Master Thesis

Exploratory Data Analysis of the Handelsregister to Estimate the Energy Consumption of Germany's Industry Sectors

In partial fulfillment of the requirements for the Master in Industrial Engineering at the TUM School of Computing and Information Technology of the Technical University of Munich.

Advisor Prof. Hamacher
Chair of Renewable and Sustainable Energy Systems

Submitted by María Martínez Escudero
Clemensstrasse 118
80796 Munich
marmartinesc@gmail.com

Submitted on Munich, September, 16, 2024

Abstract

The German industry sector is a major consumer of electricity and heat, significantly influencing the country's energy landscape and environmental footprint. The aim of this thesis is to evaluate whether the electricity and heat consumption patterns within this sector can be estimated using mainly the Handelsregister, as it is an openly available dataset of Germany's industrial and non-industrial companies.

The methodology applied integrates data analysis to use energy consumption datasets available, to then identify the sources and locations of energy consumption in Germany. This is to serve as a means to evaluate whether these locations' energy consumption can be correlated with the financial data available in Germany's commercial register, the Handelsregister. Thus giving insight into the extent and quality of openly available information of this source.

The findings are reflected in various formats and are thoroughly analyzed, in order to identify whether financial data can correlate with energy consumption as a whole, or exclusively to certain industry sectors or company types.

Despite these insights, the thesis also identifies a lack of comprehensive data, leading to the use of estimations to fill information gaps. The accuracy of these estimations is critically evaluated by comparing them with other found data sources. This approach underscores the importance of improving data transparency and availability to enhance the precision of energy consumption analyses.

All in all, this thesis contributes to identifying whether financial data of energy-intensive industrial companies can correlate with energy consumption, presenting a data source which can provide additional information to be used for optimizing energy use and supporting the transition towards a more sustainable industrial sector.

Key words - **Energy, industry, electricity, heat consumption, ...**

Statement of Academic Integrity

I,

Last name: Martínez Escudero
First name: María
ID No.: 03785279

hereby confirm that the attached thesis,

Exploratory Data Analysis of the
Handelsregister to Estimate the Energy
Consumption of Germany's Industry Sectors

was written independently by me without the use of any sources or aids beyond those cited, and all passages and ideas taken from other sources are indicated in the text and given the corresponding citation.
I confirm to respect the "Code of Conduct for Safeguarding Good Academic Practice and Procedures in Cases of Academic Misconduct at Technische Universität München, 2015", as can be read on the website of the Equal Opportunity Office of TUM.
Tools provided by the chair and its staff, such as models or programs, are also listed. These tools are property of the institute or of the individual staff member. I will not use them for any work beyond the attached thesis or make them available to third parties.

I agree to the further use of my work and its results (including programs produced and methods used) for research and instructional purposes.

I have not previously submitted this thesis for academic credit.

Munich, September, 16, 2024

Declaration for the Transfer of the Thesis

I agree to the transfer of this thesis to:

- Students currently or in future writing their thesis at the chair:
  - ☒ Flat rate by employees
  - ☐ Only after particular prior consultation.
- Present or future employees at the chair
  - ☒ Flat rate by employees
  - ☐ Only after particular prior consultation.

My copyright and personal right of use remain unaffected.

Munich, September, 16, 2024 Maria

Contents

List of Figures

Chapter 1

Introduction

The energy consumption of the industrial sector in Germany represents almost a third of its final energy demand [1], and causes therefore a high impact in the country's economy and environment. The current climatic situation has lead to set national and international climate goals which promote actions to be taken towards energy transition and improving energy efficiency in industrial sectors. Nevertheless, specific and detailed data on the energy consumption patterns of individual sectors is limited. Public and freely available databases such as Eurostat [1] and the Statistisches Bundesamt [2] provide broad overviews of energy consumption, but there is still a lack of granular data, particularly on electricity and heat consumption by specific companies of the different industry sectors. This thesis aims to analyze the data available and attempt to fill this gap by exploring data from the Handelsregister [3], an openly available registry of Germany's industrial and non-industrial companies, and estimate whether energy consumption patterns in energy-intensive sectors can be represented by using financial data.

As a consequence of the growth tendency in industrial energy consumption expected, which as the article on "The International Energy Outlook" [4] states is in non-OECD countries an average of 1.5%/year and 0.5%/year in OECD countries, locating energy sinks and sources can be of great utility. Knowing where energy is to be intensively consumed, like where basic chemicals are manufactured, a sector which accounted for the largest consumer of delivered energy in 2012 [4], can help locate renewable energy power plants to supply the needs of industries. By sharing information and collaborating to address where energy and to what degree this is to be consumed, renewable generation power plants can be integrated into these locations and a decentralized energy system can be modeled to help contribute to energy transition.

Moreover, industries can also take advantage of data availability in order to introduce a cogeneration energy system, such as in the paper and printing industry. This sector can provide its own energy with cogeneration as well as generate excess, which could be utilized by other energy-intensive companies. This, together with recycling materials for manufacture cannot only lead to reducing energy demand, but also lead the way to achieving a circular economy. All in all, contributing to the transparency in data regarding energy consumption can lead to a better integration of renewable energies and to develop an efficient and sustainable energy system.

