UNIVERSITY
INSTITUTE
OF LISBON

Public Procurement and Human Resources Data Analysis in Healthcare

Susana Filipa Raposo Ferreira

Master in Computer Engineering

Supervisor:
PhD *Luís Miguel Martins Nunes*, Associate Professor,
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD *David Alexandre Correia Ferraz*, Associate Professor,
Iscte - Instituto Universitário de Lisboa

September, 2025

[This page is intentionally left blank.]

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Public Procurement and Human Resources Data Analysis in Healthcare

Susana Filipa Raposo Ferreira

Master in Computer Engineering

Supervisor:
PhD *Luís Miguel Martins Nunes*, Associate Professor,
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD *David Alexandre Correia Ferraz*, Associate Professor,
Iscte - Instituto Universitário de Lisboa

September, 2025

[This page is intentionally left blank.]

Dedico este trabalho à minha filha Leonor.

Que seja também um exemplo de dedicação e perseverança.

I dedicate this master to my daughter Leonor.

May this work also stand as an example of dedication and perseverance.

[This page is intentionally left blank.]

Acknowledgments

I wish to express my gratitude to my supervisor and co-supervisor for their attentive guidance, scientific orientation, and constant encouragement throughout this journey.

To my daughter Leonor, I am thankful for her support and understanding, even in the moments when my dedication to this work required my absence. To my parents, Afonso e Eduarda, who have always devoted themselves unconditionally to their children and supported me in both the best and the most difficult moments of my life, I owe special thanks. To my siblings, Pedro e Ana, as well as to my uncles, aunts, cousins, and relatives, I also extend my sincere gratitude. To my nephews Afonso e Constança, and to my sister-in-law Raquel, I offer my affection and recognition. I warmly remember my grandparents Adriano and Maria Dinoria, and my cousin Daniel, whose memory remains ever present.

I am equally grateful to the professors and to my master's colleagues – Paula, Hugo, Rodrigos, Afonso, Bruno, Igor, Nuno, Rosário, Tomás, Gonçalo, Ofélia, Renata, and Inês – for sharing experiences and companionship. My thanks also go to the professors and colleagues from the Mathematics undergraduate program at Coimbra.

To my long-standing friends from Coimbra – Catarina, Marta, João, Caldeira, Guada, Lilas, Maria, Caniche, Laranja, Leite, Anas, Peixoto, Susana, Baila, Ivan, Inês, André, Raquel, Miguel, Luciana, Caçador, Maria João, Correia, and Sofia – I express my heartfelt thanks for a friendship that has endured over time. To my friends from Malveira e Venda do Pinheiro – Alcina, Cristina, Ana, Maria João, João e Carapinha – I also extend my gratitude.

To my colleagues from the various schools where I have taught, and in particular to André, Jorge, Ricardos, Nunos, Luís, Antonieta, Ângelo, Abílio, João, and the rest of my colleagues from AR, I am thankful for their friendship and collaboration.

Finally, but by no means least, to Pedro Teixeira, for showing me how truly beautiful life can be when shared with the right person by our side.

[This page is intentionally left blank.]

Agradecimentos

Desejo expressar a minha gratidão ao meu orientador e coorientador, pelo acompanhamento atento, pela orientação científica e pelo constante incentivo ao longo deste percurso.

À minha filha Leonor, agradeço o apoio e compreensão, mesmo nos momentos em que a minha dedicação implicou a minha ausência. Aos meus pais, Afonso e Eduarda, que sempre se dedicaram incondicionalmente aos filhos e me apoiaram nos melhores e piores momentos da minha vida, deixo um agradecimento especial. Aos meus irmãos Pedro e Ana, bem como aos tios, tias, primos e primas, manifesto também a minha sincera gratidão. Aos meus sobrinhos Afonso e Constança, e cunhada Raquel, o meu carinho e reconhecimento. Recordo com emoção os meus avós Adriano e Maria Dinoria e o meu primo Daniel, cuja memória permanece sempre presente.

Agradeço igualmente aos professores e colegas de mestrado – Paula, Hugo, Rodrigos, Afonso, Bruno, Igor, Nuno, Rosário, Tomás, Gonçalo, Ofélia, Renata e Inês – pela partilha de experiências e pelo companheirismo. Aos professores, e ainda aos colegas da licenciatura em Matemática de Coimbra.

Aos amigos de longa data de Coimbra – Catarina, Marta, João, Caldeira, Guada, Lilas, Maria, Caniche, Laranja, Leite, Anas, Peixoto, Susana, Baila, Ivan, Inês, André, Raquel, Miguel, Luciana, Caçador, Maria João, Correia, e Sofia -, deixo o meu sincero obrigada pela amizade que perdura no tempo. Aos meus amigos da Malveira e Venda do Pinheiro – Alcina, Cristina, Ana, Maria João, João e Carapinha -, manifesto também a minha gratidão.

Aos colegas de trabalho das várias escolas onde lecionei, e em particular aos colegas André, Jorge, Ricardos, Nunos, Luís, Antonieta, Ângelo, Abílio, João e restantes colegas da AR, agradeço pela amizade e colaboração.

Por último, mas não menos importante, ao Pedro Teixeira, pelo exemplo de como a vida pode ser verdadeiramente bonita quando partilhada com a pessoa certa ao nosso lado.

[This page is intentionally left blank.]

Resumo

Esta dissertação investiga a correlação entre contratação pública e os recursos humanos no setor da saúde em Portugal, no período de 2015 e 2024, com foco nos serviços médicos e de enfermagem. O objetivo central é avaliar em que medida os padrões de aquisição pública influenciam a evolução do número de profissionais de saúde na administração pública.

São aplicadas técnicas de ciência de dados e três fontes principais: contratos públicos extraídos do Portal Base, estatísticas de recursos humanos da DGAEP e dados legislativos do Parlamento português. O processo analítico inclui extração, preparação e limpeza dos dados, seguido de análise exploratória e visualização, com recursos a normalização por base, escalas logarítmicas e gráficos com eixo duplo.

Os resultados revelam variações significativas nos volumes de contratação ao longo do período analisado, com picos associados ao COVID-19 e a alterações nas políticas públicas. Verifica-se uma associação entre o investimento em contratação pública e a evolução dos recursos humanos, com padrões distintos entre médicos e enfermeiros.

Conclui-se que a contratação pública pode constituir um indicador complementar para a gestão estratégica de pessoal no setor da saúde, sendo o modelo de estudo proposto replicável em contextos institucionais semelhantes.

O contributo metodológico – em particular o enquadramento para extrações de dados em larga escala, agregação e tratamento – oferece um modelo replicável para futuras investigações que envolvam conjuntos de dados públicos complexos e fragmentados.

Palavras-chave: Portugal; contratação pública; inteligência artificial

[This page is intentionally left blank.]

Abstract

This dissertation investigates the correlation between public procurement and human resources in the healthcare sector in Portugal, covering the period from 2015 to 2024, with a particular focus on medical and nursing services. The central objective is to assess the extent to which public procurement patterns influence the evolution of the number of healthcare professionals in the public administration.

Data science techniques are applied to three main sources: public contracts extracted from the Portal Base, human resources statistics from Directorate-General for Administration and Public Procurement (DGAEP), and legislative data from the Portuguese Parliament. The analytical process includes data extraction, preparation, and cleaning, followed by exploratory analysis and visualization, employing baseline normalization, logarithmic scales, and dual-axis charts.

The results reveal significant variations in procurement volumes throughout the analysed period, with peaks associated with the COVID-19 pandemic and shifts in public policy. A correlation is observed between investment in public procurement and the evolution of human resources, with distinct patterns between physicians and nurses.

The dissertation concludes that public procurement may serve as a complementary indicator for strategic workforce planning in the healthcare sector, and that the proposed study model is replicable in similar institutional contexts.

Moreover, the methodological contribution – particularly the large-scale data extractions, aggregation, and cleaning framework – offer a replicable model for future research involving complex and fragmented public datasets.

Keywords: Portugal; public procurement; artificial intelligence

[This page is intentionally left blank.]

Funding and Publications

This work was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2023 and UIDP/04466/2023].

[This page is intentionally left blank.]

Contents

List of Figures

[This page is intentionally left blank.]

List of Tables

[This page is intentionally left blank.]

List of Acronyms

AI	Artificial Intelligence
ANCP	National Agency for Public Procurement
ANN	Artificial Neural Networks
API	Application Programming Interface
BE	Left Bloc (Bloco de Esquerda)
CDS-PP	Social Democratic Centre – Popular Party (Centro Democrático Social – Partido Popular)
CH	Enough (Chega)
CNN	Convolutional Neural Networks
CPV	Common Procurement Vocabulary
CSV	Comma-Separated Values: plain text format used to store tabular data
ESSPP (ESPAP)	Entity for Shared Services of Public Procurement
EU	European Union
FNN	Feed-forward Neural Networks
DGAEP	Directorate-General for Administration and Public Employment
DRE	Electronic Republic Gazette (Diário da República Eletrónico)
GDP (PIB)	Gross Domestic Product (Produto Interno Bruto)
ICT	Information and Communication Technologies
IL	Liberal Initiative (Iniciativa Liberal)
INCM	National Printing Office (Imprensa Nacional-Casa da Moeda)
HFL (LBS)	Health Framework Law (Lei de Bases da Saúde)
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
JSON	Javascript Object Notation
JPP	Together for the People (Juntos pelo Povo)

L	Free (Livre)
Latin-1	ISO/IEC 8859-1, also known as Latin Alphabet No. 1
LSTM	Long Short-term Memory
MAPE	Mean Absolute Percentage Error
NHS (SNS)	National Health Service (Serviço Nacional de Saúde)
NLP	Natural Language Processing
NSI (INE)	National Statistics Institute (Instituto Nacional de Estatística)
NPPS	National Public Purchasing System
OECD	Organization for Economic Co-operation and Development
PAN	People-Animals-Nature Party (Pessoas-Animais-Natureza)
PCC	Public Contracts Code
PCP	Portuguese Communist Party (Partido Comunista Português)
PEV	Ecologist Party "The Greens" (Partido Ecologista "Os Verdes")
PPD/PSD	People's Democratic Party / Social Democratic Party (Partido Popular Democrático / Partido Social Democrata)
PS	Socialist Party (Partido Socialista)
PTIN	Portuguese Tax Identification Number
R&D	Research and Development
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
TED	Tender Electronic Daily
UMIC	Agency for the Knowledge Society
UTF-8	Unicode Transformation Format – 8-bit
WHO	World Health Organization

CHAPTER 1

Introduction

1.1. Background and Motivation

In Portugal, public procurement has been increasing, representing 5,73% of the Gross Domestic Product (GDP) in 2023. According to the 2023 annual public procurement report in Portugal, available on Portal Base, there was a 23% increase compared to 2022. Regarding contract amounts, there was an overall increase of 20,8% (approximately 2.622 million euros compared to 2022) (Instituto dos Mercados Públicos - Portal Base, 2024).

It plays a crucial role in sectors such as healthcare, education, and infrastructure. The application of Artificial Intelligence (AI) in public procurement has become an increasingly relevant field, enabling the identification of behavioural patterns and is a crucial mechanism through which the government implements its policies and an effective tool for managing public funds (Riyad & Laila, 2024).

In an era characterized by geopolitical tensions and challenges such as the pandemic, energy concerns, and high inflation rates, the public sector assumes several important roles: ensuring public health, promoting the economy, providing national security and defence; and responding adequately to the public interest and delivering quality services (Butkus et al., 2023).

The impact of COVID-19 pandemic on procurement trends will be examined, due to the World Health Organization's (WHO) declaration, in March 2020, considering SARSV-CoV-2 as a global pandemic. This prolonged and worldwide public health crisis lead to a concern for ensuring the continuity of public services (Padeiro et al., 2021).

This study will also consider events and factors that can provide a better understanding of the context. For example, municipalities tend to issue more procurement contracts in the months leading up to elections. However, this does not always result in increased expenditure. Additionally, several key events occurred between 2013 and 2015 that can help clarify the circumstances surrounding the aforementioned transition. Portugal exited the Troika bailout program in 2014, and a new municipal finance law, which took effect on January 1, 2014, imposed stricter limits on municipalities' ability to incur debt (Curado et al., 2021).

Public procurement in Portugal is regulated by the Public Contracts Code (PCC) (Ministério das Obras Públicas, 2008), which establishes distinct legislation and procurement methods compared to other countries, based on the budget. The most important are (Portuguese Republic, 2025):

- Direct Award – a contracting authority invites a chosen supplier to submit a proposal directly.
- Prior Consultation – the authority invites at least three entities to submit proposals and

may negotiate contract aspects.

- Open Procedure – a competitive process publicly announced in the Official Gazette and the European Union (EU) Official Journal.
- Restricted Procedure - a two-stage competitive process published in official journals when contract values exceed EU thresholds, involving candidate qualification and bid evaluation.
- Negotiate Procedure – similar to the restricted procedure but allows selected candidates to refine proposals through negotiation.
- Competitive Dialogue – used when the authority knows its needs but not the solution. Includes qualification, solution presentation, and dialogue phases before proposal submission. No bid negotiation occurs.
- Partnership for Innovation – supports Research and Development (R&D) innovate products, services, or works for potential future acquisition, provided they meet agreed performance and cost levels.

The main objective is to understand the entire process of a public procurement procedure, complying with the applicable legislation and considering potential political and social events that may directly influence behavioural patterns (Curado et al., 2021). In addition to these factors, there are errors made by both the contracting authority and the awarded entity (União Europeia, 2015).

There are also other factors that may influence these behaviours, provided by the data available on Portal Base, such as: the type of procedure (open tendering, restricted tendering, request for proposals, among others); the country district and municipality of execution; and the Common Procurement Vocabulary (CPV), a code that represents the type of service contracted (Instituto dos Mercados Públicos, 2025).

The objective of the study is also based on investigating the correlation between public procurement and human resources in the healthcare sector in Portugal, between 2015 and 2024, with a focus on medical and nursing services.

1.2. Research Questions

This research seeks to answer the next two questions:

1. What behavioural patterns can be detected in public procurement, within the public health sector, from 2015 to 2024, particularly in response to the COVID-19 pandemic?
2. How is public procurement in the healthcare sector related to the creation of permanent staff positions?

1.3. Document Structure

The remainder of this dissertation is organized as follows:

- Chapter 2: Literature Review

  This chapter outlines the research foundation through a structured review of existing literature. It explains the methodology used to conduct the review, the keywords that guided the search, and the criteria for including or excluding sources.

  It also presents the research planning process, introduces key background concepts, and discusses related works and projects that inform the study.

- Chapter 3: Data Science

  This chapter presents the data-driven component of the research. It details the processes of data extraction, preparation, and cleaning, followed by a perspective overview of the dataset. It includes a focused study using the CPV vocabulary, explores patterns within the data, and analyses the relationship between public procurement and public sector employment using visualization techniques such as dual Y-axis plots, baseline normalization, and logarithmic scaling.

CHAPTER 2

Literature Review

2.1. Methodology

To ensure a comprehensive approach, this study undertakes a systematic literature review, involving the identification, selection, and synthesis of scientific research on the applications of artificial intelligence in public procurement. In the development and refinement of the research methodology, Mendeley Reference Manager plays a crucial role by organizing the collected documents, facilitating efficient retrieval, and allowing the highlighting and annotation of the most relevant text fragments. This structured approach not only supports transparency and reproducibility but also ensures that the research is grounded in a thorough understanding of prior studies, providing a solid foundation for subsequent data analysis and methodological innovations[1].

2.2. Keywords

Since the objective is centred on investigating the correlation between public procurement and human resources in the healthcare sector, with the focus on medical and nursing services; with the aim of selecting keywords that would yield search results related to the thesis topic, two separate searches were initially conducted:

- Health and “Public Procurement” and “Artificial Intelligence” (961 results).
- Portugal and “Public Procurement” and “Artificial Intelligence” (141 results).

Considering that the applicable legislation and political conditions may influence public procedures in general (Curado et al., 2021), the final keywords chosen were: Portugal and “Public Procurement” and “Artificial Intelligence”.

2.3. Exclusion Criteria

Of the 141 results given by Scopus, the following criteria were used for exclusion (Figure 1):

- Documents published before 2015.
- Not written in Portuguese and English.
- Only with the subject area: computer science or Engineering or Economics, Econometrics and Finance or Health Professions.
- That are in the final publication stage (not article in press).

[1] Elsevier. (2013). Search | Mendeley. https://www.mendeley.com/search/

2.4. Inclusion Criteria

Of the 89 results given by Scopus, the following criteria were used for inclusion Table 1:

- Documents directly related to the public procurement process and not to other topics, such as energy consumption estimation, blockchain, or transparency.

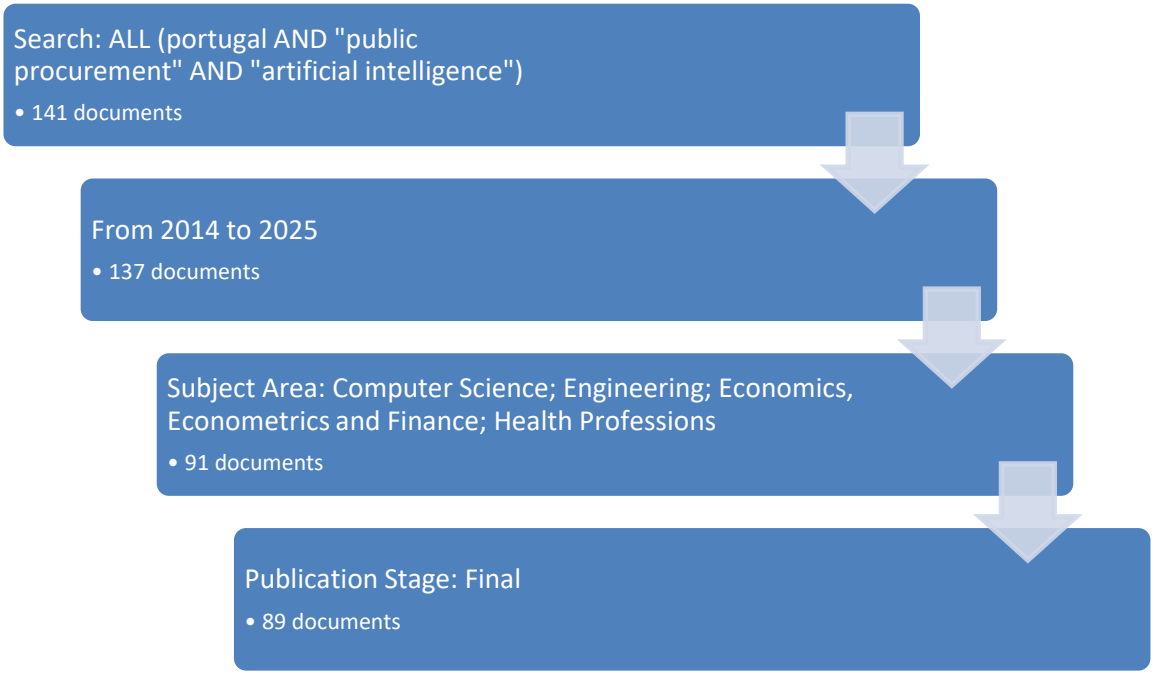


Figure 1 - Exclusion Criteria

Table 1 - Search Results and Inclusion Criteria

Database	Documents	Classification of Documents	
Scopus	89	Not Relevant	65
		Moderately Relevant	11
		Relevant	9
		Based on the abstract, it could be relevant, but it is a resource with restricted access	4

The list of the nine relevant documents is presented:

- Noce et al. (2024). *A Generic Architecture for the Digitization of Government Procurement Processes*. S.A C., G. K, E.P.
- Jacques de Sousa, Poças Martins, Sanhudo, et al. (2024). *Automation of text document classification in the budgeting phase of the Construction process: a Systematic Literature*

Review. Construction Innovation.

- Butkus et al. (2023). *Measuring quality perception of public services: customer-oriented approach*. Engineering Management in Production and Services.
- Carneiro et al. (2020). *Network Analysis for Fraud Detection in Portuguese Public Procurement*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).
- Da Rosa et al. (2015). *Portuguese public procurement governance model*. Associação Portuguesa de Sistemas de Informação, Atas da Conferencia da Associação Portuguesa de Sistemas de Informação.
- Jacques de Sousa, Poças Martins, & Sanhudo (2024). *Predicting construction project compliance with machine learning model: case study using Portuguese procurement data*. Engineering, Construction and Architectural Management.
- Ferreira & Amaral (2016). *Public e-procurement: Advantages, limitations and technological "pitfalls"*. Association for Computing Machinery, ACM International Conference Proceeding Series.
- Jacques de Sousa, Simões, et al. (2023). *Statistical Descriptive Analysis of Portuguese Public Procurement Data from 2015 to 2022*. CivilEng Journal.
- Jacques de Sousa, Poças Martins, et al. (2023). *Tackling the Data Sourcing Problem in Construction Procurement Using File-Scraping Algorithms*. Engineering Proceedings Journal.

