been highlighted). The higher the correlation between the variables, the higher the possibility of obtaining a more contained number of variables, which will represent the principal components. Thanks to the high degree of correlation among the variables the author can expect to obtain a reduced number of variables capable of providing a large amount of information through a more succinct framework.

	Corruption.Perception.Index..2020.	Absence.of.corruption..0.1.	Public.Sector.Theft..0.1...2019.
GDP..current.US..Millions.2020.	0.3224393	0.2668273	0.2944326
Corruption.Perception.Index..2020.	1	0.9084953	0.8494716
Public.Sector.Corrupt.Exchanges.0.1...2019.	0.8882617	0.9627909	0.911347
Public.Sector.Theft..0.1...2019.	0.8494716	0.9546574	1
Bribery.Incidence....of.Firms.	-0.5927504	-0.4942321	-0.4651874
Informal.payments.to.public.officials....of.firms.	-0.5612016	-0.5129027	-0.4592005
Absence.of.corruption..0.1.	0.9084953	1	0.9546574
Impact.Of.ICTs.On.Access.To.Basic.Services..1.7.	0.8661971	0.7816197	0.6970831
Availability.of.latest.technologies..1.7.	0.8277919	0.7799957	0.7098838
ICT.use...gov.t.efficiency..1.7.	0.7995556	0.7488297	0.6740191
Fixed.broadband.subscriptions..per.100.people.	0.7827077	0.7489428	0.7335449
Individuals.using.the.Internet....of.population.	0.7794148	0.7320125	0.7116694
E.Participation.Index..0.1.	0.4923668	0.4742107	0.4813831
Online.Service.Index..0.1.	0.5560367	0.5646897	0.5954178
Telecommunication.Infrastructure.Index..0.1.	0.7925654	0.7891561	0.78569
E.Government.Index..0.1.	0.7533857	0.7396216	0.7471307
Control.of.Corruption...2.5...2.5.	0.9882764	0.8900732	0.828683

Figure 7

In order to define the principal components, the author moves on from the **correlation matrix** to the **loadings matrix**. The **loadings matrix** provides the coefficients for each principal component. Seeing as a principal component is a representation of more variables "grouped" together, each coefficient represents "the weight" that each original variable takes in the construction of each principal component. The aim is to reduce the number of variables as much as possible without losing valuable information. Once the loadings matrix with all the principal components has been obtained, the author needs to decide how many principal components to retain. This will be done by looking at the **cumulative proportion**. At this point the author is interested in confirming what was established through the cluster plot during the cluster analysis; that is, two components alone explain 73.94% of the **point variability**. The **cumulative proportion** depicted in **Figure 8** below will indicate whether the Principal Components which have emerged from the **Principal Component** are responsible for the same percentage of variation of the latent variable, i.e. performance of Public Administration. Nevertheless, at this stage it is not yet possible to support the hypothesis generated through the **Cluster Analysis** as the **Principal Components** have not been yet identified.

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	3.3633	1.12148	1.00812	0.93595	0.81815	0.74864
Proportion of Variance	0.6654	0.07398	0.05978	0.05153	0.03937	0.03297
Cumulative Proportion	0.6654	0.73936	0.79915	0.85068	0.89005	0.92302

Figure 8

The results obtained from the cumulative proportion at two principal components match the output of the Cluster Analysis.

Having confirmed the output obtained from the cluster plot, it is necessary to return to the **loadings matrix** which is represented by **Figure 9** and needs to be interpreted in order to determine the labels of each new Principal Component. This will be done by taking into consideration the **coefficients** from the **loading matrix**. The original variables with the most **statistically significant coefficients** will be the ones to determine the character of the principal components.