2.5. Scival

SciVal is a research intelligence platform, leveraging data from scopus, that enables users to visualize research performance, benchmark against peers, develop collaborative partnerships, and analyse research trends. The platform comprises six modules: overview, benchmarking, collaboration, trends, grants and impact[2]. This tool was used with the aim of following a more organised and structured approach during the documentation search and the study of the state of art.

This tool has been used for exploring and identifying references since 2003. Accordingly, the search **"public procurement in Portugal with artificial intelligence"** was conducted[3], yielding the following summary:

[2] Elsevier. (2025a). SciVal | Research performance assessment solution | Elsevier. https://www.elsevier.com/products/scival?dgcid=RN_AGCM_Sourced_300005487

[3] Elsevier. (2025b, February 7). Scopus - Scopus AI. https://www.scopus.com/pages/ai?query=public%20procurement%20in%20portugal%20with%20artifical%20inteligence&isExample=false

"To understand the role of AI in public procurement in Portugal, we can explore several key areas where AI is being applied and its potential benefits and challenges."

As examples of AI applications in public procedures, five areas were provided: "auditing and supervision," "regulation and legal framework", "efficiency and anti-corruption", "human oversight", "legal and ethical issues" and "adoption and trust". Of these five topics, the following is relevant to this study: "auditing and supervision", as it addresses AI helping in data analysis, pattern detection, and improving audit efficiency.

Of the nine bibliographic references supporting these topics, two were recommended for applying AI in data analysis and detecting behavioural patterns. This focused and refined research method has proven useful for the state-of-the-art study. In addition to these bibliographic references, SciVal presents further questions for deeper exploration, such as: "How is AI being utilized in public procurement processes in Portugal?" and "What are the potential impacts of integrating AI into public procurement in Portugal?".

A concept map was created to help to organize, represent, and visualize information in a structured way, making it easier to understand and learn this topic (Figure 2).

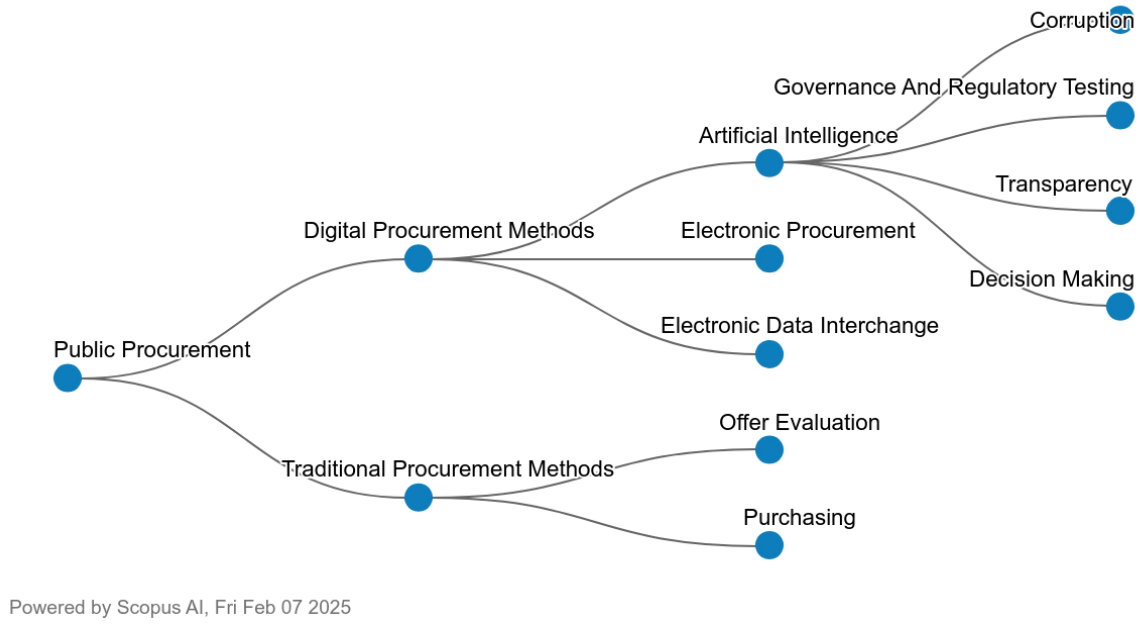


Figure 2 - Concept Map from SciVal - Scopus AI

The nine most relevant documents to the study were analysed using the online tool Voyant Tools[4] to enhance the literature review, computer-assisted text analysis tools were used to identify patterns and relationships among key concepts within the selected documents. The use of charts generated by

[4] Sinclair, S. R. G. V. C. (2025). Voyant Tools. https://voyant-tools.org/

Voyant Tools proved particularly useful for detecting terminological correlations and understanding the frequency and distribution of relevant terms across the analysed texts. These visual results complement the traditional bibliographic review by providing a quantitative and graphical perspective on the content of the studies.

Among the twenty most frequently used terms across the documents, the following were excluded from the list: https, al, et, doi.org and 2022. Strong relationships were observed between the terms “public” and “procurement”, as well as various dependencies associated with them, such as “sector”, “quality”, “production” and “portuguese” (Figure 24). These charts are dynamic and can be updated and used throughout the study by adding terms or documents (see Appendix A).

Another useful chart for finding information related to specific terms is the one shown in

Figure 25, where it is possible to identify where the required terms are used in each document.

Figure 26 allows to identify whether the terms are used in the final part of the document, which typically consists of pages dedicated to bibliographic references (see Appendix A).

2.6. Background Concepts

Public procurement is a fundamental pillar of strategic governance and service delivery by governments. It plays a key role in ensuring public sector efficiency and strengthening citizens' trust due to the significant volume of public spending it represents. Well-designed public procurement systems contribute to achieving critical policy objectives, such as environmental protection, innovation, job creation, and the growth of small and medium-sized enterprises (Da Rosa et al., 2015).

Public contracts serve as a key tool in public management, promoting public value by balancing economic efficiency with social and environmental objectives. Framed within a strategic management approach, public procurement policies aim to enhance socio-economic conditions while ensuring effective governance. A major challenge for authorities is monitoring and evaluating procurement outcomes, a task increasingly supported by Information and Communication Technologies (ICT). E-procurement has become a crucial e-government initiative, significantly impacting economic development. Portugal has made notable progress in this area, earning recognition in the European

context. However, e-procurement platforms have limitations, posing challenges in assessing public procurement's impact on value creation. Further efforts are needed to enhance transparency, accountability, and the effective use of ICT in governance (Ferreira & Amaral, 2016).

As a key tool for implementing public policy, public procurement creates public value by balancing economic efficiency with socio-economic development. A major governance challenge lies in monitoring and evaluating its impact, a process increasingly supported by information and communication technologies**.** E-procurement has emerged as a crucial e-government initiative, with Portugal recognized as a leading example in the EU due to its mandatory adoption of electronic public procurement platforms (Ferreira & Amaral, 2016).

Public procurement in Europe accounts for 19% of GDP. The 2014 European Directives emphasize the importance of governance measures in Member States. Portugal, a pioneer in electronic public procurement, is aligning with Europe's guidelines and focusing on establishing a formal governance model. Such a model is essential for improving integrity in public procurement, supporting sustainable growth, and strengthening citizens' trust. Over the past 15 years, Portugal's public procurement system has undergone significant changes, influenced by EU policies, with the creation of public entities and a reorganization of services within the National Public Purchasing System (NPPS) (Da Rosa et al., 2015).

The subject of public procurement is governed by various instruments of international and EU laws, as well as, in Portugal, the PCC (Ministério das Obras Públicas, 2008). Essentially, a public contract is an agreement formed through two or more expressions of will – one from a public contracting authority and another from a contractor – aimed at producing specific legal effects, such as the acquisition of goods or services and public works contracts (Carneiro et al., 2020).

Agency for the Knowledge Society (UMIC), National Agency for Public Procurement (ANCP), and Entity for Shared Services of Public Procurement (ESPAP) are entities involved in the evolution and development of Portugal's NPPS. Their roles and responsibilities have evolved over time, shaping the governance and structure of public procurement in the country. UMIC laid the groundwork and set the foundation of electronic procurement, NAPP formalized centralized purchasing, and ESPAP now manages and oversees the system (Da Rosa et al., 2015).

The latest EU directives on public procurement, adopted in 2014, required Member States to transpose them into national law by April 2016. These directives—Directive 2014/23/EU[5], Directive

[5] European Parliament. (2014a, February 26). Directive - 2014/23 - EN - EUR-Lex. https://eur-lex.europa.eu/eli/dir/2014/23/oj/eng

2014/24/EU[6], and Directive 2014/25/EU[7]—aimed to modernize and simplify procurement procedures while ensuring transparency, competition, and efficiency in public contracting. Portugal transposed these directives through Decree-Law No. 111-B/2017 (Ministério do Planeamento e das Infraestruturas, 2017), which amended the PCC to align with the new European legal framework. This marked the first time that Member States were subject to a comprehensive set of obligations within the scope of governance (Da Rosa et al., 2015).

2.7. Related Works / Projects

The Portal Base provides various types of information, including: the information and execution of public contracts covered by Part II of the PCC, excluding details on the performance of concession contracts, in accordance with Decree-Order No. 85/2013 of 27 February 2013, witch amends Decree-Order No. 701-F/2008 of 29 July 2008; notices regarding the initiation of procurement procedures and any subsequent notices, if applicable; orders and decisions establishing procurement priorities under exceptional regimes, specifically those outlined in Decree-Law No. 34/2009 of 6 February 2009; contracts concluded through direct award procedures (general regime), open procedures, restricted procedures, negotiation procedures, and competitive dialogue procedures; amendments to contracts that exceed 15% of the original contract value; and, accessory sanctions imposed under Article 460 of the PCC (Carneiro et al., 2020).

The adoption of government information systems has been a strategy to improve administrative efficiency, transparency, and the quality of public services. Among these systems, e-procurement (electronic public procurement) stands out for its central role, as it involves a significant share of public budgets — around 27% of government expenditure in Organisation for Economic Co-operation and Development (OECD) countries. E-procurement refers to the use of digital technologies to automate and optimize the processes of acquiring goods and services, promoting principles such as efficiency, economy, fairness, and sustainability (Noce et al., 2024).

In a statistical study conducted by the University of Porto, using data from 2015 to 2022, the contracts were organized into six thematic groups: publication and closing year, region, deadlines, award criteria, price variables, and performance**.** When the award criteria were based on multiple factors rather than just price, this feature was grouped into categories according to the weight of the price factor within the multi-criteria evaluation. Similarly, both price and performance were also grouped into categories (Jacques de Sousa, Simões, et al., 2023).

[6] European Parliament. (2014b, February 26). Directive - 2014/24 - EN - EUR-Lex. https://eur-lex.europa.eu/eli/dir/2014/24/oj/eng

[7] European Parliament. (2014c, February 26). Directive - 2014/25 - EN - EUR-Lex. https://eur-lex.europa.eu/eli/dir/2014/25/oj/eng

Each contract identified involves two parties: the contracting authority and the contracted entity. Various other properties are obtained from Base Portal, including the contract date, contract value, contract purpose, publication date, execution location, and competing entities (Carneiro et al., 2020).

A related study on fraud detection in Portuguese public procurement suggests that the highest layer of the architecture is the decision-support layer. The primary objective of this layer is to assist auditors and other users, such as those preparing framework agreements or other contracts in compliance with Portuguese public procurement law, in their decision-making process. The system can therefore be used in a proactive and preventive manner, enabling entities currently renegotiating new contracts to verify their legal validity. This feature is particularly beneficial for framework agreements, which involve multiple entities (Carneiro et al., 2020).

Based on the publications and contract completions per year from 2009 to 2022, the decline in the number of published contracts from 2020 onward is primarily due to the exclusion of ongoing contracts that were still in progress at the time of data aggregation. Additionally, the impact of the COVID-19 pandemic may have contributed to this decrease, affecting procurement processes and contract publication (Jacques de Sousa, Simões, et al., 2023).

A descriptive statistical analysis of 5.172 public procurement contracts closed between 2015 and 2017 provided key insights into the performance of construction projects in Portugal. The study revealed that overall project performance during this period was positive, though several trends and challenges emerged: award criteria showed little correlation with final project prices; multifactor assessment criteria did not necessarily result in better performance compared to projects awarded solely based on price; high-value projects (above 1.1 million United States Dollars - approximately 1.066.890 €) awarded solely based on price tended to perform worse than those evaluated using multiple factors; contract execution was predominantly concentrated in coastal areas and major cities; the number of contracts submitted to public procurement has generally increased; extended submission deadlines did not guarantee improved financial performance in construction projects; errors and omissions in construction reporting were common, leading to information gaps in public tender repositories; public procurement data repositories plays a crucial role in summarizing and providing insights into project outcomes; error mitigation tools and enhanced awareness of data submission accuracy are essential for improving public procurement databases. These findings highlight the importance of improving procurement processes, refining award criteria, and enhancing data accuracy to optimize public contract performance and decision-making (Jacques de Sousa, Simões, et al., 2023).

A study conducted by the University of Porto analysed the applicability of AI algorithms Feed-forward Neural Networks (FNN), multilayer perceptron, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-term Memory (LSTM), and Radial Basis Function

(RBF)) in the budgeting phase of the construction process**.** The data preparation followed several stages: data preparation and pre-processing, feature selection, algorithm selection and training, model testing, and model evaluation**.** The study used Python, MATLAB, R, and SPARQL as programming languages and employed AI libraries/packages such as TensorFlow, PyTorch, and Keras**,** with the latter specializing in Artificial Neural Networks (ANN)**.** Additionally**,** SpaCy, NLTK, and FastTextAI were utilized as Natural Language Processing (NLP)-support repositories (Jacques de Sousa, Poças Martins, Sanhudo, et al., 2024).

There are two main approaches to implementing AI in construction project budgeting: ANN, primarily LSTM models, to predict key budgeting variables, and NLP to categorize project specifications and improve budget accuracy. Researchers typically obtained data from private companies and government institutions, then developed and tested algorithms, with Python and SpaCy being the most used tools. Highway construction was the most frequent project type analysed. Performance evaluation relied on metrics such as F1-Score, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), with CNN achieving up to 90% accuracy when combined with other techniques. However, no single approach consistently outperformed others, as results depended heavily on data quality. One major limitation was the lack of complete, high-quality datasets, leading researchers to focus on specific project types or tasks rather than generalizing solutions (Jacques de Sousa, Poças Martins, Sanhudo, et al., 2024).

The study highlights key research gaps, including the need to generalize AI models across various construction tasks, establish standardized contract formatting, and develop larger open-source datasets. Continuous learning models could help improve prediction accuracy over time. A major implication for future ANN and NLP applications in the architecture, engineering, and construction sector is the necessity of better data accessibility before AI tool development. To fully integrate AI in construction, a cultural shift in data-sharing practices is required, along with reduced reliance on human verification and enhanced algorithm applicability across different construction tasks. While ANNs have been extensively studied for cost prediction, NLP applications remain underexplored, signalling a need for further research in this area (Jacques de Sousa, Poças Martins, Sanhudo, et al., 2024).

CHAPTER 3

Data Science

3.1. Data Extraction

The data utilized in this research were extracted from the Portal Base (Instituto dos Mercados Públicos, 2025), the official Portuguese government platform for public procurement information. This site serves as the central repository for all public procurement activities in Portugal and plays a key role in promoting transparency and accountability in the management of public funds. The platform offers various categories of data, including public contracts, procurement notices, institutional profiles, contract amendments, acquisitions of movable goods, cancellations, and legal disputes. For the purpose of this thesis, only the "contracts" category was selected, as it provides structured and detailed records essential for analysing patterns, procedures, and values in public procurement activities over time.

Given the size and complexity of the database, and since there was no Application Programming Interface (API) available for direct data access, the extraction of contract data across the ten years from 2015 to 2024 was carefully planned and executed using a robust methodological framework (Figure 3). It is important to note that the portal allows advanced search filtering based on three distinct dates: the contract date, the publication date, and the closure date. For the purposes of this search, the contract date was used as the primary temporal reference, as it is the most accurate time reference, and it reflects the moment of legal and financial commitment between the contracting authority and the supplier. This choice ensures that the resulting dataset is aligned with the actual timeline of public expenditure, rather than administrative publication delays or procedural closures that may occur significantly later.

All Python code and Jupyter Lab notes are available on the public GitHub repository https://github.com/sfrfa/ThesisAppendices/tree/main.

The extraction began with defining the scope of the data. The period from January 2015 to December 2024 was selected to provide a comprehensive longitudinal perspective on public contracting patterns. The Portal Base offers access to a range of procurement-related information, including contracts, procurement notices, contracting entities, contract modifications, acquisitions of movable goods, cancellations of contract awards, and legal disputes or challenges. However, given that the objective of this study is to examine human resources contracting in the healthcare sector, only entries classified as formal contracts were considered, as they include structured metadata regarding contracting authorities, contractors, contract values, procedural types, and timelines, making them suitable for quantitative and comparative analysis.

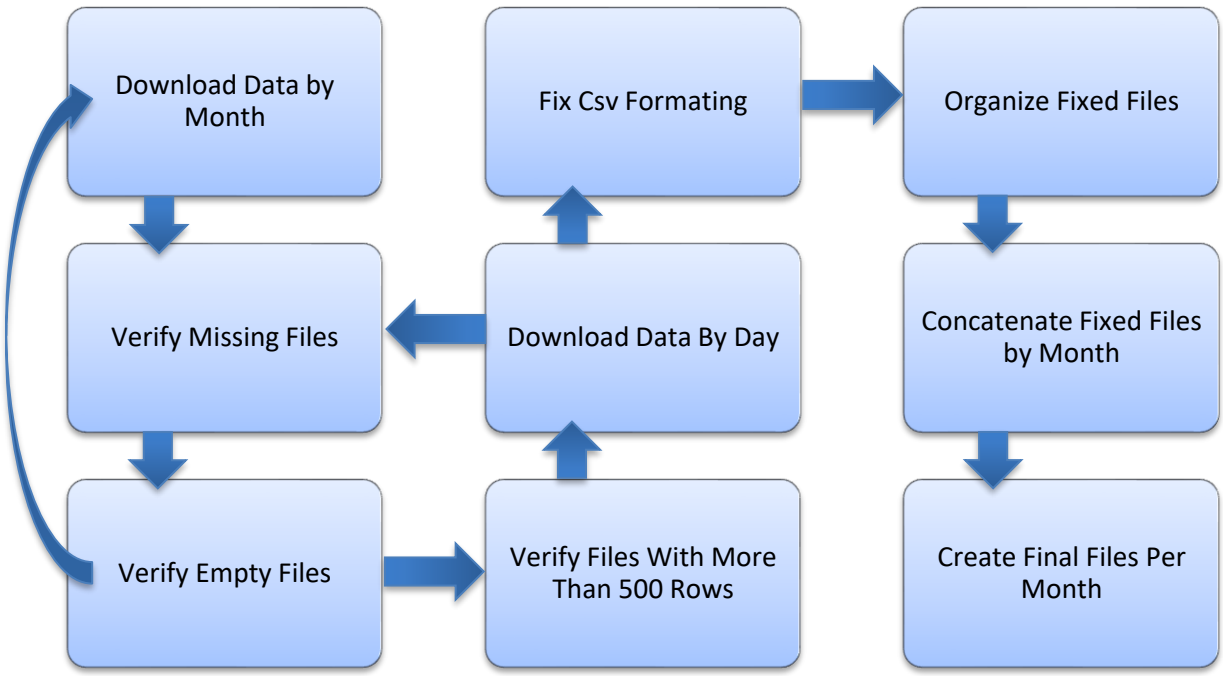


Figure 3 - Diagram for Downloading Monthly Data

One of the central technical challenges was the Portal Bases's limitation on data export: a maximum of 500 contracts can be downloaded per search. This restriction required a comprehensive strategy to ensure that all relevant contracts were retrieved without exceeding this limit.

Although the portal includes numerous contract procedures, this research extracted every available type of public contracting process. These included open tenders, direct awards, simplified procedures, prior consultations, negotiated procedures, competitive dialogues, innovation partnerships, frameworks agreements, and procedures conducted under exceptional legal regimes. Even less common or recently introduced procurement procedures – such as those implemented under emergency legislative measures – were included to ensure coverage of the full legal and procedural landscape.

To extract data within the download constraints, filtering techniques were employed. The search interface of the Portal Base allows users to filter results not only by procedure type and contract date, but also by geographic criteria such as country, district, and municipality. The data collection strategy involved systematically applying these filters in succession. For each year and month in the designated time frame, queries were executed separately for each contract type. For higher-density regions or periods, municipality-level filters were applied to reduce the number of results per query. If the number of results still exceeded the maximum allowable threshold of 500, the searches were further broken by day, effectively limiting each query to 24-hour periods. This approach ensured that no record was omitted due to the portal's technical limitations.

Once the initial download was completed, a verification script checked for missing files – those

that had not been downloaded successfully due to connectivity issues – and re-initiated the download for any missing data. Another script then scanned the dataset for empty files and identified those that contained only column headers without any contract records. These files were typically the result of portal behaviour where searches yielding no results still generated a Comma-Separated Values (CSV) file with only the header row. Such files were downloaded again.

In cases where the portal returned an error during manual re-download – typically because the original information was not in the right format, a script generated series of individual daily queries was used to identify which contract was in error.

Given the iterative nature of this filtering process, the data extractions involved the generation of a large number of individual queries. Each query returned a file, typically in CSV format – though in this case, the fields were separated by semicolons-containing a subset of the contracts. The resulting dataset consisted of millions of files, each corresponding to a unique combination of procedural type, time interval, and geographic unit.

Conceptually, the data extraction process can be visualized as a multi-layered funnel in which the broad dataset hosted on the Portal Base is filtered through successive stages of refinement. Initially, the entire database is restricted to the contracts category. From there, the data is sliced by contract type, then segmented by year and month, further divided by geographic area, and finally, when needed, broken down by individual days. At each stage, the objective was to reduce the number of records returned by each query to a manageable size while ensuring that no relevant data was excluded. The result of this process was a fully comprehensive and structured dataset of public contracts awarded in Portugal over a ten-year period.

This workflow represents a significant methodological commitment to data completeness and replicability. It also reflects the real-world limitations of publicly accessible governmental data platforms, which are often not designed for large-scale academic research. By developing a rigorous extraction and organization pipeline, this study not only builds a robust empirical foundation for analysis but also contributes to the methodological literature on public procurement data collection, offering a replicable model for future researchers working with constrained public datasets.

3.2. Data Preparation

Once the raw files were acquired, a structured data organization system was implemented. All files were categorized into a hierarchical filename system based on year, month, procedure type, and geographic area. File naming conventions were standardized to embed metadata such as the contract date, location, and procedure type directly into the filenames, which facilitated downstream processing and traceability. The files were then parsed using data manipulation tools capable of

handling large volumes of tabular data, such as Python with Pandas library. This step included concatenating files into a unified dataset, checking for duplicates, standardizing column formats, and identifying any inconsistencies or gaps in the data.

Following the data extraction phase, a comprehensive data preparation and cleaning process was undertaken to transform the raw output into a structured, analysable dataset. Given the scale of the project – comprising millions of individual CSV files resulting from filtered downloads across time periods, locations, and contract types – this step was crucial for ensuring the integrity, consistency, and usability of the data. The raw files, although machine-generated, were not immediately ready for analysis due to various formatting inconsistencies and the fragmented nature of the download structure imposed by the portal's technical limitations.

The first task in the cleaning workflow involved correcting the CSV (Figure 4). One of the first issues encountered was inconsistency in the formatting of the raw CSV files, they were delimited by semicolons which is a common regional variant. Additionally, a subset of the files contained encoding errors, malformed characters, or partial rows, often due to variations in encoding between Unicode Transformation Format – 8 bit (UTF-8) and Latin-1 (ISO/IEC 8859-1, also known as Latin Alphabet No. 1). To address these problems, a dedicated Python script was developed to automate the cleaning of files by reading the CSV using Pandas, initially attempting UTF-8 decoding and falling back to Latin-1 when necessary.

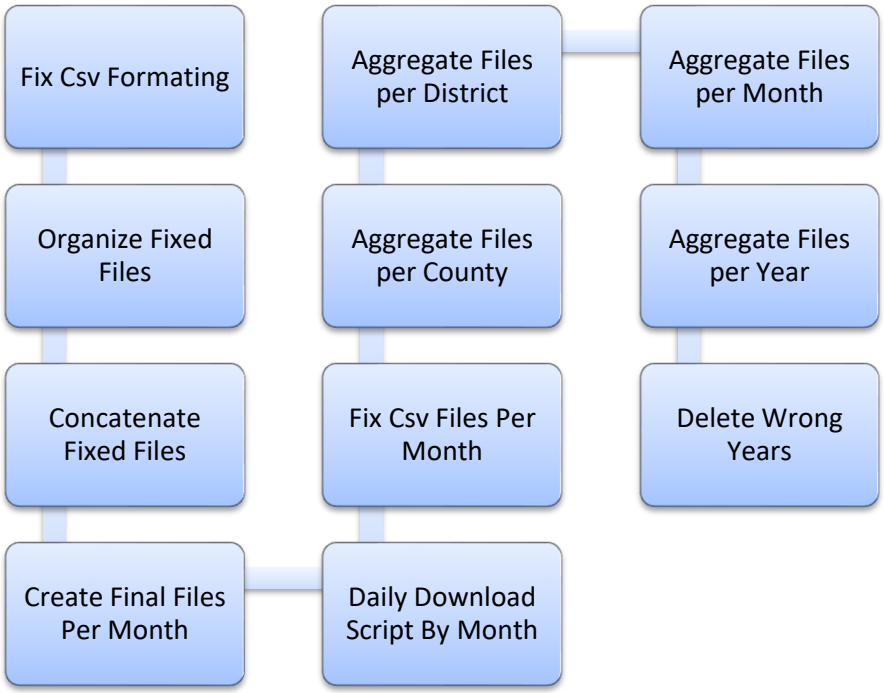


Figure 4 - Diagram for Data Preparation

The script processed files stored in a designated source folder and saved the corrected versions in a separate output directory. Logging was implemented to track each operation, capturing both

successful files and files that raised decoding expressions. This logging was essential for transparency, error tracking, and reproducibility. Importantly, the script also avoided overwriting already corrected files, ensuring that previously cleaned data would not be redundantly processed or inadvertently modified.

Once the CSV files were successfully cleaned and re-encoded in UTF-8, they were reorganized into a consistent structure that mirrored their original filtering dimensions – by county, and municipality district. A second set of scripts then concatenated individual files into aggregated monthly datasets. During this process, a new variable was introduced to each record, indicating the specific filter used during data extraction (e.g., the contract type or geographical level). This ensured that the contextual metadata from the original query was retained in the final dataset, supporting traceability and future disaggregation.

Subsequent stages of the cleaning process focused on structured aggregation. The fixed monthly files were further merged to generate broader analytical dataset: one per type, one per county, one per district, and one per month. This multi-level structure enabled both special and temporal analysis, speed and ease of handling. At the end of this process, all data was aggregated by year, producing one master file for each year between 2015 and 2024. These annual files included a comprehensive set of variables describing the contract value, contracting authority, supplier, procedural method, geographic location, and relevant dates.

To facilitate the automation of monthly queries on the Portal Base, the data extraction process was initially designed to fetch contract records from the first day of each month to the first day of the following month. This approach simplified the coding logic by maintaining uniform start and end points for all time ranges. However, an important peculiarity of the portal's behaviour became apparent: results from the first day of the year were incorrectly included in the dataset to the preceding year (as well as in the correct year). To resolve this overlap and ensure temporal accuracy, a post-processing script was implemented that filtered each file to retain only those records whose contract date strictly belonged to the intended year.

3.3. Data Description

Table 2 provides a detailed description of the dataset fields, including their name, a brief description of the contents, data type, the unit of measurement, a representative example and the null or empty percentage. In addition to the original variables, several derived fields were created throughout the study to support the analysis. These derived features were based on transformations or groupings of the original variables. For example, the procedure's base price was categorized into predefined intervals aligned with the official thresholds for public procurement procedures: 'Up to 5.000 €', '5.001

€ to 50.000 €', '50.001 € to 139.000 €', '139.001 € to 5.350.000 €', and 'Above 5.350.000 €'. Such derived fields enabled a more structured and meaningful interpretation of the data during the exploration and analysis phase.

Table 2 - Description of the Raw Data Fields

Description	**Description**	**Data Type**	**Unit**	**Example(s)**	**Null or empty percentage**
Contract ID	Contract Identification Number	Integer	Not applicable	26533	Null: 0% Empty: 0%
Announcement Number	Unique identifier automatically assigned to each public procurement procedure that has been published in the official platforms (e.g., Electronic Public Gazette (DRE), Portal Base, or the Tender Electronic Daily (TED)).	String	Not applicable	20210/2024	Null: 0% Empty: 88,42%
Type of Announcement	Nature or purpose of the public notice published by the contracting authority. Also indicates what stage or update of the procedure is being made public.	String	Not applicable	Procedure announcement Notice of extension of deadline Urgent tender announcement	Null: 0% Empty: 88,41%
INCM ID	Identifier assigned by the National Printing Office (INCM)	Integer	Not applicable	415808522	Null: 0% Empty: 88,41%
Type of Contract	Category or nature of the contract	String	Not applicable	Purchase of movable goods Hiring services	Null: 0% Empty: 0%
Procedure ID	Procedure Identification Number	Integer	Not applicable	6818690	Null: 0% Empty: 0%

Type of Procedure	Method or legal process used to award the public contract	String	Not applicable	Direct award Prior Consulting Public tender	Null: 0% Empty: 0%
Contract Object	Indicates the specific goods, services, or work that the contract covers	String	Not applicable	Aquisição de Serviços Médicos APROV-PRT04/2016/ARSN	Null: 0% Empty: 0%
Contract Description	Short description of the object of the contract	String	Not applicable	Purchase of medicines Various urology materials	Null: 0% Empty: 0%
Contracting Authority	Public organization or entity responsible for issuing and managing the contract. This field has the fiscal number and the entity name.	List of String	Listof(<fiscal number> - <entity designation>)	[501442600 - Institute of Employment and Vocational Training, I. P.]	Null: 0% Empty: 0%
Awardees / Contractors	Company or individual that wins the contract and is responsible for delivering the goods, services, or works agreed upon with the contracting authority	List of String	Listof(<fiscal number> - <entity designation>)	[508080975 - Ricoh Portugal Unipessoal, Lda]	Null: 0,12% Empty: 0%
Publication Date	Official date when the contract notice or award information was made publicly available	Date	dd/mm/yyyy	30/05/2024	Null: 0% Empty: 0%
Contract Signing Date	Date of contract signing	Date	dd/mm/yyyy	22/07/2024	Null: 0% Empty: 2,53%
Contract Value	Financial worth of the contract as established at the time of the contract	Double	€	142632.54	Null: 0% Empty: 0%

CPV	CPV code and description	List of string	Listof(<cpv code> - <cpv designation>)	[33140000-3 - Material médico de consumo]	Null: 0% Empty: 0%
Execution Period	Timeframe during which the contractor must deliver the goods, complete the works, or provide the services specified in the contract	Integer	Days	365	Null: 0% Empty: 0%
Place of Execution	Geographical place of contract performance	List of string	Listof(<local designation>)	[Portugal, Vila Real, Chaves]	Null: 2,45% Empty: 0%
Justification	Legal/technical basis for choosing a specific procurement procedure or award decision	String	Not applicable	Artigo 259.º do Código dos Contratos Públicos	Null: 0% Empty: 3,29%
Centralized Procedure	Indicated whether the contract was awarded through a central purchasing body or joint procurement system, rather than by the contracting authority alone	Boolean	Not applicable	Yes/No	Null: 0% Empty: 0%
Framework Agreement Number	Unique identifier of the framework agreement under which the contract was awarded	String	Not applicable	5032541	Null: 88,18% Empty: 0%
Framework Agreement Description	Short summary of the subject and scope of the framework agreement	String	Not applicable	CP 2019/61 - Medicamentos diversos	Null: 88,18% Empty: 0%
Base Price of the Procedure	Estimated budget set by the contracting authority	Double	€	266982.65	Null: 0% Empty: 0%

Award Decision Date	Date on which the contracting authority officially decided to award the contract to a specific contractor	Date	dd/mm/yyyy	13/09/2024	Null: 0% Empty: 2,71%
Contract Closing Date	Date on which the contract is officially competed or terminated	Date	dd/mm/yyyy	24/09/2015	Null: 0% Empty: 65,73%
Effective Total Price	Actual amount paid under the contract including adjustments, taxes, or additional costs	Double	€	9799.53	Null: 0% Empty: 0%
Regime	Legal framework or set of rules governing the contracts	String	Not applicable	Public Contracts Code (Decree-Law 18/2008)	Null: 0% Empty: 0%
Justification for not Reducing to Written Contract	Why a formal written contract was not required	List of string	Not applicable	[Article 11(1) – Procurement excluded in special sectors]	Null: 69,53% Empty: 0%
Type of Contract Termination	How or why the contract was ended before or at the end of its execution	String	Not applicable	Contract Termination / Resolution	Null: 0% Empty: 66,67%
Material Criteria	Indicates whether specific material or substantive criteria were used to evaluate bids	Boolean	Not applicable	Yes/No	Null: 0% Empty: 0%
Competitors	Companies or entities that participated in the procurement process but did not win the contract	List of String	Listof(<fiscal number> - <entity designation>)	[508080975 - Ricoh Portugal Unipessoal, Lda]	Null: 60,97% Empty: 0%
Link to Procedure Documents	URL or web link to the official documents	String	Not applicable	https://community.vortal.biz/Public/contract-notice-	Null: 0% Empty: 73,89%

				view/PT1.NTC.3073745/	
Observations	Additional notes, comments, or relevant information	String	Not applicable	Assumption of Multiannual Commitments – INF 137359/2024/L-EFSA	Null: 0% Empty: 93,38%
Ecological Contract	Indicated whether the contract incorporates environmental or sustainability criteria in its execution or procurement process	Boolean	Not applicable	Yes/No	Null: 0% Empty: 0%
Year	Year associated to the contract closing date	Date	yyyy	2023	Null: 0% Empty: 0%
Justification for Direct Award	Reason why the contract award was directly to a supplier without a competitive tender	String	Not applicable	Lack of own resources	Null: 0% Empty: 62,12%

3.4. Data Cleaning

The files were saved in Parquet format because it offers efficient data compression and encoding, enabling faster read and write operations compared to text-based formats like Javascript Object Notation (JSON) or CSV. This is especially beneficial for handling large datasets, as Parquet's columnar storage structure allows for optimized querying and reduced storage footprint. Moreover, it is widely supported in data processing frameworks such as Pandas, Spark, and SQL engines, making it a practical choice for scalable analytics workflows[8].

The data cleaning process was carried out prior to applying any filtering based on CPV codes, with the objective of facilitating the detection of inconsistencies and ensuring high data quality. The initial dataset comprised 1.575.991 public procurement procedures registered between 2015 and 2024, including a wide range of contract types and attributes. A detailed schema of the dataset was produced to understand its structure, and each feature was systematically validated, including the assessment of missing and empty values – both in absolute terms and as percentages.

[8] Apache Parquet. (2025, January 17). Overview | Parquet. https://parquet.apache.org/docs/overview/

During the cleaning stage, features with generic object data type were converted to more appropriate types such as datetime for date fields and float for monetary values, improving consistency and enabling more efficient downstream processing. Additionally, new features were created from logical transformations and pattern extraction, such as splitting the Portuguese Tax Identification Number (PTIN) and entity name, which were originally stored together in the same field for both awarding and awarded entities. This separation enabled more accurate entity-level analysis and easier cross-referencing with external datasets, such as CPV catalogues or fiscal registries.

During the data cleaning phase, several issues were identified, such as columns with high percentage of missing or empty values (e.g. nAnuncio with 88.42%, TipoAnuncio with 88.41%, idINCM with 88.41%, dataFechoContrato with 65,73%, and PrecoTotalEfetivo with 1.017.852 zero values); as well as inconsistent values, such as in the prazoExecucao (execution period) field, which, although expressed in days, contains extreme values such as 509.948.286 days (equivalent to 1.397.118 years) and 365.744.304 days.

3.5. Common Procurement Vocabulary Study

The CPV establishes a single classification system for public contracts, aiming to standardize how contracting authorities describe the subject of their procurements and consists of: a main vocabulary, used to define the object of a contract; and, a supplementary vocabulary, which adds qualitative details. It uses a tree-structured numeric code of up to 9 digits: 8 digits describe the item or service; and, the 9th digit is a check digit for validation[9].

The structure includes: first 2 digits – divisions (e.g., XX000000-Y), first 3 digits – groups (XXX00000-Y), first 4 digits – classes (XXXX0000-Y), first 5 digits – categories (XXXXX000-Y). The last 3 digits (of the 8) offer increasing detail within each category.

To analyse public procurement data and identify contracts related to human resources in the health sector, the CPV codes were used as primary classification system. CPV codes offer a standardized way to describe the subject of public contracts across the EU. The CPV codes used in this analysis were extracted from two main sources: Portal Base (Instituto dos Mercados Públicos, 2025) and TED[10], the online supplement to the Official Journal of the EU, where EU- -wide public procurement notices are published.

The analysis of CPV code was conducted using data primarily from Portal Base. While both portal base and TED provide CPV information, the data from TED includes multiple columns with CPV

[9] European Union. (2025). Main and Supplementary Vocabulary of CPV - ted.europa.eu - TED. https://ted.europa.eu/en/simap/cpv

[10] European Union. (2025). Main and Supplementary Vocabulary of CPV - ted.europa.eu - TED. https://ted.europa.eu/en/simap/cpv

description in all official EU languages and would complicate text processing and introduce redundancy when identifying relevant procurement categories. To ensure consistency, clarity, and efficient data handling, only Portuguese-language CPV data from portal base was used for the classification process.

The analysis was carried out in Jupyter Lab, following a structured approach based on the hierarchical nature of CPV codes: divisions, used to broadly filter contract areas related to health and social services; groups, helped isolate services possibly involving human resource components; classes, narrowed the focus to labour-related services, including staffing and recruitment; categories, provided precise identification of CPV codes directly associated with human resources in the healthcare sector, such as the hiring of medical staff or outsourcing of healthcare professionals.

Of the 30.434 CPV codes analysed, 222 were selected. From the 222 selected cases, regardless of the type of contract (e.g., service acquisition, public service concessions, others), there are 112 distinct final CPVs. The Jupyter Lab code and the final complete list of selected codes can be consulted in the appendix. The CPV list is organized by divisions, groups, classes and categories in Table 2 (see Appendix B).

In order to strengthen the selection of the appropriate CPV codes, a dictionary of terms was created and used to search for CPVs based on the contract object and awarding entity. It was found that 101 codes appeared in both lists, 11 only in the manually selected list, and 4.240 exclusively in the list obtained through the term dictionary search. The 11 codes present solely in the manual list were retained, as they refer to specific medical practices not included in the dictionary, such as homeopathy, urology, or medical imaging. The 4.240 codes identified exclusively through the dictionary search mostly relate to procedures involving public health entities but not directly linked to healthcare provision – for example, facility maintenance or furniture acquisition.

The final list of CPVs is organized into five main divisions, as presented in Table 3. Based on an analysis of the descriptions of groups, classes, and categories within each of these divisions, as well as a reviews of the contract object descriptions for selected procedures, it was concluded that the CPV codes corresponding to procedures involving the procurement of external services in the healthcare sector are the following: 79624000-4 (Supply services of nursing personnel) and 79625000-1 (Supply services of medical personnel).

Table 3 - Final CPV Divisions

CPV started with	**Description**	**Examples**
73	Research and development services and related consultancy services	Experimental development and research services Services related to research laboratories

75	Services related to public administrations, defence, and social security	Emergency and fire services Community services provision
79	Business services – legal, marketing, consultancy, recruitment, printing, and security	Nursing staff services Medical staff supply services
80	Education and training services	Higher education services Youth training services Medical education services
85	Health and social work services	Hospital surgical services Hospital gynaecology services

3.6. Exploration and Analysis of the Data

Figure 5 presents the annual distribution of public contracts awarded to private entities for the provision of medical and nursing within the public sector in Portugal, spanning the period from 2015 to 2024. Overall, the distribution exhibits a non-linear trend, with noticeable fluctuations across the years. The number of contracts peaked in 2016, with a total of 76 contracts, followed by another significant rise in 2024, which recorded 57 contracts. In contrast, the lowest point occurred in 2020, with only 14 contracts identified.