	PC1	PC2	PC3
GDP..current.US..Millions.2020.	-0.1126306	-0.37167443	-0.462772711
Corruption.Perception.Index..2020.	-0.2802862	0.18256029	-0.038304680
Public.Sector.Corrupt.Exchanges.0.1...2019.	-0.2594905	0.24928913	0.066678992
Public.Sector.Theft..0.1...2019.	-0.2619857	0.09006900	0.084688226
Bribery.Incidence....of.Firms.	0.1909400	-0.17183415	0.191728509
Informal.payments.to.public.officials....of.firms.	0.1650470	-0.13241640	0.601784885
Absence.of.corruption..0.1.	-0.2724560	0.17973876	0.053615612
Impact.Of.ICTs.On.Access.To.Basic.Services..1.7.	-0.2656213	0.18670970	-0.087223329
Availability.of.latest.technologies..1.7.	-0.2579976	0.19107681	0.041408209
ICT.use...gov.t.efficiency..1.7.	-0.2551751	0.11507881	-0.104961229
Fixed.broadband.subscriptions..per.100.people.	-0.2594710	-0.08206794	0.174826869
Individuals.using.the.Internet....of.population.	-0.2546810	-0.16925907	0.270752701
E.Participation.Index..0.1.	-0.1773347	-0.41733409	-0.377967144
Online.Service.Index..0.1.	-0.2194202	-0.45741993	0.121306565
Telecommunication.Infrastructure.Index..0.1.	-0.2711901	-0.20202290	0.236245327
E.Government.Index..0.1.	-0.2653106	-0.32112418	0.182489626
Control.of.Corruption...2.5...2.5.	-0.2773181	0.17768727	-0.008766576

Figure 9

According to the **loadings matrix** the variables which have the greatest weight within the first Principal Component are indicators related to the presence of corruption in the Public Administration such as: **Corruption Perception Index**, **Control of Corruption** and **Absence of Corruption**; therefore, it would be reasonable to label the first principal component "**Corruption**". This goes on to confirm the first expected finding. On the other hand, the original variables which have the greatest weight, i.e. the highest loadings in the second Principal Component are Online Service Index, E-Participation Index, and E-Government Index, all these indicators are closely related to the level of digitalization in the Public Administration and the interaction between the public sector/authorities and other stakeholders. Therefore, it would be appropriate to name the second Principal Component "**Digitalization**". Considering that the threshold established for the number of Principal Components to be retained is 75% of the variation, as seen by the cumulative proportion, a third variable needs to be kept. However, for the purposes of this brief analysis, which is intended to be more instrumental to the confirmation of the results of the cluster analysis rather than a stand – alone analysis, no focus will be placed on the third Principal Component. The optimal number of principal components to be retained can also be determined through a **scree plot**, which is shown in Figure 10 below.

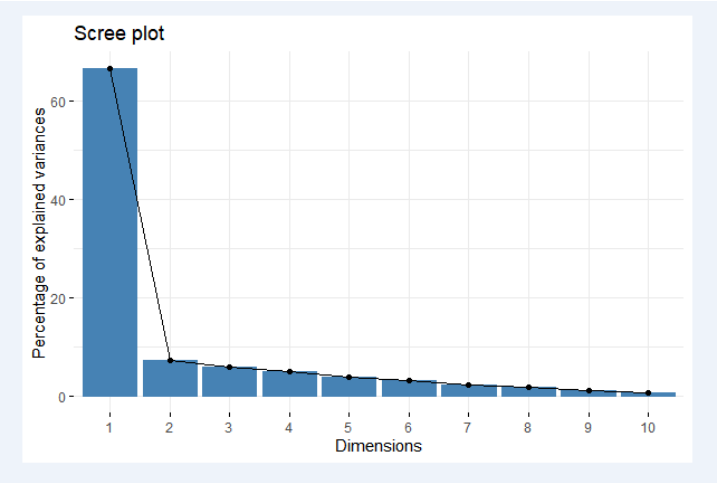

Figure 10

At this point the second expected finding can be confirmed. That is, not only do two components have a great influence in the clustering of the units of observation, but these two components are precisely the ones expected, namely the level of corruption and the level of digitalization. Therefore, the results obtained through the Principal Component Analysis are in favour of the patterns and hypotheses generated through the cluster.

It is noteworthy to assess the influence of the two Principal Components, which have been chosen, namely Digitalization and Corruption on the latent variable, i.e. Performance of the Public Administration. Analyzing the loadings matrix for the first principal component, it can be noticed that all three indicators taken into consideration have a negative coefficient. Although the author is aware of the fact that correlation does not mean causation, it would be logical to state that the presence of corruption is negatively correlated to the performance of the Public Administration. On the other hand, looking at the loadings matrix for the second principal component, it emerges that the indicators have a negative coefficient. Seeing as the second principal component was labelled "Digitalization", it would be reasonable to state that the absence of digitalization in the Public Administration would have a negative impact on the overall performance of the Public Administration, or inversely, the presence of digitalization would positively influence the latent variable.