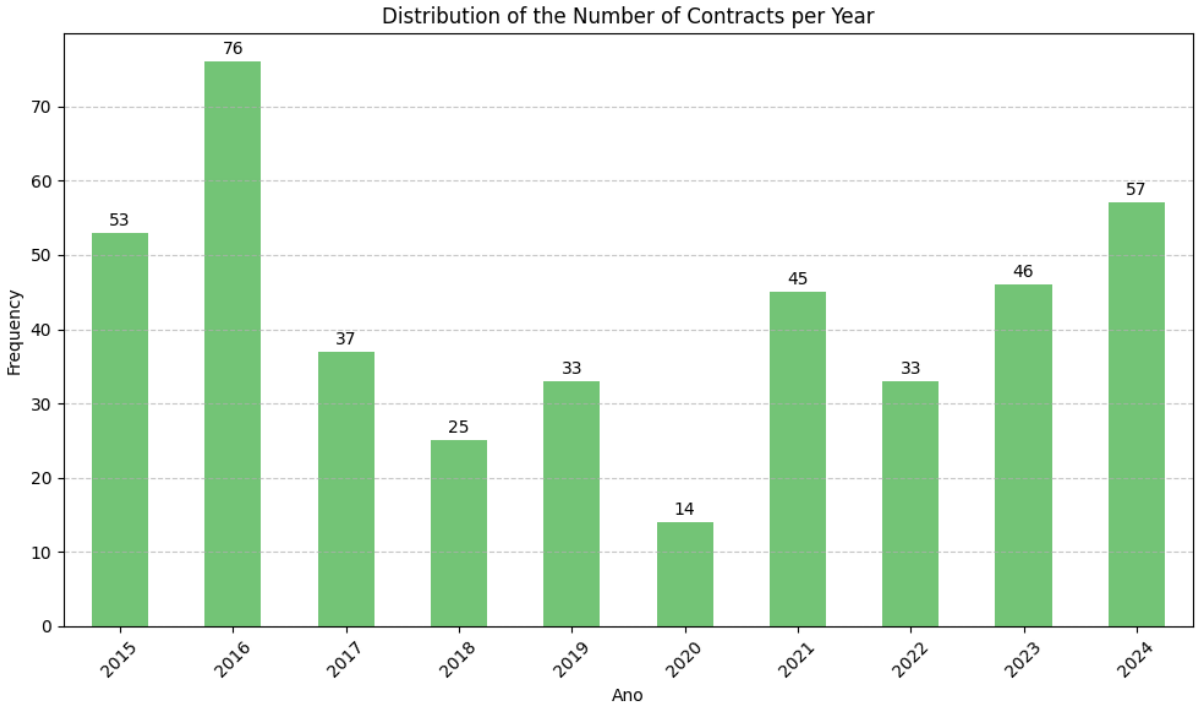


Figure 5 - Chart - Distribution of the Number of Contracts per Year

Several factors may explain the observed fluctuations; however, to complete the data, it is necessary to check not only the number of contracts but also the annual amount spent. In the data provided by the Portal Base, there are three data points with values: the contractual price, the base price of the procedure, and the effective total price. The base price of the procedure refers to the maximum amount set by the contracting authority as the upper spending limit during the tender planning phase, serving as a benchmark for bid evaluation. The contractual price is the amount established at the time of award and formalized in the contract, reflecting the most economically advantageous tender and constituting the legally binding commitment between the parties. The effective total price, in turn, represents the final amount paid after contract execution, including any revisions, amendments, or adjustments, thus reflecting the actual cost borne by the Public Administration.

In the graphs of Figure 6 and Figure 7, the values are organized by contractual ranges (these ranges are based on the values of each type of procedure). A significant number of procedures with an effective total price and base procedure price of zero are observed.

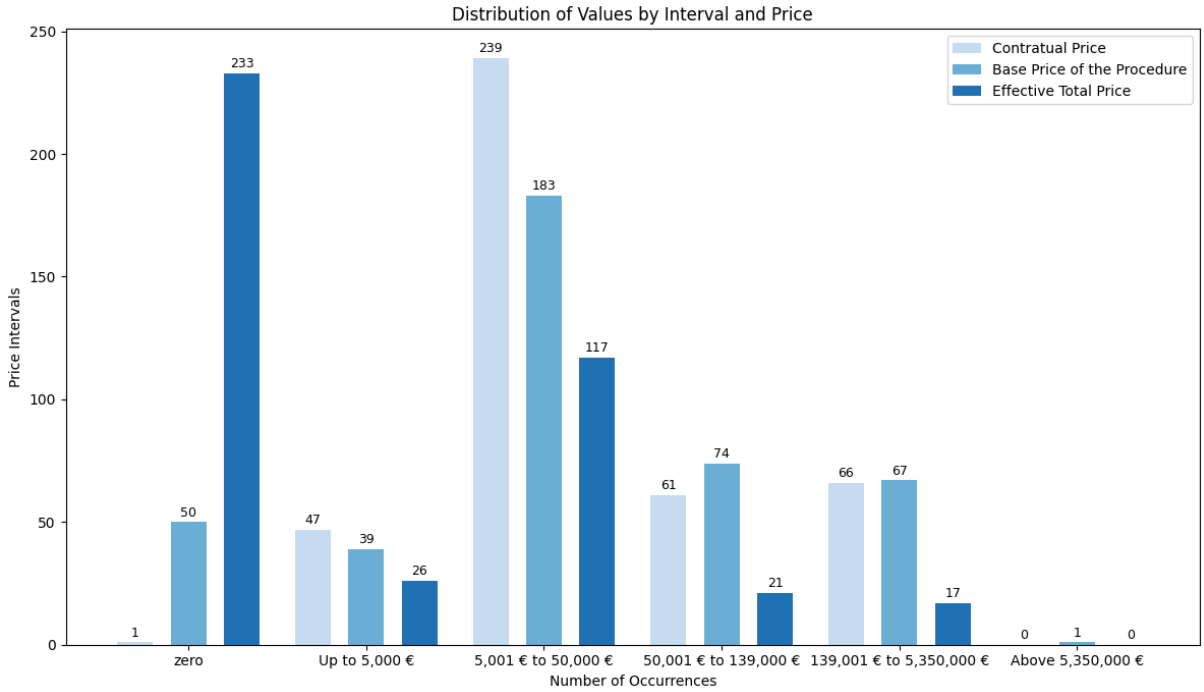


Figure 6 - Distribution of Values by Interval and Price

Analysing the sum of values per year (Figure 8), we find that the three columns show disparate data. The correlation between the three price columns and the other numerical features in the dataset is similar (Figure 9), which indicates that their relevance and impact in relation to those columns are similar.

The Base Portal was created in 2008, in the context of the implementation of the PCC, approved by Decree-Law No. 18/2008 of January 29, which came into force on July 30, 2008 (Ministério das

Obras Públicas, 2008). Several ordinances have defined the model for interactive data entry in the Portal Base, notably Ordinance No. 701-E/2008 of July 29 (Ministérios das Finanças e da Administração Pública e das Obras Públicas, 2008), No. 57/2018 of February 26 (Ministério das Finanças e Planeamento e das Infraestruturas, 2018), and No. 318-B/2023 of October 25 (Ministério das Finanças, 2023). As shown in Table 4, not all data fields were mandatory between 2015 and 2024. Moreover, the three price-related fields are filled in at different stages of the procurement process, which may have contributed to the incorrect entry of 0 € in these fields.

Thus, the contractual price value will be used for the remainder of the study.

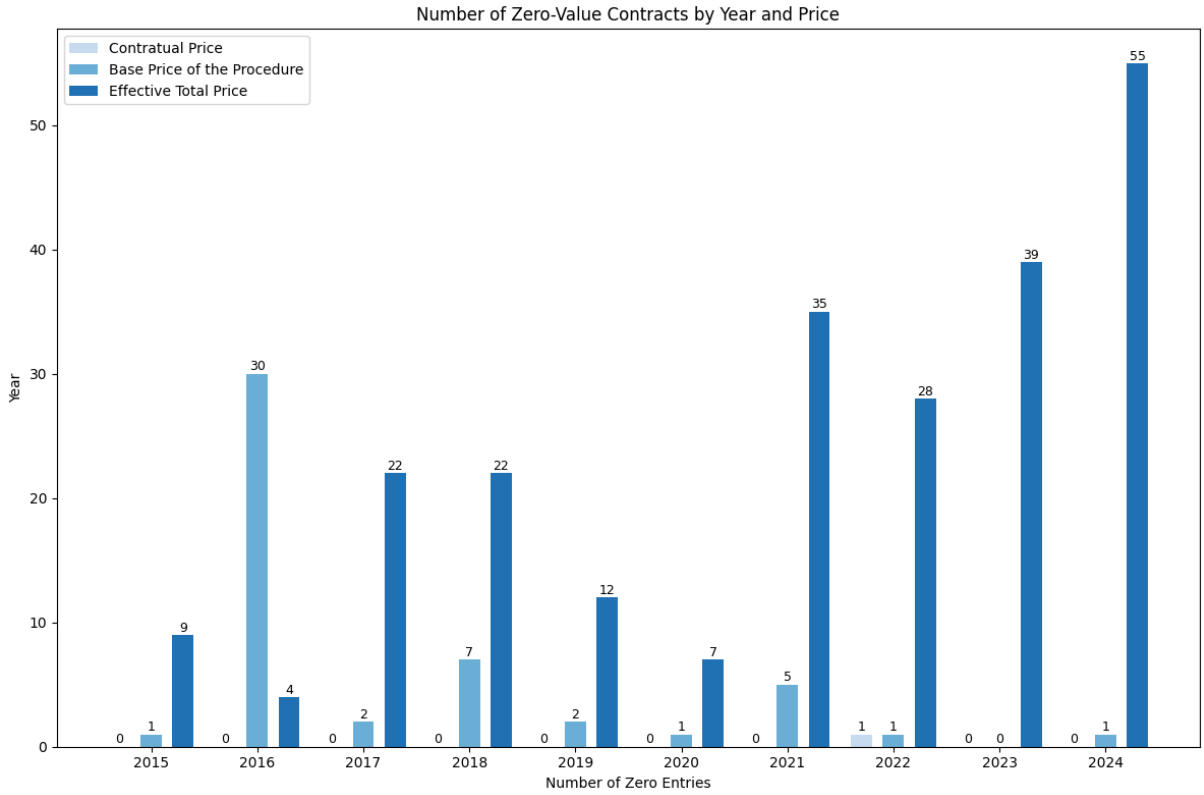


Figure 7 - Number of Zero Value Contracts by Year and Price Type

A procedure was identified in 2022 with a duration of four years but no recorded contract value (Figure 7). Upon reviewing the procedure documents – specifically Annex I of the specifications document – it was found that the contracting authority requested 18 specialist doctors, expected to provide a total of 189.100 hours over the four-year period, and 21 non-specialist doctors, expected to deliver 274.092 hours over the same period. By accessing the contract details, it was found that a rate of 30,98 € per hour was established for specialist doctors and 25,47 € per hour for non-specialist doctors, resulting in a total estimated contractual value of 12.839.441,24 €. Therefore, the contractual price for this procedure will be updated to this amount in the dataset under analysis (Direção Regional da Saúde - Portugal, 2022).

From 2015 to 2024, five procedures were identified that simultaneously include both medical and

nursing services, making it difficult to determine contract values by specific area. In terms of data processing, it was decided to duplicate these five rows corresponding to this record in the dataset – one entry for medical services and another for nursing services – and update the respective contract values accordingly.

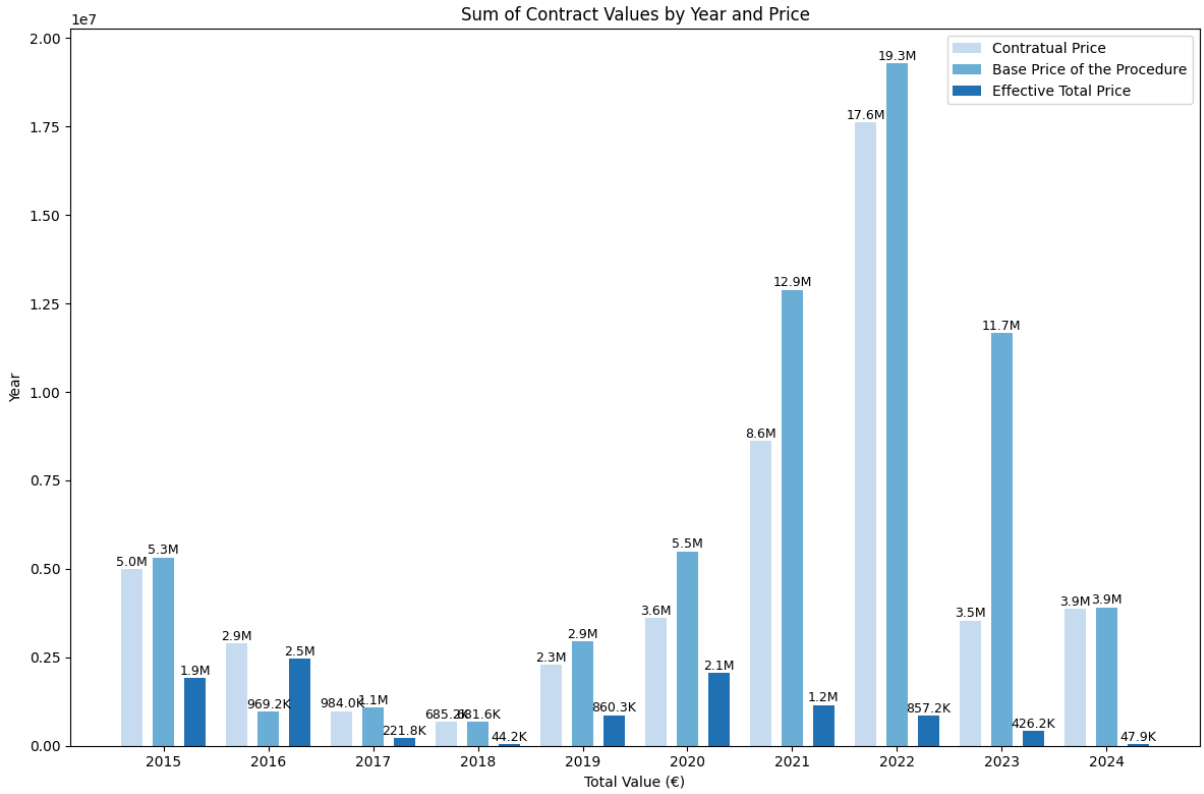


Figure 8 - Sum of Contract Values by Year and Price Type

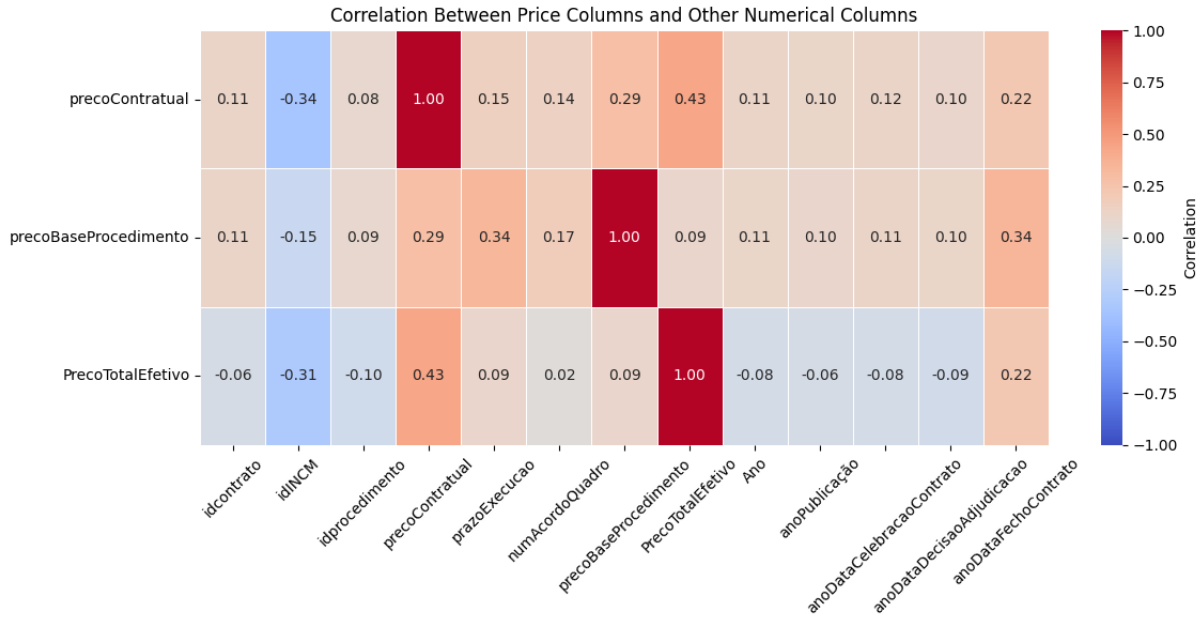


Figure 9 - Correlation Between Price Columns and Other Numerical Columns

Table 4 - Comparative Table of Mandatory Completion of Base, Contracted, and Effective Price on the Portal Base

	Base Price of the Procedure	Contractual Price	Effective Total Price
Ordinance No. 701-E/2008, of July 29	Mandatory for contracts whose contractual price is equal to or greater than 75.000. Filled in at the announcement stage.	Mandatory for all contracts. Filled in after the award.	Not explicitly required.
Ordinance No. 57/2018, of February 26	Mandatory for all non-competitive procedures. Filled in at the announcement and invitation stage.	Mandatory for all non-competitive procedures. Filled in at the contract communication stage.	Optional. Filled in at the contract completion.
Ordinance No. 318-B/2023, of October 25	Mandatory for all contracts. Filled in at the initial stage.	Mandatory for all contracts. Filled in at the stage of communication of the decision to contract and award.	Mandatory when applicable. Filled in at the contract execution stage.

In the first of these procedures, with a contractual value of 21.248,40 € and a one-year duration, an analysis of the signed contract revealed that services were requested in the areas of anaesthesiology (4.254,40 € annually), gastroenterology (4.430,40 € annually), paediatrics (1.704,00 € annually), and nursing (10.859,60 annually). This results in a total of 10.388,80 € allocated to medical services and 10.859,60 € to nursing services (Estado-Maior-General das Forças Armadas (EMGFA), 2023).

Another combined contract from 2023, with a base procedure price of 97.892,00 €, has recorded contractual value of 16.823,04 €. Of the requested lots (anaesthesiology, dermatology, gastroenterology, paediatrics, UCII, general emergency HFAR – PL, ICU HFAR – PP, urology, and nursing), the contract available specifies service acquisition only in the areas of gastroenterology, paediatrics, and nursing, totalling 16.823,04 €. However, it does not break down the contractual value by lot. According to the specifications document, the base prices presented were 4.680 € for gastroenterology, 1.800 € for paediatrics, and 11.220 € for nursing, amounting to a total of 17.700 €. The difference between this base total and the contractual value is 876,96 €. Given the small size of

this discrepancy relative to the total value of contracts under study, it was decided to proportionally distribute this difference across the three lots and subtract it from the base prices to estimate the contractual value per lot. As a result, the estimated contractual value is 5.895,36 € for medical services and 16.823,04 € for nursing services (Estado-Maior-General das Forças Armadas, 2023).

The third combined contract presents a situation similar to the previous one, but concerning the lots for UCI, General Emergency HFAR – PL, and nursing; with a base procedure price of 103.964 €. Accordingly, it totals 31.534,72 € for medical services and 10.745,36 € for nursing services. (ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS, 2023a). Similarly, the fourth contract allocates 43.331,30 € to medical services and 10.686,82 € to nursing services (ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS, 2023a) ; and the fifth contract allocates 27.037,74 € to medical services and 10.672,32 € to nursing services (ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS, 2023b).

The bar chart titled “Total Contract Value by Year” (Figure 10) compares the total contract values per year for medicine and nursing from 2015 to 2024. The values are shown in euros (€), with a y-axis scaled to millions. However, contracts may extend for more than one year. Thus, in Figure 11, the same values are shown but distributed across the years of the contract’s duration, using the feature representing the contract duration in days, rather than attributing the entire amount to the year in which the contract was awarded.

From 2015 to 2020, medicine generally registered higher contract values than nursing. In 2015, medicine peaked at around 4,69 M€, compared to nursing’s 0,27 M€. Only in 2021, medicine reached 4,38 M€ and nursing rose even higher, to 4,63 M€, surpassing medicine for the first time.

Expenditure on physician recruitment declined until 2018, reaching 0.67 M€, subsequently increasing steadily through 2024 to 6.77 €M. The decrease observed in 2025 solely reflects the recognition of multi-year recruitment contracts formalized in prior years. With regard to nursing services, the expenditure values are not linear, showing increases and decreases over shorter intervals.