Simple Linear Regression Analysis

The purpose of this section is to confirm the interaction between the two variables chosen and consequently the potential impact their interaction could have on the performance of Public Administrations. More specifically, the interaction expected is a negative correlation between digitalization and corruption, which would highlight the role of digitalization as a tool in the reduction of corruption in the Public Administration. The implementation of Digitalization and ICT technologies in the Public Administration is expected to have a positive effect on the latent variable through an increase of transparency, efficiency, and citizen/enterprise participation and interaction. Corruption on the other hand is known to be one of the most negatively contributing factors to good governance, rule of law and transparency and accountability as it undermines state capacity, drains precious economic resources and leads to social inequality and civil unrest. The expected positive effect of digitalization and innovation on the performance of the Public Administration will be analyzed through four **linear regression models** which aim to observe and scrutinize the relationship between digitalization and corruption. Seeing as both digitalization and corruption are complex and composite variables, four pairs of indicators have been chosen in order to illustrate the potential interaction between the two. The indicators chosen for each model are discretionally chosen by the author as deemed fit for a potential correlation. In the description of each linear regression analysis the author will make reference to the graphs containing the lines of best fit for each of the four pairs of variables chosen to test the interaction between the two variables i.e. digitalization and corruption. The author will in addition make reference to the $\mathbf{R^2}$ that is also known as the **coefficient of determination**. $\mathbf{R^2}$ is a statistical measure that represents the proportion of the variance for a **dependent variable** that can be explained by an **independent variable** or variables in a linear regression model.

Linear Regression A: In this model the author has taken into consideration the following two variables: "**Telecommunication Infrastructure Index**" (this variable is scaled from 0 which represents the lowest value to 1 which represents the highest value) and "**Public Sector Theft**"(originally this variable was scaled and 0 represented the lowest score and 1

the highest). However, seeing as the original scale was counter-intuitive for the illustrative purposes of this model, the author inverted the range. Therefore, 0 now represents lack of public sector theft, whereas 1 represents prominent presence of this phenomenon. The **line of best fit** clearly shows a negative trend between the two variables. The $\mathbf{R^2}$ value is equal to 0.62, which means that approximately 62% of the variance of the dependent variable can be explained through the independent variable. Seeing as correlation does not necessarily imply causation, the safest statement to be made is that **Telecommunication Infrastructure Index** affects **Public Sector Theft** due to the fact that a linear regression model attempts to show how a dependent variable varies as a consequence of changes in the independent variable. Therefore, countries with a higher **Telecommunication Infrastructure Index** are more likely to have a lower **Public Sector Theft**.

Linear Regression B: In this model the author has taken into consideration the following two variables: "**Absence of Corruption**" (this variable is scaled from 0 which represents the lowest value to 1 which represents the highest value) and "**E-Government Index**" (this variable is scaled from 0 which represents the lowest value to 1 which represents the highest value).The **line of best fit** clearly shows a positive trend between the two variables. The $\mathbf{R^2}$ value is equal to 0.55, which means that approximately 55% of the variance of the dependent variable can be explained through the independent variable. This linear regression model shows that although less significantly than the first case, the dependent variable is influenced by the independent variable. The most reasonable statement to be made in this case is that **"E-Government Index"** affects **"Absence of Corruption"** due to the fact that a linear regression model attempts to show how a dependent variable varies as a consequence of changes in the independent variable. Therefore, countries with a higher **"E-Government Index"** are more likely to have a higher value of **"Absence of Corruption"** and are therefore more free of the phenomenon.

Linear Regression C: In this model the author has taken into consideration the following two variables: "**ICTs impact on access to basic services**", as the independent variable, this scaled variable measures in each country, to what extent ICTs enable access for all citizens to basic services (e.g., health, education, financial services, etc.) (1 = not at all; 7 = to a great

extent)" and "**Corruption Perception Index**" as the dependent variable (this index originally uses a scale of 0-100, where 0 is highly corrupt and 100 is very clean). However, seeing as the original scale was counter-intuitive for the illustrative purposes of this model, the author inverted the range. Therefore, 0 now represents lack of corruption, whereas 100 represents a dominant presence of this phenomenon. The **line of best fit** clearly shows a strong negative trend between the two variables. The $\mathbf{R^2}$ value is equal to 0.75, which means that approximately 75% of the variance of the dependent variable can be explained through the independent variable. In this case, the most logical deduction is that the presence of ICTs has a strong impact on access to basic services offered by the Public Administration and as a consequence has an impact also on the perception of corruption. As a result, in countries with a higher availability of ICTs it is more likely that corruption will be perceived as less present and access to basic services will be positively impacted.

Linear Regression D: In this model the author has taken into consideration the following two variables: "**ICT use & gov't efficiency"** as the independent variable, (this scaled variable measures the extent to which the use of ICTs by the government improve the quality of government services to the population (1 = not at all; 7 = to a great extent) and **"Public Sector Corrupt Exchanges"** as the dependent variable (this index originally uses a scale of 0-1, where 0 indicates a high incidence of corrupt exchanges and 1 indicates a country free of this phenomenon). However, seeing as the original scale was counter-intuitive for the illustrative purposes of this model, the author inverted the range. Therefore, 0 now represents lack of corrupt exchanges in the public sector, whereas 1 represents a dominant presence of this phenomenon. The **line of best fit** clearly shows a negative trend between the two variables although not as strong as the other models. The $\mathbf{R^2}$ value is equal to 0.51, which means that approximately 51% of the variance of the dependent variable can be explained through the independent variable. In this case, the most reasonable deduction is that the use of ICTs by the government has an impact on corrupt exchanges in the public sector. Therefore, it can be stated that it is more likely for countries where the government makes extensive use of ICTs to have a lower degree of corrupt exchanges in the public sector.

The linear models that has been adopted, as was expected, have confirmed the negative association trends between the variables corruption and digitalization. The author has taken into account the parameters most representative of digitalization (to be used as independent variables) which might have the greatest influence on level of corruption and consequently the performance of Public Administrations. However, this simple linear regression approach is limited since it is difficult to predict the exact values of the dependent variable i.e. “Public sector corrupt exchanges” by taking into consideration only the unitary changes of each independent variable selected. Therefore, the author recognizes that this explorative quantitative analysis is useful for illustrative purposes and should be further developed. Through this analysis the author cautiously proceeds to confirm the hypotheses raised by the Hierarchical Cluster Analysis and supported by the Principal Component Analysis; Hence, can confirm that digitalization does have an effect on corruption, as was predicted, and could be a potential tool to improve the performance of Public Administrations by encouraging citizen/enterprise participation and promoting transparency, accountability and reducing the negative and inevitable effects deriving from a high incidence of corruption.

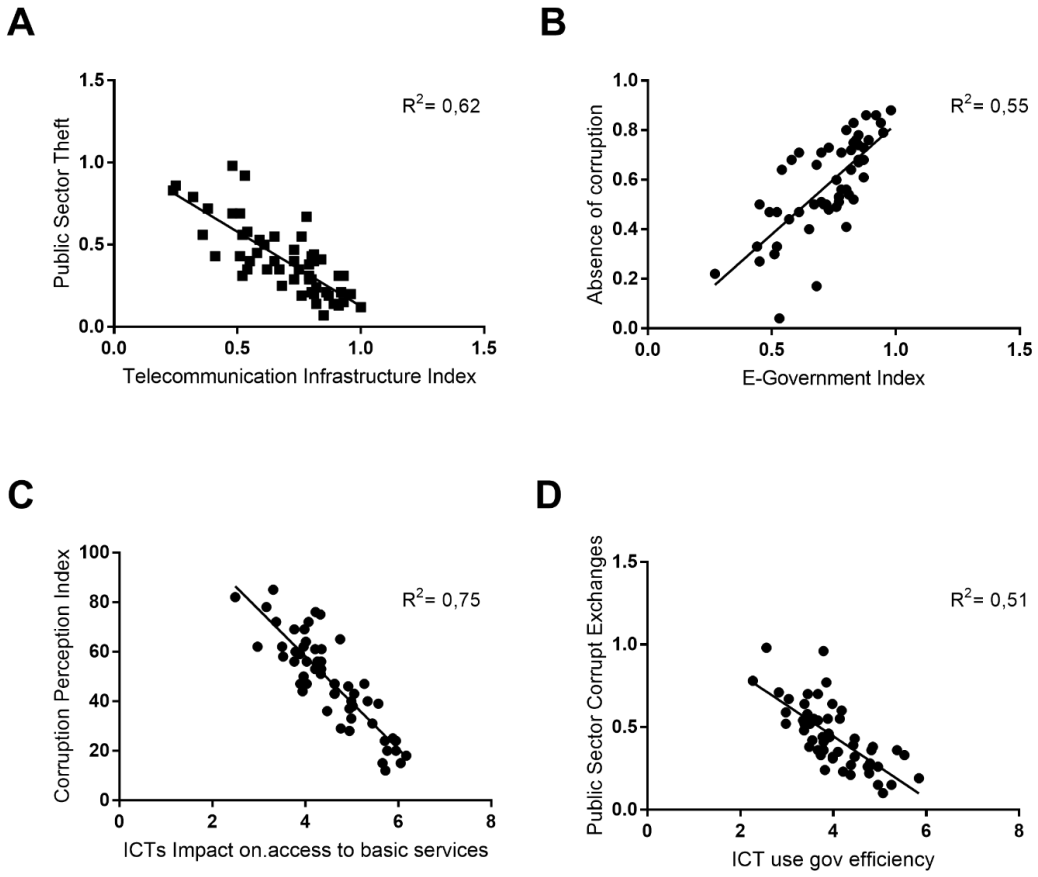
A
Public Sector Theft
Telecommunication Infrastructure Index
R2= 0,62
B
Absence of corruption
E-Government Index
R2= 0,55
C
Corruption Perception Index
ICTs Impact on.access to basic services
R2= 0,75
D
Public Sector Corrupt Exchanges
ICT use gov efficiency
R2= 0,51