The COVID-19 pandemic had a significant impact on healthcare contracts during 2021-2022, likely driving the sharp rise observed, particularly in nursing. This surge can be attributed to the urgent need for additional frontline nurses to manage the growing patient load, as well as expanded efforts in testing, vaccination, and hospital staffing. The peak in 2021 coincides with the height of pandemic response spending, reflecting the intensified demand for healthcare services and personnel during this critical period (Presidência do Conselho de Ministros, 2020).

According to the document on budgetary analysis and health revenues published by the National Statistics Institute (NSI), the financing of National Health Service (NHS) relies almost entirely on funds from the State Budget. This limited diversification of funding sources, as outlined in the current Health Framework Law (HFL), poses sustainability challenges for the NHS, especially considering the growing health needs of the population due to aging, a high burden of disease, and the potential for more

frequent pandemics driven by climate change. The expenditure of the NHS has been rising sharply, particularly in areas such as personnel and hospital medications, putting pressure on both the health system and public finances. In 2022, NHS expenditure reached 13.168 billion, representing 5.5% of GDP and 13.3% of total expenditure – the highest share since 2014. While the NHS-to-GDP ratio remained stable at around 5% from 2014 to 2018, it peaked at 5.9% in 2021 due to pandemic-related costs, and then declined slightly in 2022, reflecting strong nominal GDP growth outpacing health expenditure (Conselho de Finanças Públicas, 2023).

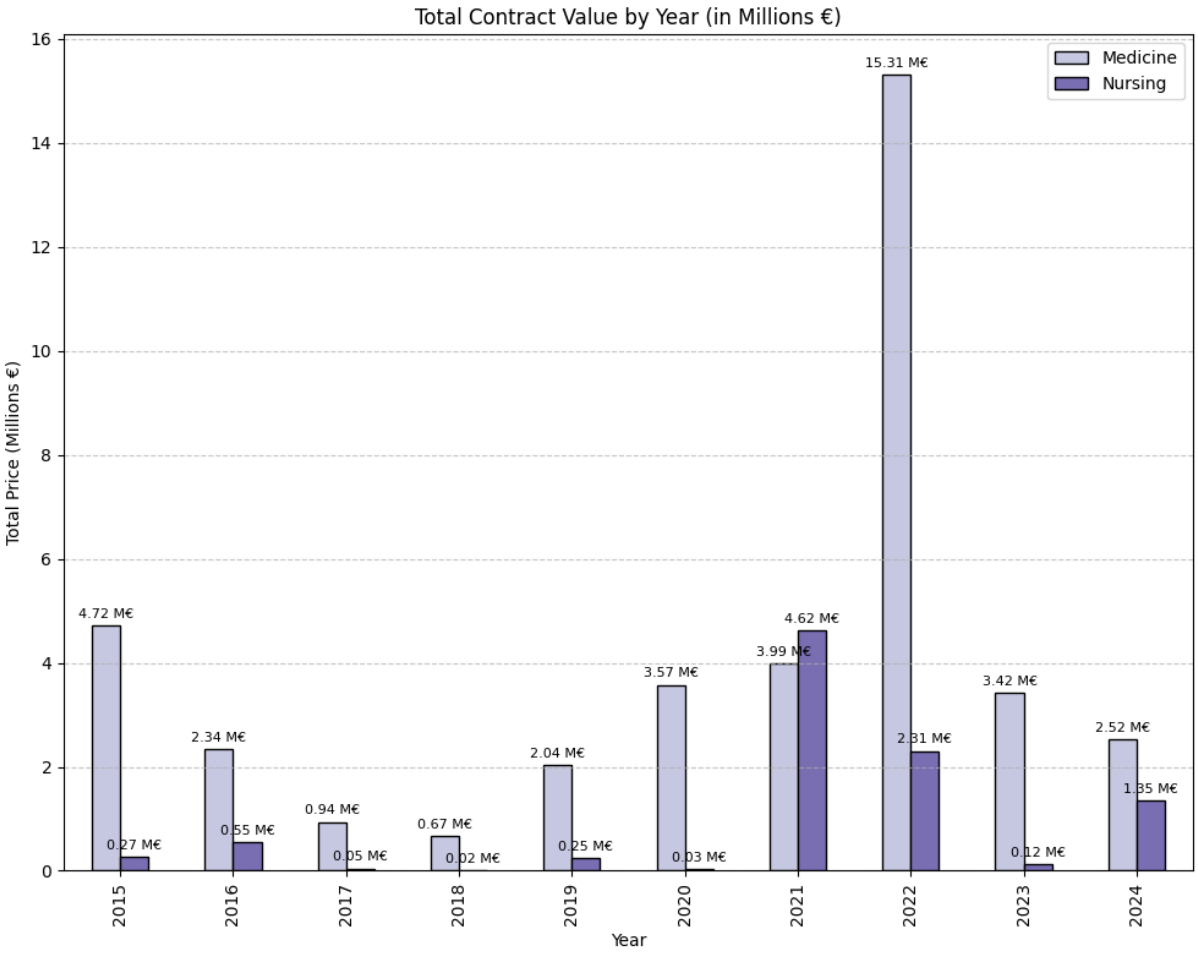


Figure 10 - Total Contract Value by Year

3.7. Analysis of the Relationship between Public Procurement and Public Employment

At this stage, the aim is to analyse the data collected so far on public procurement in the healthcare sector in relation to the number of doctors and nurses employed in the public sector. For this purpose, data provided by DGAEP was used (Direção-Geral da Administração e do Emprego Público, 2025), which includes a survey of employment in the public administration sector by subsector, according to position/career/group, education level, and gender.

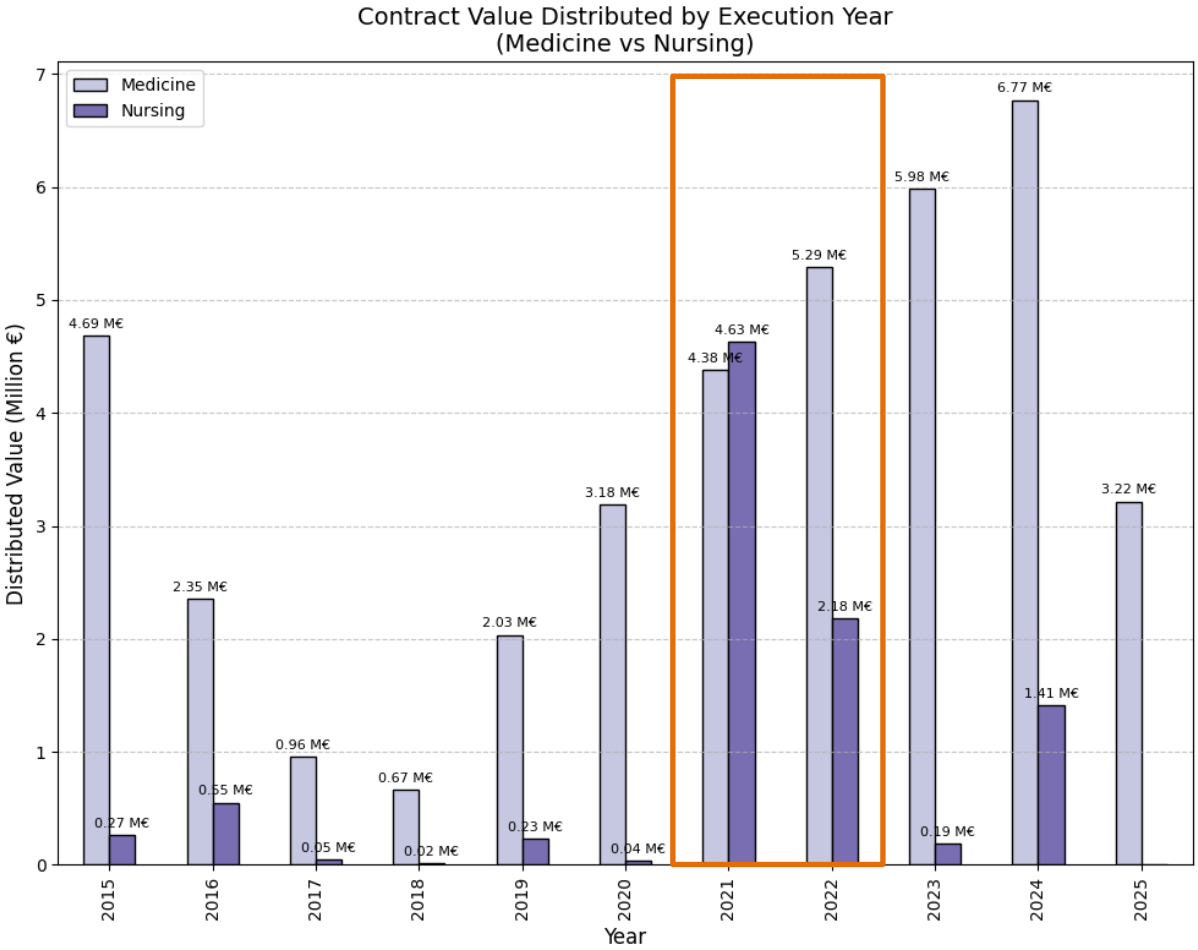


Figure 11 - Total Contract Value Distributed by Execution Year

The datasets analysed in this study represent two distinct but related dimensions of the healthcare system: the economic side of service provision, captured through nursing and medical contract values, and the workforce dimension, represented by the number of nurses and doctors in the public administration. A careful statistical and graphical examination was conducted to better understand the relationship between these domains, with particular attention to minimum and maximum values, the role of outliers, and the implications of applying different normalization and scaling methods for comparative visualization.

The inspection of ranges reveals significant differences in variability across the datasets: nursing contracts exhibit the broadest range, extending from 0.0 to nearly 4.63 million. The difference of over 4.6 million highlights the strong heterogeneity in contract values, with one contract far exceeding the others (Figure 12). Medical contracts follow a similar distributional pattern, showing wide variability. Although no single observation emerges as strongly as in the nursing dataset, the data are nonetheless dispersed across large values, reflecting the financial scale of hospital procurement. Nurses in public administration show values ranging from 86.634 to 114.861, with a difference of 28.227. This indicates a relatively stable workforce, with gradual yearly adjustments but no drastic fluctuations. Doctors in

public administration present the narrowest range, from 54.596 to 68.072, with a difference of only 13.476, further confirming the workforce's stability over time.

Boxplots were used to assess the presence of outliers (Figure 12 and Figure 13). Results confirm the existence of a single extreme outlier in the nursing contracts dataset, corresponding to an exceptionally large procurement value. This outlier exerts disproportionate influence in methods such as min-max normalization, which rely heavily on maximum values to set scales. The medical contracts dataset contains no visible extreme anomalies, although the values remain highly variable. By contrast, the nurses and doctor's workforce datasets display compact distributions, with non-evidence of outliers. This reflects the institutional stability of public workforce planning, which contracts with the occasional occurrence of extraordinary contrasts in the healthcare procurement domain. The asymmetry between datasets – with one outlier in contracts but none in workforce counts – underscores the importance of selecting appropriate comparison methods that do not distort interpretations due to a single extreme case.

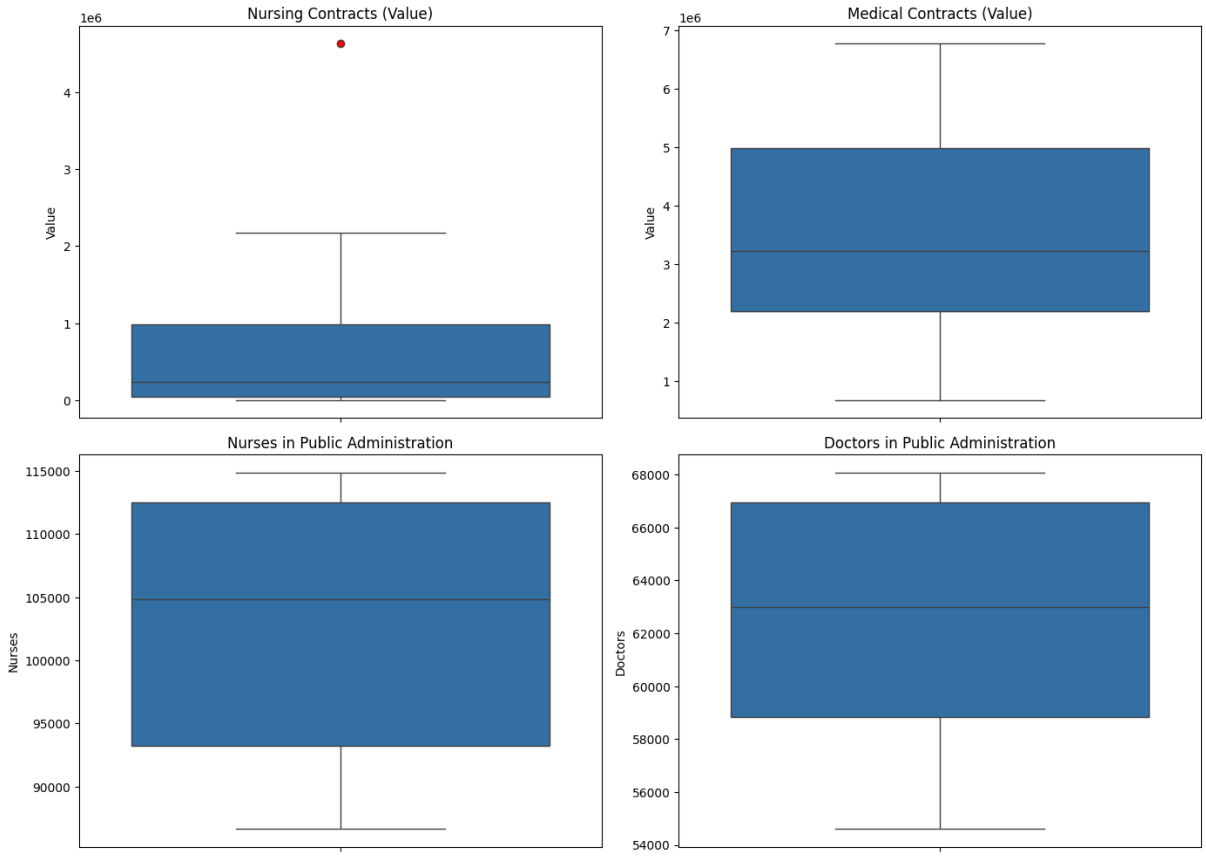


Figure 12 - Outliers of the Contract Values (in euros) and Number of Nurses and Doctors in Public Administration

To ensure a reliable comparison, it is advisable to compare magnitudes of the same order and type. Thus, using data from DGAEP, the average monthly salary of nurses and doctors was extracted for the three distinct regions corresponding to NUTS 1: Mainland Portugal, the Autonomous Regions

of Madeira and the Autonomous Regions of the Azores[11]. After determining the number of doctors and nurses in each NUTS 1 region, the total annual expenditure on salaries for doctors and nurses was estimated.

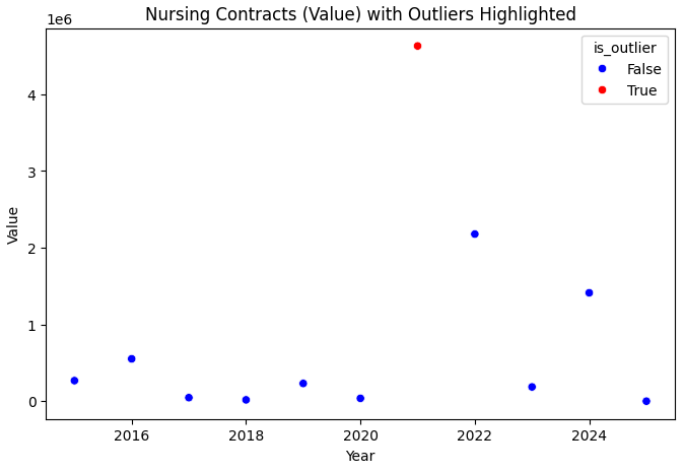


Figure 13 - Specific Outlier Study in Nursing Contracts

In this first visualization (Figure 14 and Figure 15), total annual earnings and public procurement values for doctors and nurses are displayed together on a single axis, without any normalization or scale manipulation. This approach preserves the absolute magnitudes of both measures, making the stark disparity immediately visible: annual earnings, measured in billions, dominate the scale, while procurement values, in millions, appear comparatively small. Although this method clearly illustrates the magnitude gap between the two financial dimensions, it makes detailed fluctuations in procurement activity difficult to discern.

For doctors, the correlation analysis based on absolute values revealed a moderate positive association between earnings and expenditure (Spearman's correlation: ρ=6,636 and p = 0,048). This result is statistically significant at the 5% level, suggesting that higher levels of public expenditure on procedures are associated with higher earnings for doctors. The finding indicates that doctors' remuneration appears closely linked to systematic trends in public spending, reflecting a structural connection between the allocation of resources and income levels within this professional group.

In contrast, the relationship for nurses proves weaker, possibly due to the impact of the COVID-19 period. The correlation between nurses' earnings and public procedure expenditure, although positive (Spearman ρ = 0,321), was not statistically significant (Spearman p = 0,365). This implies that changes in public expenditure are not systematically associated with nurses' earnings in the same way they are for doctors. Nurses' income appears less directly influenced by overall expenditure trends, potentially reflecting differences in contractual arrangements, bargaining structures, or the

[11] Eurostat. (2024). Overview - NUTS - Nomenclature of territorial units for statistics - Eurostat. https://ec.europa.eu/eurostat/web/nuts

mechanisms through which public funds are allocated to nursing salaries.

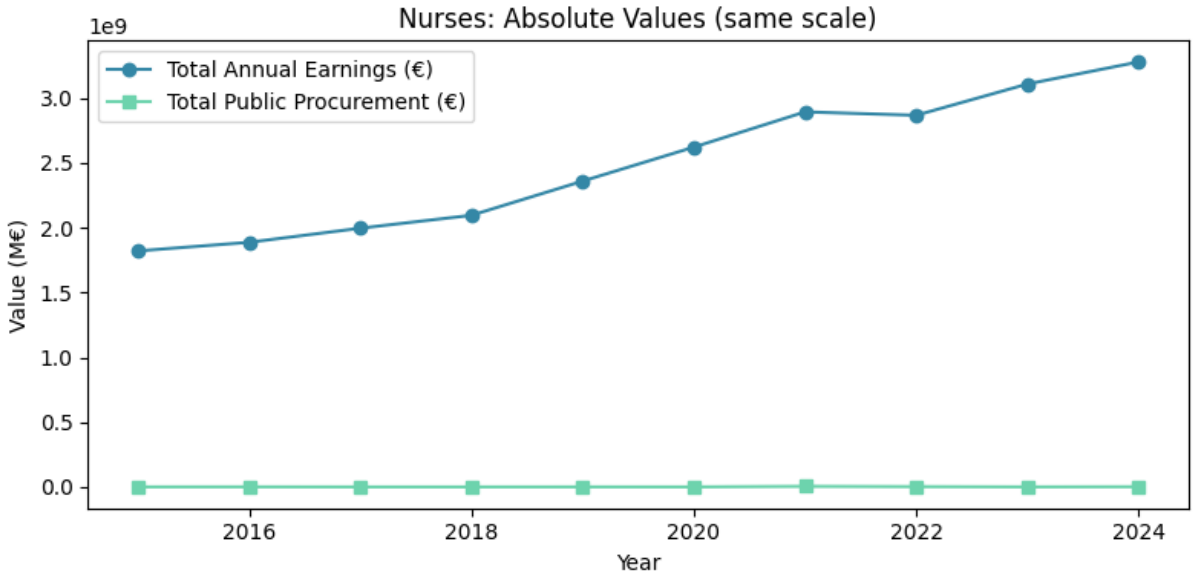


Figure 14 - Absolute Values Comparison – Nurses

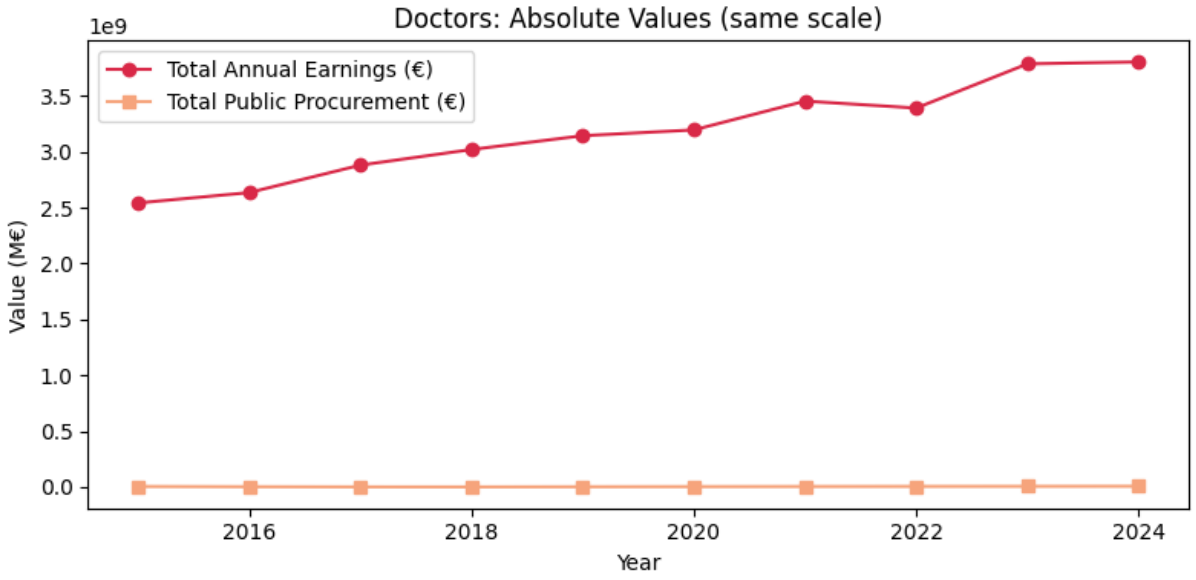


Figure 15 - Absolute Values Comparison – Doctors

The regression line plots (Figure 16) provide a complementary perspective on the relationship between total annual earnings and total public procedure expenditure. For doctors, the fitted regression line indicates a clear upward trajectory, reinforcing the positive association already identified through the Spearman correlation. The visual alignment of the data points with the regression line suggests that increases in public expenditure are systematically accompanied by increases in doctors' earnings. While Spearman's correlation captures monotonic trends without assuming linearity, the regression line highlights that this relationship also approximates a linear form.