Figure 11

Conclusion

This preliminary experimental study was carried out with the intention of exploring the role of digitalization as a tool in reducing corruption in the Public Administration. The author chose to take a quantitative approach and decided to adopt three statistical methods; namely, AGNES Hierarchical Cluster Analysis, Principal Component Analysis and Linear Regression Analysis to explore this relationship. Throughout the paper, the author has maintained a skeptical and cautious tone, focusing principally on analyzing the relationship between the two composite variables and raise questions regarding the viability of this potential tool rather than trying to prove a pre-established belief or enforce the idea of digitalization as a panacea against corruption. The author has refrained from expressing personal beliefs intentionally in order to explore the issue along with the reader and avoid inducing bias. Against this background, the choice of statistical analysis methods is not accidental but is intended as a step by step guide from the generation of a hypothesis through an unsupervised method such as the Cluster Analysis to the confirmation of the hypothesis through a Principal Component Analysis and Linear Regression. Notwithstanding the data – driven nature of this study, the author does not aim to provide definitive conclusions nor reliable empirical evidence to be used as an assessment for or against the use of digitalization to combat corruption in the Public Administration. The scope of this research question is too ample for the author to claim an exhaustive and thorough analysis and evaluation. Thus, what can be stated is that the aim of this paper is to apply statistical means in order to possibly highlight the potential of digitalization as a tool against corruption in the Public Administration and more importantly lay the groundwork for further more in-depth studies in the field.

Cluster Analysis: The first statistical method to be used was an AGNES Hierarchical Cluster Analysis. The output obtained from the analysis suggests that the units of observation be clustered into two distinct groups. Based on the interpretation of the data and the output, the two clusters suggested by the analysis are conceptually meaningful. As a matter of fact, the first cluster contains countries which can be considered as best-in-class performers in terms of high level of digitalization and low level of corruption. The first cluster consists of European Union member states and two LAC countries; namely Chile and Uruguay which

are current leaders in the region as regards the implementation of Digital Agendas and digitalization in the Public Administration. Instead, the second cluster contains mainly LAC countries which tend to lag behind in the implementation of Digitalization and are often characterized by high levels of corruption, frail institutional frameworks, poverty, inequality and civil dissatisfaction. Therefore, the elements contained in the second cluster are worst-in-class performers. Secondly, the output obtained from the cluster analysis through a cluster plot suggests that there is dissimilarity between the two clusters despite a slight overlapping which may be due to data limitations. Furthermore, the cluster plot suggests that there are two components which explain approximately 74% of the point variability. Through this analysis, the author detected relevant and conceptually meaningful patterns in the data, and hypothesized as had expected an interaction between the two variables corruption and digitalization. The regional element is also evident in the two clusters, but a correlation cannot be inferred as there is a high risk of running into spurious associations due to the numerous socio-economic and institutional endogenous and exogenous elements that characterize the members of the two regions.