Together, these findings strengthen the interpretation that doctors' remuneration is structurally linked to expenditure levels, with the regression plot illustrating the practical magnitude of this association.

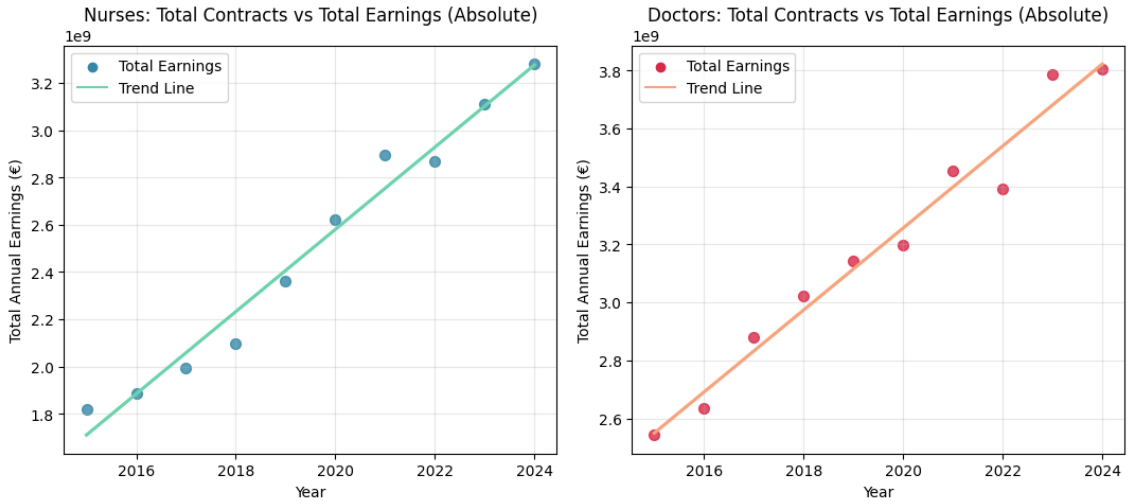


Figure 16 - Regression Line – Absolute Values

The regression line for nurses also shows a linear increase in total earnings over the period, though starting from a lower absolute value compared to doctors. While growth is steady and predictable, the absolute difference in earnings between nurses and doctors remains significant in every year observed.

For doctors, the regression line demonstrates a consistent and linear increase in total earnings, fitting the data points very well and indicating a strong and predictable relationship between year and total earnings.

In comparison, doctors consistently earn more, and the absolute growth in euros is faster than for nurses. Both graphs suggest that the number of contracts and/or average remuneration is rising over time, but earnings trajectories show that while nurses experience stable growth, doctors continue to lead in absolute earnings.

Several comparative visualization strategies were tested and evaluated, each offering distinct advantages and limitations when applied to healthcare datasets. These strategies were designed to explore the relationship between workforce numbers – specifically doctors and nurses in public administration – and the financial contracts associated with their services.

To enrich and complete this comparison, two additional factors were included: the government in office and the distribution of members of parliament by parliamentary group or by single-party representative in the Republic Assembly. Based on information from the Portuguese Parliament's official website (Parlamento, 2025), the number of elected deputies in each election from 2015 to 2024 was determined, and from government sources[12], the composition of the government in office was

[12] Governo da República Portuguesa. (2025). Governos Anteriores - XXV Governo Constitucional. https://www.portugal.gov.pt/pt/gc25/governo/governos-anteriores

identified, whether formed independently or through coalitions.

The goal is to determine which visualization method most effectively communicates the interplay between these two dimensions: workforce volume and expenditure. This is particularly relevant in policy analysis, where understanding not just absolute values but proportional dynamics and temporal trends is essential.

In summary, when applied to the datasets: the dual y-axis plots expose the magnitude gap workforce monetary expenditure and contract values, but they obscure proportional changes over time; the baseline normalization reveals that contracts grew and fluctuated far more than workforce expenditure, offering a clear view of relative dynamics and highlighting key policy implications; the logarithmic scaling proves to be the most effective overall, as it mitigates the impact of outliers – especially in nursing contracts – and allows both series to be visually comparable despite scale differences; the min-max normalization emphasizes the shape of trends, but its effectiveness is undetermined by the presence of extreme values, which distort the visual representation.

In this case, dual Y-axis, baseline-normalized, and min-max normalized plots all present similar visual patterns and insights. These three methods emphasize proportional relationships and temporal alignment between earnings and public procurement, while the logarithmic scale offers a distinct lens, smoothing volatility and revealing subtler shifts.

This comparative evaluation provides a foundation for selecting appropriate visualization techniques in future analyses, especially when dealing with datasets that combine financial and human resource metrics across disparate scales.

The trajectory of nurses' earnings in Portugal from 2015 onward cannot be fully understood without acknowledging the lingering effects of the Troika program – a period of external financial oversight imposed by the European Commission, European Central Bank, and International Monetary Fund. During the Troika years (2011-2014), Portugal underwent severe austerity measures, which included wage freezes, reduced public sector hiring, and cuts to healthcare spending. Although the data begins in 2015, the aftermath of these policies is still felt in the early years of the graphs.

From 2015 to 2024, nurses' earnings in Portugal don't reveal a consistent and tightly coupled relationship with public procurement spending. Across the dual Y-axis, baseline-normalized, and min-max normalized plots, the visual story is misaligned (Figure 17).

The early years of the timeline reflect the post-Troika recovery. Though the Troika program formally ended in 2014, its austerity legacy—wage freezes, hiring limits, and constrained healthcare budgets—lingered into 2015. Nurses' earnings begin at a suppressed level, and procurement is modest. With the rise of a Socialist Party (PS) - led government supported by Left Bloc (BE), Portuguese Communist Party (PCP), and Ecologist Party "The Greens" (PEV), a deliberate pivot away from austerity emerges. Procurement begins to climb, and nurses' earnings follow in close rhythm. This alignment is

visible in all three plot styles: dual Y-axis, baseline-normalized, and the min-max chart.

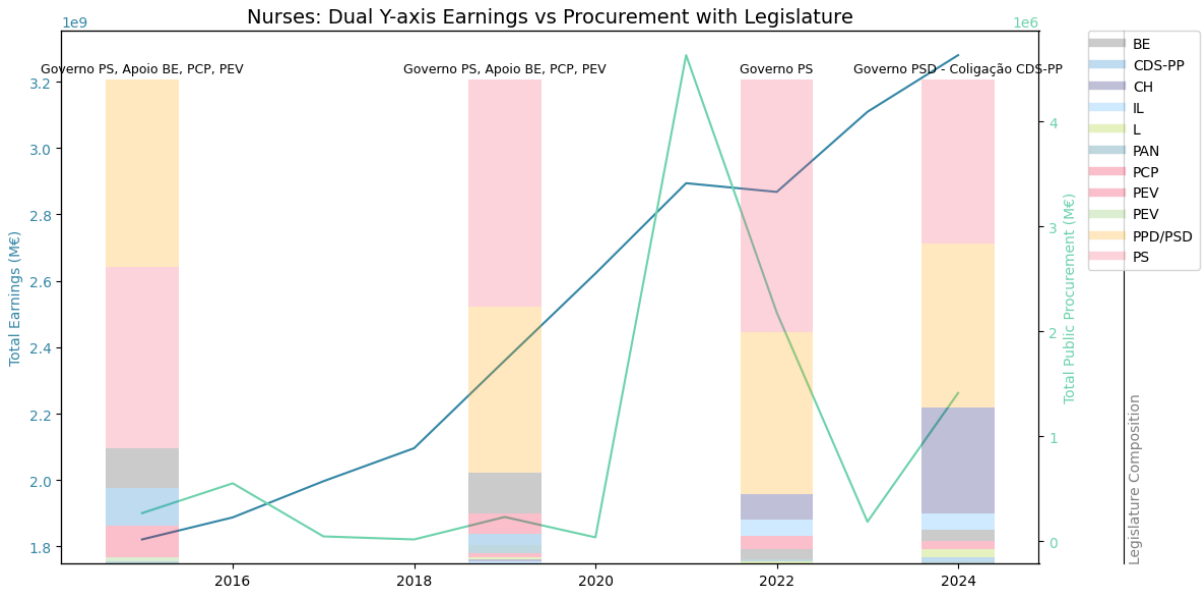


Figure 17 - Dual Y-axis Comparison – Nurses

The COVID-19 pandemic, beginning in 2020, marks a dramatic inflection point. Procurement spikes sharply reflecting emergency spending on equipment, testing, and hospital capacity. Nurses' earnings also rise, though more gradually. The three plots capture this moment with synchronized peaks, showing how the pandemic catalysed investment and elevated compensation. However, the rise in earnings is tempered by structural inertia: while procurement can be scaled rapidly, salary adjustments require policy reform and negotiation.

After 2022, the impact of COVID-19 pandemic decreases, and procurement begins to decline. Nurses' earnings growth slows in tandem. The three plots reflect this contraction clearly: procurement dips, and earnings flatten. The visual consistency across the three methods reinforces the conclusion that nurses' earnings are not sensitive to procurement. In 2024, a Social Democratic Party - Social Democratic Centre (PSD-CDS) coalition takes power and procurement rises again.

Doctor's earnings over the same period show a more complex relationship with procurement. While the three plot styles – dual Y-axis, baseline-normalized, and min-max normalized – still reveal a general alignment, the correlation is looser and punctuated by sharper jumps and pauses (Figure 18).

In the post-Troika years, doctors begin with higher earnings than nurses but still face the constraints of prior austerity. As the PS - led coalition takes office, procurement rises steadily, and doctors' earnings follow – but with more volatility. These patterns suggest that doctors' compensation is shaped not only by procurement but also by other factors like contractual negotiations, union influence, and targeted reforms.

During the COVID-19 pandemic, procurement surges dramatically. Doctors' earnings also rise, but not in perfect sync. The three plots show earnings increasing, but with lag and irregularity. This reflects

the nature of pandemic spending: while procurement focused on infrastructure and frontline staffing, doctors—often in specialized roles—may have seen delayed or uneven compensation changes. Nevertheless, the upward trend is visible, and the three plot styles capture the general alignment.

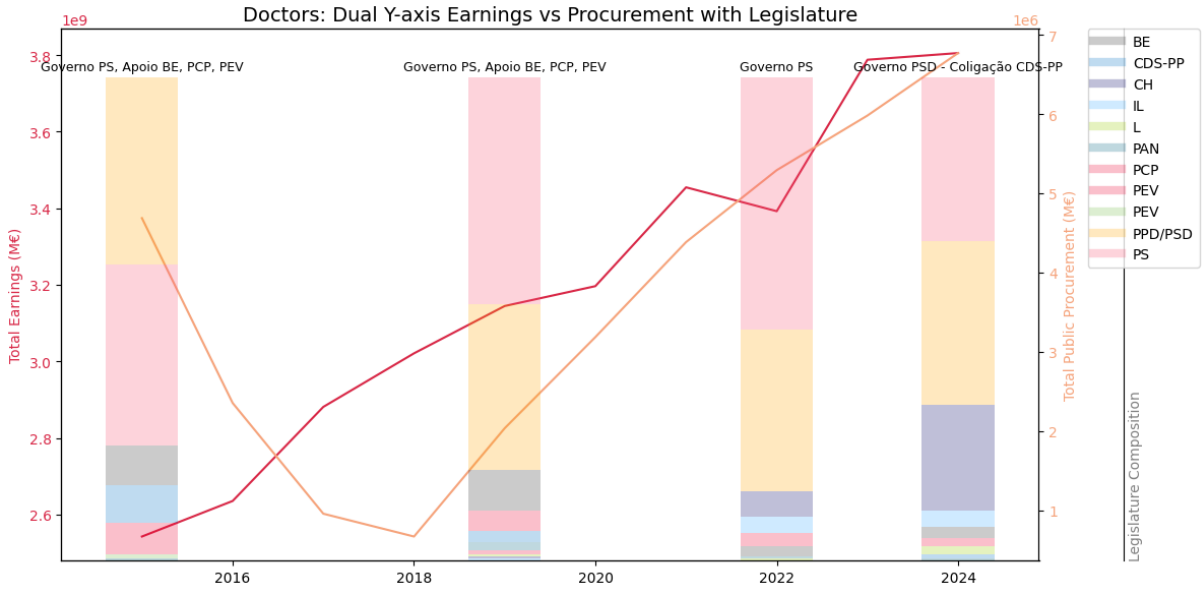


Figure 18 - Dual Y-axis Comparison – Doctors

Post-2022, under PS government, procurement dips and earnings growth slows. Yet doctors' earnings remain more stable than nurses', suggesting institutional resilience. The three plots show a softening of growth, but not a reversal. This indicates that doctors' compensation, while influenced by procurement, is buffered by longer-term contracts and strategic positioning within the healthcare system.

In the log-scaled chart, nurses' earnings grow steadily from 2015 onward, with fewer abrupt jumps. Procurement shows more fluctuation, especially around the pandemic years. The log scale reveals that while procurement surged in 2021, nurses' earnings maintained a linear trajectory—suggesting that emergency spending did not immediately translate into salary increases. This view underscores the structural lag between investment and compensation and highlights the resilience of nurses' earnings even aimed political fragmentation (Figure 19).

Doctors' earnings in the log-scaled chart show sharper inflections, particularly around 2018 and 2022. These jumps may correspond to contract renegotiations or policy reforms, rather than procurement shifts alone. Procurement still rises during the pandemic, but doctors' earnings respond with delay and irregularity. The log scale makes clear that doctors' compensation is less reactive and more strategically managed. Post-pandemic, earnings continue to grow, even as procurement flattens—suggesting autonomy from short-term budget cycles. (Figure 20)

When the analysis is extended to year-to-year differences, both professional groups exhibit weak and statistically non-significant associations (Figure 21 and Figure 22). For doctors, the correlation

between annual changes in earnings and expenditure was slightly negative (Spearman ρ = -0,310 and p = 0456), indicating no consistent monotonic relationship between short-term fluctuations in the two variables. A similar pattern was observed among nurses, where the correlation was also negative (Spearman ρ = -0,262 and p = 0,531), again with no evidence of statistical significance. These results suggest that while the overall levels of doctors' earnings and public expenditure are moderately related over time, short-term variations do not exhibit the same degree of alignment. For nurses, both long-term and short-term associations appear weak and largely independent of changes in expenditure.

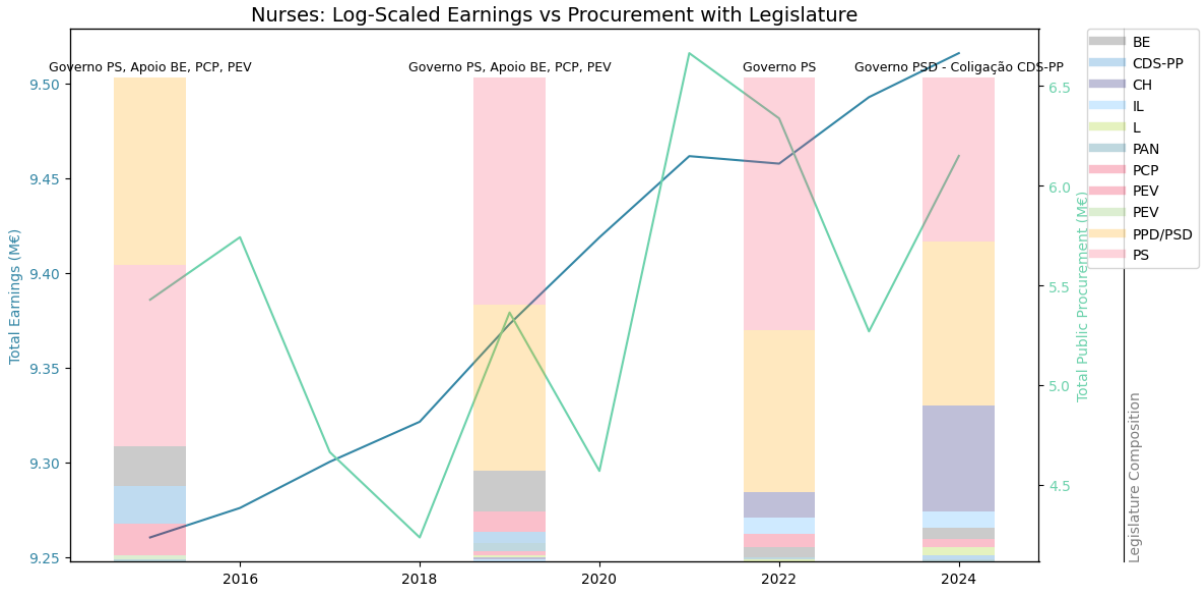


Figure 19 - Logarithmic Scale Comparison – Nurses

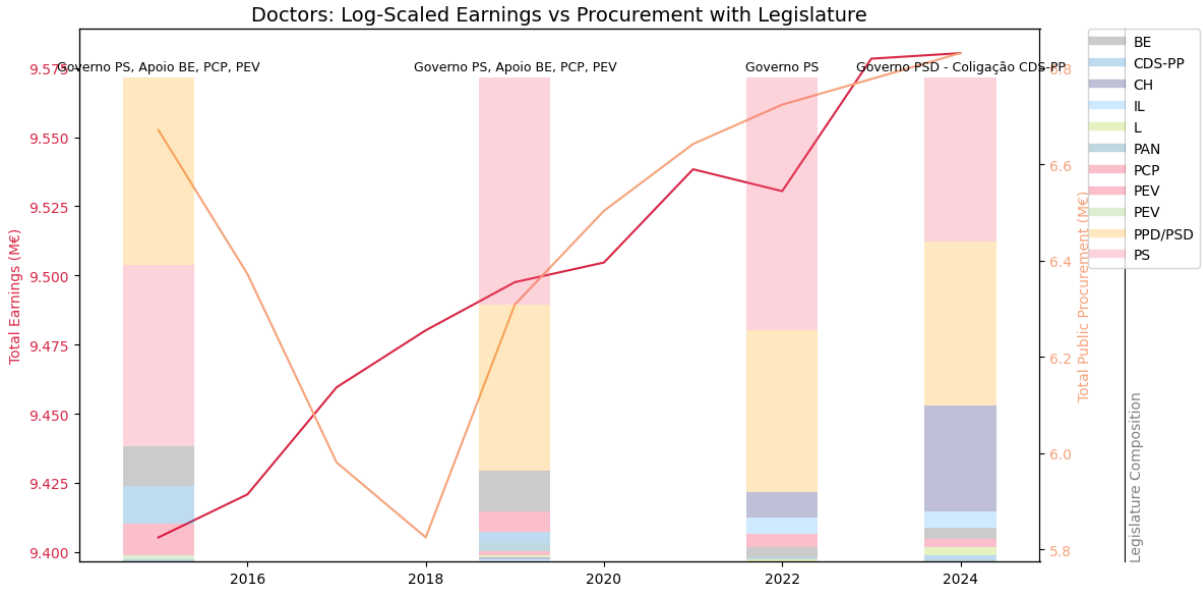


Figure 20 - Logarithmic Scale Comparison – Doctors

When year-to-year differences are plotted (Figure 23), the regression lines for both doctors and

nurses show almost no discernible slope, reflecting the absence of a systematic linear relationship between annual changes in earnings and expenditure. This observation aligns with the negative and non-significant Spearman coefficients. The flatness of the regression lines highlights that short-term fluctuations in expenditure do not translate into proportional changes in professional earnings for either group. The visual evidence therefore confirms that the meaningful association lies in the long-term levels of expenditure and earnings (particularly for doctors), rather than in their year-to-year volatility.

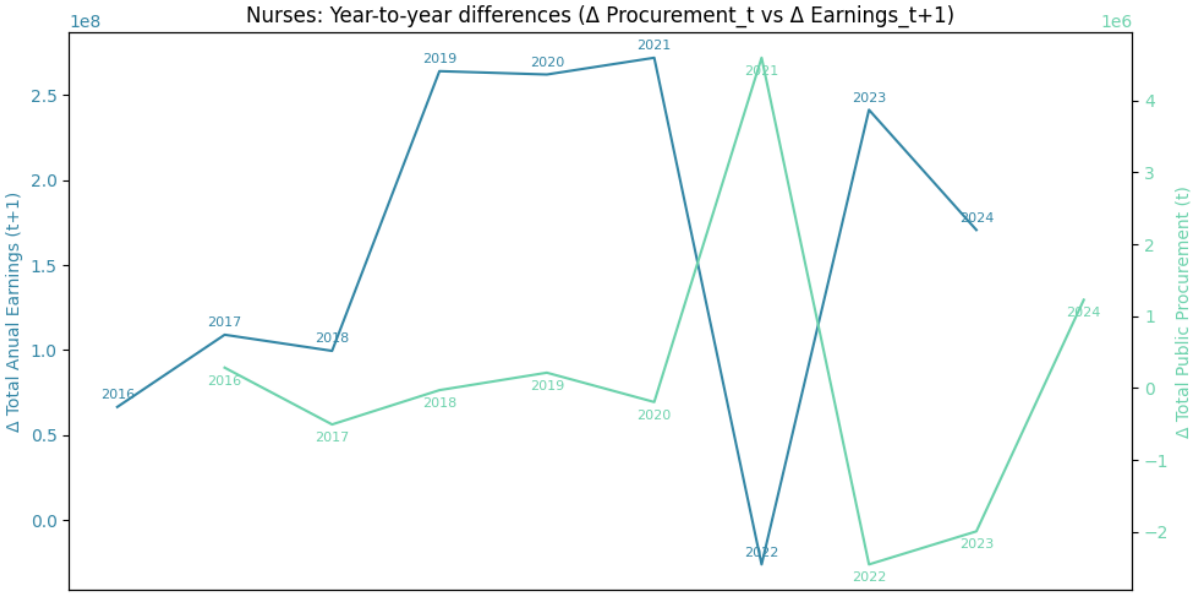


Figure 21 - Year-to-year difference - Nurses

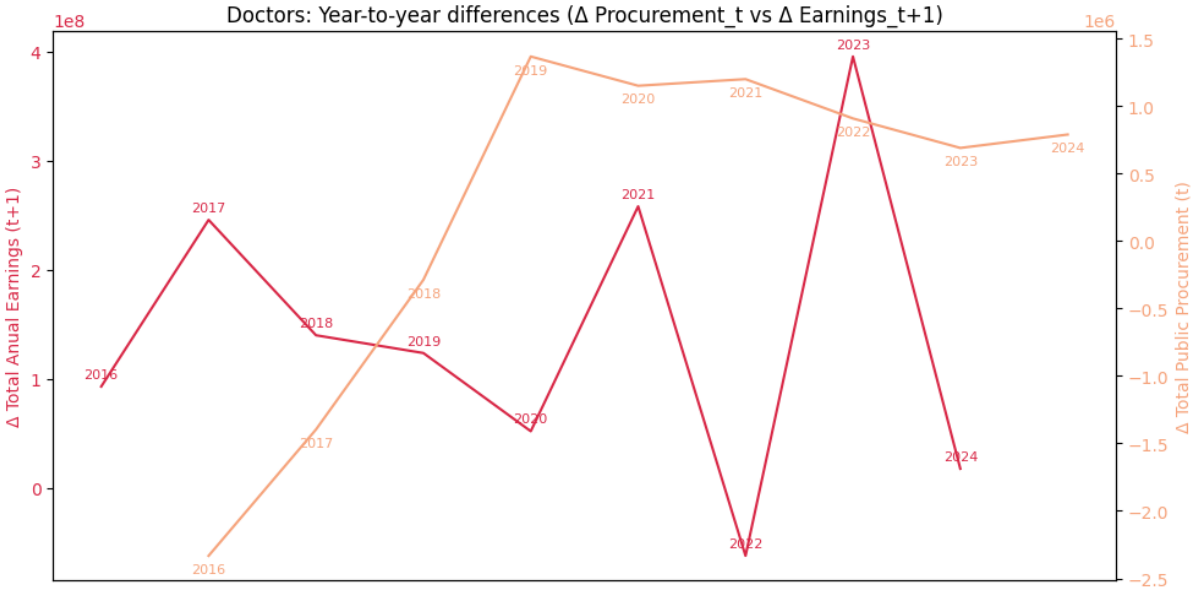


Figure 22 - Year-to-year difference – Doctors

Taken together, the regression plots complement the correlation analysis by emphasizing both the strength and direction of the relationships. For doctors, the regression line illustrates a robust, approximately linear association between public expenditure and earnings, suggesting a structural dependency. For nurses, the line is weak and unstable, illustrating the lack of a consistent link. For both groups, the year-to-year regressions confirm that short-term variations are largely independent of expenditure shifts. These patterns support the broader conclusion that doctors' incomes are significantly conditioned by systemic trends in public procedure spending, whereas nurses' earnings follow a more autonomous trajectory.

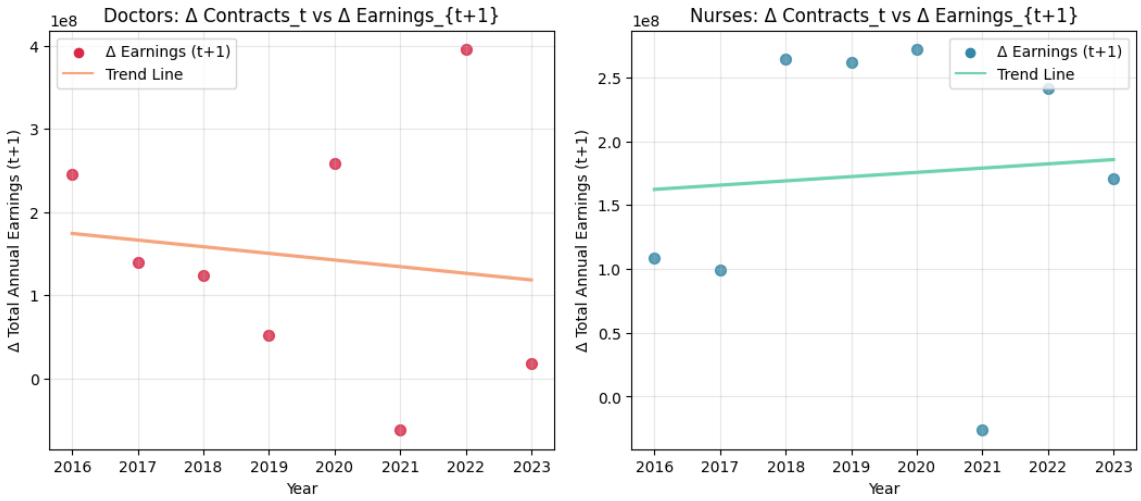


Figure 23 - Regression Line – Year-to-year difference

CHAPTER 4

Conclusions

This dissertation has shown that patterns of public procurement in Portugal's healthcare sector between 2015 and 2024 reflect both structural reforms and crisis-driven pressures. The procurement modest and gradual growth from 2015 to 2016/17 coincided with post-Troika recovery measures, while the COVID-19 pandemic triggered an unprecedented surge in 2020–2021, reshaping expenditure dynamics across the system.

The statistical analyses demonstrated that physicians' total earnings are significantly associated with levels of public procedure expenditure. Both correlation and regression models confirmed a moderate, positive, and statistically significant relationship, indicating that doctors' remuneration tends to rise in parallel with systemic increases in procurement spending. This reflects the centrality of physicians' contracts and fee structures in translating macro-level funding into individual income.

For nurses, however, the results revealed a weaker and non-significant association. Although an exceptional rise in contracts occurred during the pandemic years to meet urgent frontline needs, regression models indicated that such expenditure fluctuations do not consistently explain variations in nurses' annual earnings. Once the pandemic outlier was excluded, the association remained statistically insignificant, suggesting that nurses' remuneration is more strongly shaped by institutional wage agreements and structural labour market conditions than by procurement cycles.

Taken together, these findings illustrate a divergence in how public procurement expenditure translates into professional earnings. Doctors' remuneration is closely tied to systemic expenditure flows, reflecting stronger links between macro-level funding and contractual arrangements. By contrast, nurses' earnings follow a more autonomous trajectory, with procurement peaks largely reflected in contract volumes rather than direct income adjustments. This asymmetry underscores broader differences in bargaining power, wage-setting mechanisms, and the professional hierarchies shaping the healthcare labour market.

4.1. Methodological Contributions

One of the most significant challenges of this research was the data acquisition stage. Extracting information from the Portal Base extended over a period of more than two months and required the generation of millions of individual CSV files, overcoming the platform's 500-record limit by filtering queries by procedure type, region, and date intervals. This step, although essential to guarantee completeness and reliability, consumed far more time than initially anticipated.

The subsequent stages of data aggregation and cleaning also demanded extensive effort, consolidating fragmented monthly datasets into annual Parquet files, converting text fields to numeric

and datetime types, splitting combined entity identifiers, and enabling reproducible analyses with visualization techniques. Handling fragmented downloads, identifying missing or empty files, and retrying downloads, correcting encoding inconsistencies resolving UTF-8/Latin-1 encoding issues, and ensuring temporal accuracy were labour-intensive tasks that delayed the transition to the analytical phase. Ultimately, while the methodological framework developed in this dissertation proved robust and replicable, the disproportionate amount of time devoted to data acquisition and preparation highlights an important limitation. Future studies will benefit greatly from improved data accessibility, standardized formats, and the availability of APIs in public procurement platforms.

4.2. Study Limitations

Although the methodology produced a comprehensive view of medical and nursing procurement, it relied exclusively on Portal Base's public CSV downloads, subject to inconsistent field mandates and occasional zero-value entries that demanded manual corrections. The focus on healthcare professionals limits immediate generalizability to other public sector roles, and the absence of complementary API public endpoints constraint real-time updating and scalability.

4.3. CPV Classification of Healthcare Contracts

To isolate procurements for medical and nursing services, every Portuguese-language CPV entry in Portal Base was processed hierarchically in Jupyter Lab. First, divisions corresponding to social, defence, business and health services were selected. Next, within those divisions, groups and classes implying human – resource support – such as recruitment and staffing – were filtered, and finally categories explicitly naming medical or nursing personnel were extracted. An initial manual shortlist of 222 CPV codes was then validated against a dictionary – based term search over contract descriptions, producing 112 final codes. From this end, two principal codes emerged as encompassing all external staffing contracts in the public health sector: 79624000-4 for nursing personnel supply and 79625000-1 for medical personnel supply.

4.4. Exploratory Data Analysis

Exploration began by charting the annual count of medical and nursing contracts from 2015 through 2024, revealing peaks in 2015 and 2024 for doctors and in 2016 and 2024 for nurses. Price-range histograms uncovered clusters of zero-value procedures, prompting a decision to focus on the contractual price field for consistency.

In Portugal, several ministerial ordinances defined which price fields were mandatory at different times. Ordinance 701-E/2008 (July 29) required recording the contractual price for all contracts above

75.000 €, Ordinance 57/2018 (February 26) extended mandatory entry for base and contractual prices to all non-competitive procedures, and for every contract while requiring the effective total price when applicable. Despite these legal requirements, many records were found with empty or zero-value entries. This occurred because data-entry operators sometimes left even mandatory fields blank – planning to complete them later – or because earlier regimes allowed optional fields to remain unpopulated.

When plotting yearly sums of base, contractual and effective prices, divergent trends became apparent, but their similar correlations with other numerical features indicated that each price metric tracked the same underlying patterns. Boxplot-base outlier analysis highlighted one exceptionally large nursing contract, which in turn guided the selection of normalization methods – standard baseline, min-max scaling and logarithmic transformation – to ensure comparability across series.

4.5. Linking Procurement and Workforce Data

To assess how investment in external staff translated into public employment, DGAEP head-count data for nurses and doctors in public administration were merged with annual procurement values. This combined dataset was further enhanced with electoral and government information: on October 4, 2015, a Socialist Party government took office with parliamentary support from BE, PCP and PEV; following the October 6, 2019 election, the same arrangement continued under a PS minority administration; after the January 30, 2022 ballot, the Socialist Party won an absolute majority and governed alone; finally, the March 20, 2024 election ushered in a coalition led by PSD together with Social Democratic Centre – Popular Party (CDS-PP).

When procurement and head-count series were plotted alongside these legislative milestones, procurement spikes – most notably during the COVID-19 emergency, while physician head-count and pay rose more gradually, reflecting slower contractual renegotiations. This analysis demonstrates that shifts in public procurement serve as leading indicator for health-sector workforce expansion, and that changes in parliamentary majorities and coalition arrangements have shaped both scale and timing of these staffing responses.

Overall, the findings point to a differential relationship between public procedure expenditure and professional earnings across occupational groups. Doctors' income displays a significant and positive association with expenditure levels, reinforcing the interpretation that systematic increases in public funding are mirrored in doctors' remuneration. Nurses, by contrast, possibly due to the outlier COVID-19, show no such consistent relationship, highlighting a weaker or less direct connection between macro-level spending decisions and their earnings. This divergence may reflect structural differences in how compensation is determined, as well as broader disparities in how public investments in

healthcare procedures translate into income distribution across professional categories.

Research Question 1. What behavioural patterns can be detected in public procurement, within the public health sector, from 2015 to 2024, particularly in response to the COVID-19 pandemic?

The analysis of Spearman correlations reveals distinct behavioural patterns in public procurement across the studied period. Procurement activity shows clear responsiveness to macro-political and crisis events. The values in 2015/2016 aligns with post-Troika recovery measures, while an unprecedented surge in 2020–2021 reflects the extraordinary demands of the COVID-19 pandemic.

When paired with human resources data, Spearman's correlation underscores a differential relationship across professional groups. For doctors, the correlation between total annual earnings and total public procedure expenditure is moderate and statistically significant (ρ = 0,636, p = 0,048). This indicates a monotonic relationship in which systemic increases in procurement expenditure tend to be mirrored by higher physician earnings, pointing to a structural connection between macro-level funding flows and doctors' remuneration. By contrast, the relationship for nurses is weaker and non-significant (ρ = 0.321, p = 0.365), suggesting that procurement surges — although visible in contract volumes — do not systematically translate into changes in nurses' aggregate earnings.

When year-to-year differences are considered, the correlations for both groups turn negative and remain non-significant (doctors: ρ = –0.310, p = 0.456; nurses: ρ = –0.262, p = 0.531). This absence of a consistent monotonic relationship indicates that short-term fluctuations in procurement, particularly during crisis years, do not directly explain contemporaneous changes in remuneration. In substantive terms, the COVID-19 shock represents an exceptional procurement event, but its translation into earnings appears uneven across professions, with stronger long-term effects for physicians and more immediate but transient effects for nursing contracts.

These results were reported and discussed in Section 3.6 (Exploration and Analysis of the Data) and in Section 3.7 (Analysis of the Relationship between Public Procurement and Public Employment) of the dissertation.

Research Question 2. How is public procurement in the healthcare sector related to the creation of permanent staff positions?

The Spearman correlation results demonstrate a statistically significant monotonic association between procurement expenditure and doctors' total annual earnings, but not for nurses. While this provides evidence of a link between procurement levels and remuneration structures, it does not establish a direct relationship with the creation of permanent positions. For nurses, the weak and non-significant correlation suggests that procurement-driven increases, particularly during the pandemic, were likely absorbed through temporary contract expansions rather than long-term structural

adjustments to earnings. For doctors, the significant positive correlation indicates a closer alignment of income with systemic expenditure, but this may reflect collective bargaining or contract renegotiations rather than net additions to permanent staff.

When focusing on year-to-year differences, the absence of significant correlations for both groups further supports the interpretation that procurement peaks did not translate into stable, monotonic changes in remuneration that would be consistent with permanent staff expansion. Instead, the evidence suggests that procurement growth was more closely tied to short-term needs, with nurses responding through rapid contract increases during the pandemic, while doctors' earnings evolved more gradually in response to structural adjustments.

Accordingly, while procurement expenditure clearly influenced labour market dynamics, the Spearman correlations alone cannot demonstrate a causal or systematic link with the creation of permanent posts.

These results were reported and discussed in Section 3.7 (Analysis of the Relationship between Public Procurement and Public Employment) of the dissertation.

4.6. Future Research Directions

Despite these constraints, and the fact that few scientific studies have systematically addressed this topic, underscoring the originality and relevance of this research, the research successfully achieved its objectives and demonstrated that public procurement data can serve as a complementary indicator for strategic workforce planning in the healthcare sector. The methodological contribution of this work – particularly the large-scale extraction, aggregation, and cleaning pipeline – constitutes a valuable foundation for future research, especially in contexts where public data is fragmented and not readily accessible.

This project is founded on a strategic collaboration with public agencies aimed at establishing standardized API access and enhancing the quality of procurement metadata. By reducing extraction overhead and enabling more systematic analyses, this approach ensures that procurement data are both comprehensive and reliable. The analytical framework is further extended to encompass additional health-related professions, including pharmacists and diagnostic technicians, through the careful refinement of CPV code mappings, thereby providing a nuanced understanding of sector-specific procurement patterns.

Central to this initiative is the integration of advanced AI and machine-learning techniques, which serve to augment the public procurement process in a proactive and intelligent manner. These algorithms facilitate the detection of behavioural patterns, identification of errors, and uncovering of potential fraud, while simultaneously contributing to the optimization of adjudication procedures.

Beyond procedural improvements, AI-driven insights inform more effective management of human resources by highlighting workforce requirements and operational inefficiencies, thus fostering a more strategic allocation of personnel within public institutions.

To ensure the generalizability and robustness of the methodology, cross-sector analyses are conducted across areas such as education and infrastructure. This validation demonstrates the replicability of the large-scale data pipeline and its capacity to enhance transparency, efficiency, and evidence-based decision-making across diverse public procurement contexts. By harmonizing standardized data access, AI-enhanced analytics, and cross-sector validation, the study provides a comprehensive and forward-looking framework for modernizing procurement practices and supporting sustainable improvements in public sector governance.