Principal Component Analysis: Following the cluster analysis, a principal component analysis was carried out in order to support the results obtained from the previous model. Moreover, a Principal Component Analysis was conducted in order to identify the two components highlighted in the Cluster analysis and to potentially confirm the hypothesis of digitalization playing a role in the reduction of corruption in the Public Administration. The results obtained from the Principal Component Analysis went on to confirm the hypothesis raised in the precedent section. Furthermore, the Principal Component Analysis underlined the influence of digitalization and corruption on a third latent variable called "Performance of Public Administration". As was expected, digitalization positively impacts the performance of public administration, whereas corruption has a negative effect.

Linear Regression Analysis: This method was included in the study to further confirm and illustrate the results obtained from the two previous analyses. Considering the more visual nature of this model, it was used to assess the strength of the interaction, in terms of variability, between digitalization and corruption. In particular, the selection of certain

indicators which represent the aforementioned composite variables allowed for the illustration and establishment of a relationship between the two. Of course a linear model and the choice of single indicators is extremely limited and cannot provide a thorough evaluation of the interaction between the two variables, but considering the reductionist approach taken in this paper, the model provides some insightful information which goes in favour of the hypothesis that digitalization can be used as a tool against corruption in the Public Administration and allows the reader to have a visual representation of the association.

At this point, after numerous tests and confirmations, regardless of the limitations imposed by the data and methods, the author accepts the potential of digitalization as a tool in the fight against corruption in the Public Administration. The benefits of digitalization are well-known and potentially countless; however, global leaders and policy makers must take into account also the limitations and drawbacks of this instrument. Should initiatives regarding digitalization be implemented in the Public Administration and beyond, concrete and elaborate strategies must be formulated, in order to make the policy as applicable and effective as possible in each context and most importantly to overcome obstacles such as the digital divide which persists not only in terms of infrastructure but also digital literacy.

In conclusion of this study, the author proposes further research into the quantitative aspect of anti-corruption measures, one of which being digitalization. The introduction and implementation of any initiatives and tools must be made in compliance and coordination with the Sustainable Development Agenda in order to ensure uniform growth and prosperity for all.

Bibliography

Baller, S., Dutta, S. and Lanvin, B. (2016). *The Global Information Technology Report, Innovating in the Digital Economy*. World Economic Forum.

Council of Europe (1999). Criminal Law Convention on Corruption. *International Legal Materials*, 38(3), pp.505–516.

Council of Europe (2009). *Convention on Access to Official Documents*. [online] Available at: https://rm.coe.int/1680084826 [Accessed 10 Jul. 2021].

Eggers, W.D. and Bellman, J. (2015). *The journey to government's digital transformation*. [online] Deloitte Insights. Available at: https://www2.deloitte.com/za/en/insights/topics/digital-transformation/digital-transformation-in-government.html [Accessed 10 Jul. 2021].

European Commission (2017). *Report, Attitudes towards the impact of digitisation and automation on daily life*. [online] Available at: https://europa.eu/eurobarometer/surveys/detail/2160 [Accessed 18 Jul. 2021].

European Commission (2021). *Shaping Europe's Digital Future, The Digital Europe Program*. [online] Available at: https://europa.eu/eurobarometer/surveys/detail/2160 [Accessed 18 Jul. 2021].

Eurostat, Data Browser (2021). *Individuals using the internet for interaction with public authorities*. [online] Europa.eu. Available at: https://ec.europa.eu/eurostat/databrowser/view/tin00012/default/table?lang=en [Accessed 5 Jul. 2021].

Gazzetta Ufficiale della Repubblica Italiana (1990). *Legge 7 agosto 1990, n. 241*. [online] Available at: https://www.gazzettaufficiale.it/eli/id/1990/08/18/090G0294/sg.

Gazzetta Ufficiale della Repubblica Italiana (2013). *Decreto Legislativo 14 marzo 2013, n. 33*. [online] Available at: https://www.gazzettaufficiale.it/eli/id/2013/04/05/13G00076/sg.