References

Butkus, M., Rakauskiene, O. G., Bartuseviciene, I., Stasiukynas, A., Volodzkiene, L., & Dargenyte-Kacileviciene, L. (2023). Measuring quality perception of public services: customer-oriented approach. *Engineering Management in Production and Services*, *15*(2), 96–116. https://doi.org/10.2478/emj-2023-0015

Carneiro, D., Veloso, P., Ventura, A., Palumbo, G., & Costa, J. (2020). Network Analysis for Fraud Detection in Portuguese Public Procurement. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12490 LNCS*, 390 – 401. https://doi.org/10.1007/978-3-030-62365-4_37

Conselho de Finanças Públicas. (2023). *EVOLUÇÃO DO DESEMPENHO DO SERVIÇO NACIONAL DE SAÚDE EM 2022*. www.cfp.pt,

Curado, A., Damasio, B., Encarnacão, S., Candia, C., & Pinheiro, F. L. (2021). Scaling behavior of public procurement activity. *PLoS ONE*, *16*(12 December). https://doi.org/10.1371/journal.pone.0260806

Da Rosa, I., Tribolet, J., & Da Silva, M. M. (2015). Portuguese public procurement governance model. *Atas Da Conferencia Da Associacao Portuguesa de Sistemas de Informacao*, *15*, 529–554. https://doi.org/10.18803/capsi.v15.529-554

Direção Regional da Saúde - Portugal, T. do H. (2022). *Concurso Público para a celebração de Contratos Públicos de Aprovisionamento relativos à prestação de serviços médicos às Unidades de Saúde de Ilha do Serviço Regional de Saúde*. https://www.base.gov.pt/Base4/pt/detalhe/?type=contratos&id=9203164

Direção-Geral da Administração e do Emprego Público. (2025). *DGAEP - Direção-Geral da Administração e do Emprego Público*. https://www.dgaep.gov.pt/index.cfm?OBJID=C0F56E62-5381-4271-B010-37ECE5B31017

Estado-Maior-General das Forças Armadas. (2023). *Aquisição de Serviços Médicos e Enfermagem para HFAR PL e PP*. https://www.base.gov.pt/Base4/pt/detalhe/?type=contratos&id=9824036

ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS. (2023a, March). *Aquisição de Serviços Médicos e Enfermagem HFAR PL e PP março 2023*. https://www.base.gov.pt/Base4/pt/detalhe/?type=contratos&id=9876962

ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS. (2023b, May 19). *Serviços Médicos e de Enfermagem para HFAR PL e PP*. https://www.base.gov.pt/Base4/pt/detalhe/?type=contratos&id=10020116

Estado-Maior-General das Forças Armadas (EMGFA). (2023). *Aquisição de Serviços Médicos e de Enfermagem HFAR PL e PP*. https://www.base.gov.pt/Base4/pt/detalhe/?type=contratos&id=9722622

Ferreira, I., & Amaral, L. A. (2016). Public e-procurement: Advantages, limitations and technological "pitfalls." In E. E., B. J., & M. S. (Eds.), *ACM International Conference Proceeding Series* (Vols. 01-03-March-2016, pp. 9 – 12). Association for Computing Machinery. https://doi.org/10.1145/2910019.2910089

Instituto dos Mercados Públicos, do I. e da C. I. P. (2025). *Portal BASE - Contratos Públicos Online*. https://www.base.gov.pt/base4

Instituto dos Mercados Públicos - Portal Base. (2024). *Contratação Pública em Portugal 2023*. https://www.base.gov.pt/Base4/media/1elfqicv/relat%C3%B3rio-anual-da-contrata%C3%A7%C3%A3o-p%C3%BAblica-2023.pdf

Jacques de Sousa, L., Poças Martins, J., & Sanhudo, L. (2023). Tackling the Data Sourcing Problem in Construction Procurement Using File-Scraping Algorithms. *Engineering Proceedings*, *53*(1). https://doi.org/10.3390/IOCBD2023-15190

Jacques de Sousa, L., Poças Martins, J., & Sanhudo, L. (2024). Predicting construction project compliance with machine learning model: case study using Portuguese procurement data. *Engineering, Construction and Architectural Management*, *31*(13), 285–302. https://doi.org/10.1108/ECAM-09-2023-0973

Jacques de Sousa, L., Poças Martins, J., Sanhudo, L., & Santos Baptista, J. (2024). Automation of text document classification in the budgeting phase of the Construction process: a Systematic Literature Review. *Construction Innovation*, *24*(7), 292–318. https://doi.org/10.1108/CI-12-2022-0315

Jacques de Sousa, L., Simões, M. L., Poças Martins, J., Sanhudo, L., & Moreira da Costa, J. (2023). Statistical Descriptive Analysis of Portuguese Public Procurement Data from 2015 to 2022. *CivilEng*, *4*(3), 808–826. https://doi.org/10.3390/civileng4030045

Ministério das Finanças, I. e H. (2023, October 25). *Portaria n.º 318-B/2023 | DR*. https://diariodarepublica.pt/dr/detalhe/portaria/318-b-2023-223338607

Ministério das Finanças e Planeamento e das Infraestruturas. (2018, February 26). *Portaria n.º 57/2018 | DR*. https://diariodarepublica.pt/dr/detalhe/portaria/57-2018-114766031

Ministério das Obras Públicas, T. e C. (2008). *Decreto-Lei n.º 18/2008, de 29 de janeiro. Diário da República n.º 20/2008, Série I - Código dos Contratos Públicos - CCP | DR*. https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-lei/2008-34455475

Ministério do Planeamento e das Infraestruturas. (2017, August 31). *Decreto-Lei n.º 111-B/2017 | DR*. https://diariodarepublica.pt/dr/detalhe/decreto-lei/111-b-2017-108086621

Ministérios das Finanças e da Administração Pública e das Obras Públicas, T. e C. (2008, July 29). *Portaria n.º 701-E/2008 | DR*. https://diariodarepublica.pt/dr/detalhe/portaria/701-e-2008-575338

Noce, I., Carvalho, J. Á., Amaral, L., Cerejeira, J., Camões, P. J., Ferreira, I., & Cunha, S. (2024). A Generic Architecture for the Digitization of Government Procurement Processes. In C. S.A., K.-S.

G., & P. E. (Eds.), *ACM International Conference Proceeding Series* (pp. 190 – 200). Association for Computing Machinery. https://doi.org/10.1145/3680127.3680211

Padeiro, M., Bueno-Larraz, B., & Freitas, Â. (2021). Local governments' use of social media during the COVID-19 pandemic: The case of Portugal. *Government Information Quarterly*, *38*(4). https://doi.org/10.1016/j.giq.2021.101620

Parlamento. (2025). *Resultados Eleitorais*. https://www.parlamento.pt/Parlamento/Paginas/ResultadosEleitoraisAnteriores.aspx

Portuguese Republic. (2025). *Participation in public contracts - gov.pt*. https://www2.gov.pt/en/cidadaos-europeus-viajar-viver-e-fazer-negocios-em-portugal/participacao-em-contratos-publicos

Presidência do Conselho de Ministros. (2020). *Decreto-Lei n.º 10-A/2020 | DR*. https://diariodarepublica.pt/dr/detalhe/decreto-lei/10-a-2020-130243053

Riyad, B., & Laila, E. A. (2024). The Artificial Intelligence and Public Procurement. *2024 IEEE 15th International Colloquium of Logistics and Supply Chain Management, LOGISTIQUA 2024*. https://doi.org/10.1109/LOGISTIQUA61063.2024.10571429

União Europeia. (2015). *Contratação PúbliCa-Guia prático*. 1–106. https://doi.org/10.2776/303330

[This page is intentionally left blank.]

Appendices

APPENDIX A

Voyant Word Analysis

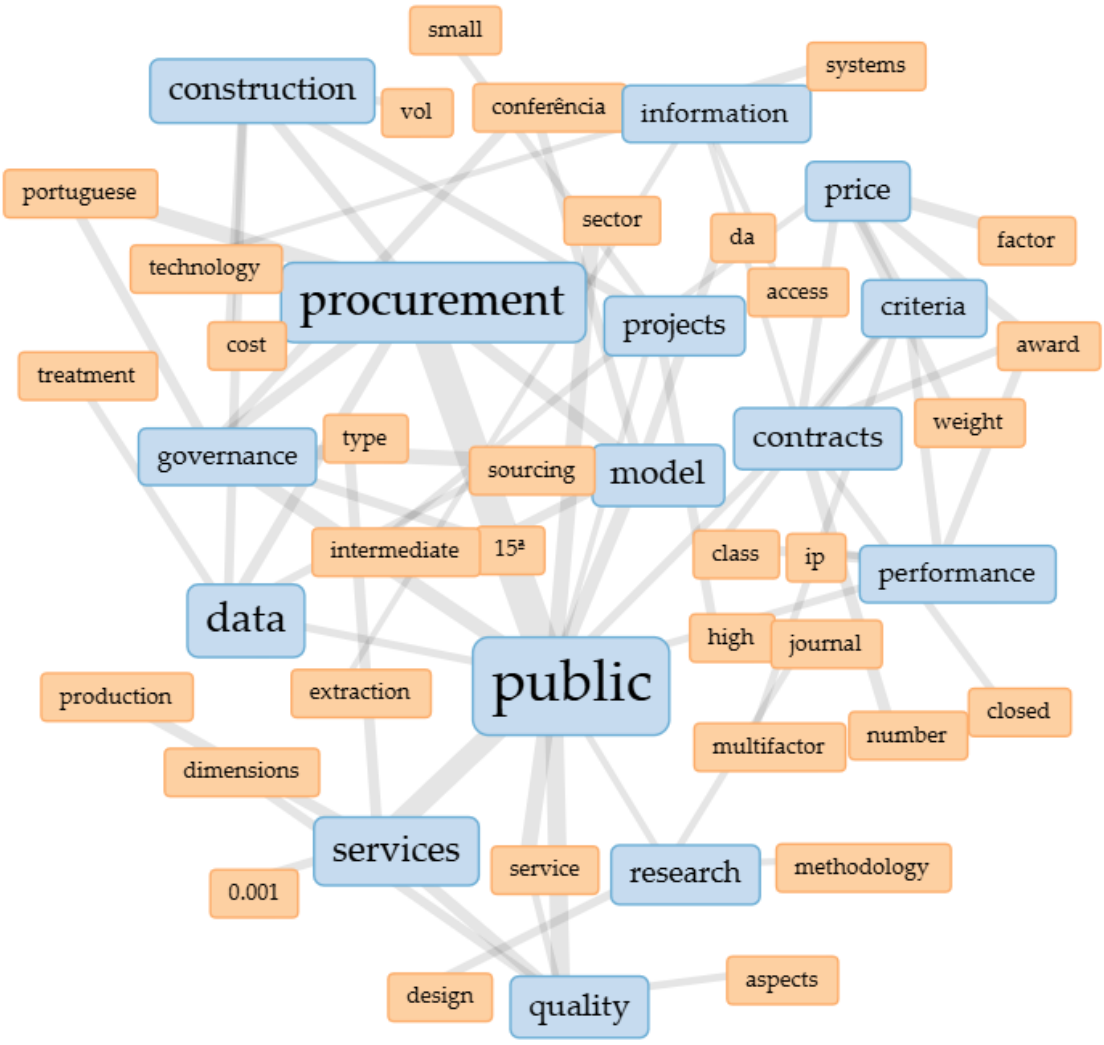
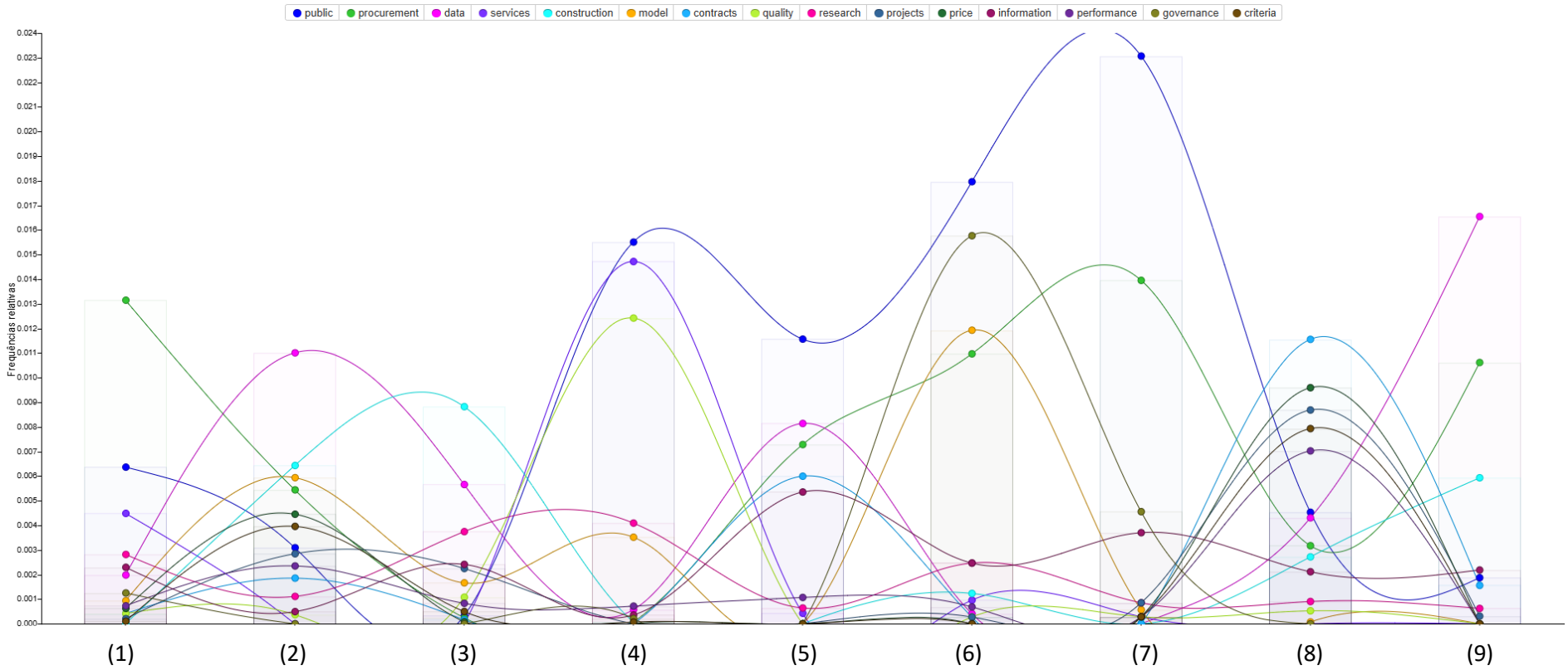


Figure 24 - Voyant Link Visualization for the Most Pertinent Documents - Word Analysis



(1) Noce et al. (2024)

(2) Jacques de Sousa, Poças Martins, & Sanhudo (2024)

(3) Jacques de Sousa, Poças Martins, Sanhudo, et al. (2024)

(4) Butkus et al. (2023)

(5) Carneiro et al. (2020)

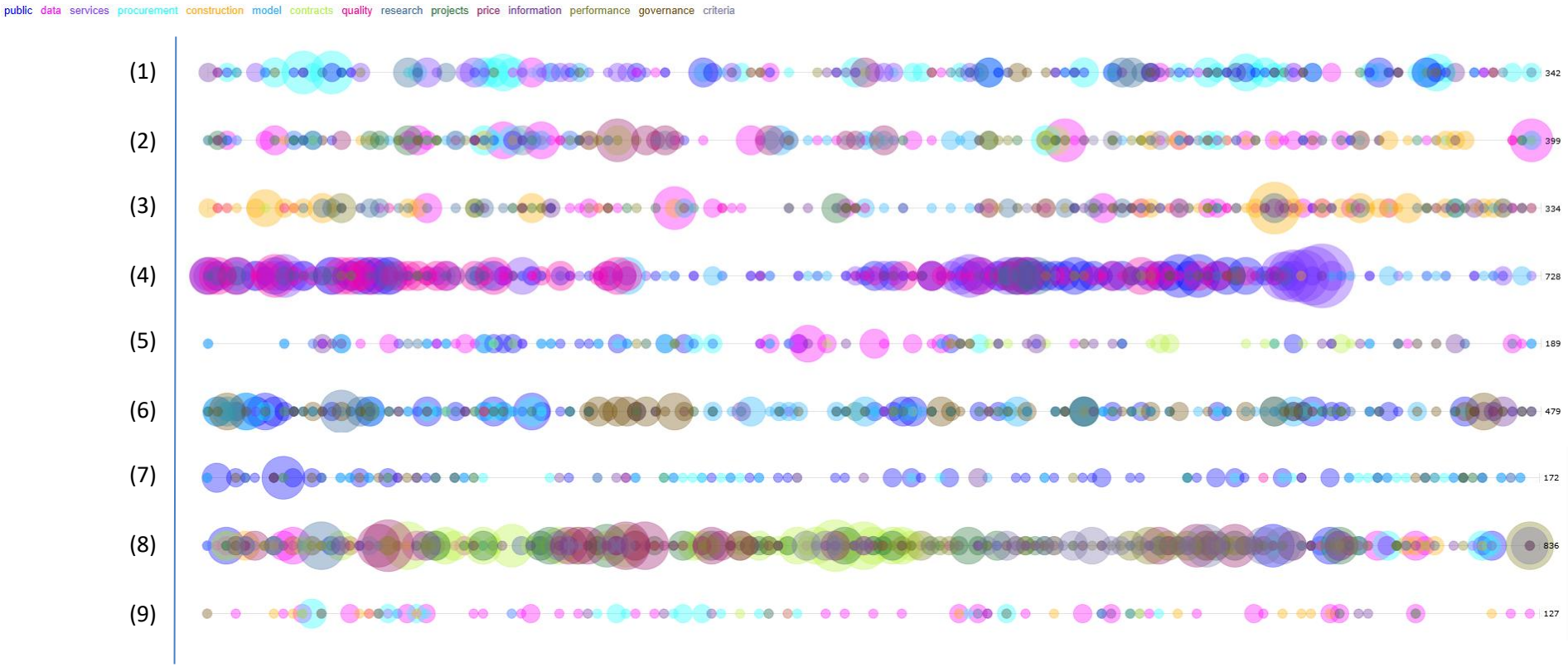
(6) Da Rosa et al. (2015)

(7) Ferreira & Amaral (2016)

(8) Jacques de Sousa, Simões, et al. (2023)

(9) Jacques de Sousa, Poças Martins, et al. (2023)

Figure 25 - Voyant Trends Visualization for the Most Pertinent Documents - Word Analysis



(1) Noce et al. (2024)

(2) Jacques de Sousa, Poças Martins, & Sanhudo (2024)

(3) Jacques de Sousa, Poças Martins, Sanhudo, et al. (2024)

(4) Butkus et al. (2023)

(5) Carneiro et al. (2020)

(6) Da Rosa et al. (2015)

(7) Ferreira & Amaral (2016)

(8) Jacques de Sousa, Simões, et al. (2023)

(9) Jacques de Sousa, Poças Martins, et al. (2023)

Figure 26 - Voyant Bubble Chart Visualization for the Most Pertinent Documents - Word Analysis

[This page is intentionally left blank.]

APPENDIX B
List of Topics Derived from CPV Codes

Table 5 - List of Topics Derived from CPV Codes

Research and development services and related consultancy services
Experimental development and research services
Research services
Services related to research laboratories
Experimental development services
Research and development consultancy services
Research consultancy services
Development consultancy services
Design and execution in research and development
Services related to public administration, defence, and social security
Services related to public administration
Community services
Emergency and fire services
Emergency services
Business services: legal, marketing, consultancy, recruitment, printing, and security
Recruitment services
Staff placement services
Personnel supply services, including temporary staff
Nursing staff supply services
Medical staff supply services
Personnel-related services excluding placement and supply
Staff training services
Staff development services
Recruitment assessment services
Investigation and security services
Search and investigation services
Education and training services
Higher education services
Youth training services

- Medical education services
- Special education services
- Training services
- Specialized training services
- Executive training services
- Health and first aid training services
- Health training services
- First aid training services
- Personal improvement training services

Health and social care services

- Health services
- Hospital and related services
- Hospital services
- Surgical hospital services
- Medical hospital services
- Gynecological hospital services
- In vitro fertilization services
- Obstetrics hospital services
- Rehabilitation hospital services
- Psychiatric hospital services
- Ortho therapy services
- Oxygen therapy services
- Pathology services
- Blood analysis services
- Bacteriological analysis services
- Hospital dialysis services
- Hospital support services
- Hospital bed linen supply services
- Outpatient care services
- Medical practice and related services
- Medical practice services
- General medical practice services
- Specialized medical services

- Gynecology or obstetrics services
- Specialized nephrology or nervous system services
- Cardiology or specialized pulmonary services
  - Cardiology services
  - Specialized pulmonary services
- Otolaryngology or audiology services
- Gastroenterology and geriatrics services
  - Gastroenterology services
  - Geriatrics services
- Psychiatry or psychology services
  - Home services for people with psychological disorders
- Ophthalmology, dermatology or orthopedics services
  - Ophthalmology services
  - Dermatology services
  - Orthopedics services
- Pediatrics or urology services
  - Pediatrics services
  - Urology services

Specialized surgical services

Dental and related services

Dental services

Orthodontics services

Surgical orthodontics services

Various health services

Services provided by medical personnel

Midwifery services

Nursing services

Home medical treatment services

Home dialysis services

		Nursing advisory services
	Paramedical services	
		Physiotherapy services
		Homeopathy services
		Hygiene services
		Home delivery of incontinence products
	Ambulance services	
	Health services with accommodation	
		Nursing services with accommodation
	Services provided by medical laboratories	
	Services provided by blood banks	
		Services provided by sperm banks
		Services provided by organ banks
	Occupational health services	
	Clinical analysis services	
	Pharmaceutical services	
Medical imaging services		
Optics services		
Acupuncture and chiropractic services		
	Acupuncture services	
	Chiropractic services	