Hartigan, J.A. (1975). *Clustering algorithms*. New York, N.Y.: Wiley.

International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database (2019a). *Fixed broadband subscriptions (per 100 people)*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/IT.NET.BBND.P2 [Accessed 3 Jul. 2021].

International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database (2019b). *Individuals using the Internet (% of population)*. [online] Worldbank.org. Available at: https://data.worldbank.org/indicator/IT.NET.USER.ZS [Accessed 27 Jun. 2021].

Ministro per la Pubblica Amministrazione (2016). *Dlgs 25 maggio 2016, n. 97 - FOIA e Trasparenza*. [online] . Available at: http://www.funzionepubblica.gov.it/articolo/ministro/12-02-2016/trasparenza [Accessed 16 Jul. 2021].

OECD (2009). *Convention on Combating Bribery of Foreign Public Officials in International Business Transactions - OECD*. [online] . Available at: https://www.oecd.org/corruption/oecdantibriberyconvention.htm [Accessed 18 Jul. 2021].

Pacheco, J., Santiso, C., Gaviria Halaby, L., Tejerina, L. and Moschini, S. (2020). *The Dialogue, Can Latin America Find a Faster Path to Digital Government?* [online] . Available at: https://www.thedialogue.org/analysis/can-latin-america-find-a-faster-path-to-digital-government/ [Accessed 9 Jul. 2021].

Transparency International (2020). *Corruption Perceptions Index 2020*. [online] Transparency.org. Available at: https://www.transparency.org/en/cpi/2020/index/nzl [Accessed 1 Jul. 2021].

Transparency International (2021). *What is corruption?* [online] Transparency.org. Available at: https://www.transparency.org/en/what-is-corruption [Accessed 16 Jul. 2021].

Tufis, C.D. (2019). *The Global State of Democracy Indices Codebook*. 3rd ed. Stockholm: International IDEA.

UNDP (2019). *UNDP Digital Strategy*. [online] UNDP. Available at: https://digitalstrategy.undp.org/ [Accessed 16 Jul. 2021].

United Nations, Department of Economics and Public Affairs, Division for Public Institutions and Digital Government (2020a). *Data Center*. [online] Un.org. Available at: https://publicadministration.un.org/egovkb/en-us/Data-Center [Accessed 2 Jul. 2021].

United Nations, Department of Economics and Public Affairs, Division for Public Institutions and Digital Government (2020b). *E-Government Development Index*. [online] Un.org. Available at: https://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government-Development-Index [Accessed 5 Jul. 2021].

United Nations, Department of Economics and Social Affairs, Statistics Division (2021). *The sustainable development goals report*. [online] Available at: https://unstats.un.org/sdgs/report/2021/ [Accessed 15 Jul. 2021].

United Nations. Office On Drugs And Crime (2004). *United Nations Convention Against Corruption*. Vienna: United Nation, Office On Drugs And Crime.

UNODC,The Doha Declaration, Promoting a Culture of Lawfulness (n.d.). *Anti-Corruption Module 1 Key Issues: Corruption - Baseline Definition*. [online] www.unodc.org. Available at: https://www.unodc.org/e4j/en/anti-corruption/module-1/key-issues/corruption---baseline-definition.html [Accessed 11 Jul. 2021].

World Bank, Enterprise Surveys (2020a). *Bribery incidence (% of firms experiencing at least one bribe payment request)*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/IC.FRM.BRIB.ZS [Accessed 28 Jun. 2021].

World Bank, Enterprise Surveys (2020b). *Informal payments to public officials (% of firms)*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/IC.FRM.CORR.ZS [Accessed 28 Jun. 2021].

World Bank Group (2021). *Worldwide Governance Indicators*. [online] Worldbank.org. Available at: https://databank.worldbank.org/source/worldwide-governance-indicators/preview/on [Accessed 5 Jul. 2021].

World Bank national accounts data, and OECD National Accounts data files (2020). *GDP (current US$) - European Union, Latin America & Caribbean | Data*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=EU-ZJ&view=map [Accessed 1 Jul. 2021].

World Bank, Quality of Government (2019). *GovData360: Political corruption index*. [online] GovData360. Available at: https://govdata360.worldbank.org/indicators/h7b6f8df1?country=BRA&indicator=44172&viz=line_chart&years=1946 [Accessed 2 Jul. 2021].