The primary aim of this thesis is to evaluate the extent to which a comprehensive understanding of energy consumption within Germany's industrial sectors can be achieved through the use of the financial data contained in the publicly available source which is the Handelsregister [3], additionally using other sources for validation. The specific goals of the research include:

- **Data quality and availability:** Analyzing public data sources, their extent and limitations.
- **Energy consumption estimation:** Identifying the correlation between the variations in electricity and heat consumption and the sources' capital across energy-intensive industrial sectors in Germany.
- **Geographical Distribution:** Representing the spatial distribution of energy consumption to visualize the data obtained and provide a source to identify hotspots to take into account for energy transition and infrastructure.
- **Outlook on available data**: Assessing the quality and thoroughness of the available information and suggesting improvements.

In order to approach this thesis, the first step was to collect the required data. This was done by using public reference databases and data analysis tools, as a means to get an idea of those industries where data should be obtained from in the Handelsregister. This commercial register was handled via the application of web scraping techniques, to automate repetitive search requests. Additionally, other methods such as geocoding and mapping where applied in order to access further geographical information.

As for the final steps, the results obtained were analyzed in order to evaluate the hypothesis proposed in this thesis. To do so, scatter plots were used to compare the results of the financial data extracted from the Handelsregister and the energy consumption estimated obtained from two reference data sources. This will reveal how the hypothesis is to be evaluated distinguishing between sectors and will only prove to be valid for selected industries. Moreover, the resulting new dataset created from the information obtained from the Handelsregister was collected into a database. This can be freely and publicly accessed in SQLite format and can be visualized in a map, by entering the GitHub repository attached to this research project [5].

Chapter 2

State of the art

Energy consumption within Germany's energy-intensive industries presents an unexpected gap in publicly available and specific data. While general statistics with information distinguishing between sectors and energy carriers are accessible through public sources like Eurostat [1] and the Statistisches Bundesamt [2], detailed information about individual industries is notably scarce, suggesting a lack of legal mandates to regulate transparency in this area. As a consequence, several studies have attempted to estimate energy consumption using both bottom-up and top-down approaches. These take reference on publicly available data, including sectoral databases, emissions data, and major industrial processes, to make estimations.

The existing literature underscores the importance of this thesis, as it shines the light on the common interest to provide a deeper understanding of energy consumption patterns in German industries by analysing databases and applying interactive tools like geospatial maps from statistical offices. As an example, the Statistiches Bundesamt [2], inform about regional energy use, as can be observed further in this section, in Figure 2.1. This thesis utilizes these and other foundational studies, employing a range of programming tools and data analysis techniques to further investigate and visualize energy consumption trends across Germany's industrial landscape by using Germany's publicly available data.

Literature review

From reviewing the literature published, the main objective was to find whether a public data source with detailed information about energy consumption in Germany was available. After thoroughly doing so, it has been noted that there is no openly available sources portraying the actual electricity and heat consumption in Germany, in terms of specific granular data reflecting the consumption of actual energy-intensive industries. Information can be found regarding the consumption of energy distinguishing between countries, energy carriers and the different sectors, which are mainly households, transport, industry and agriculture. Sources found representing this information are to be hereunder mentioned.

However, there is literature related to this topic, which presents an estimation on energy consumption values. These follow bottom up and top down approaches, using sources of information openly available. Bottom up approaches use detailed and granular data at the micro-level as a starting point, to then aggregate it so as

Source	Information
AG Energiebilanzen [6]	Yearly energy balances for Germany, including electricity consumption by sector.
BDEW Strom-Kennzahlen [7]	Monthly power consumption data for Germany.
Statistisches Bundesamt [2]	Monthly power consumption data for Germany (official government source).
ENTSO-E Transparency Platform [8]	Hourly load data, actual total load for European countries.
Eurostat [1]	Monthly electricity consumption data for European countries (nrg_105m dataset).
IEA Monthly Electricity Statistics [9]	Monthly electricity consumption data for OECD member countries.

Table 2.1 Public Data Sources for Energy Consumption in Germany and Europe

to provide a comprehensive overview. This method involves collecting and analyzing data from individual units or processes, such as specific industrial plants, equipment, or sub-sectors. Whereas, a top-down approach in energy consumption analysis starts with macro-level data and disaggregates it to infer insights about smaller units or processes. This typically involves using aggregate data from national or sectoral statistics and breaking it down into components. Applying these, they provide information about energy consumption based on calculations and assumptions where sectoral databases, pollutant's emissions and main energy consuming processes are taken as a starting point. Some of the information used by these sources is easily accessible to all users. while more specific data related to individual companies from distinct sectors is not freely available. This suggests that the issue studied in this thesis is of interest, as companies have taken time and resources to elaborate detailed databases with information of companies from the different sectors in the industry, such as in the chemical, paper, refineries, metal and mineral industries.

Another approach to find more information about the research topic was to explore the different energy-intensive sectors, and verify the data availability. As exposed in Table 2.2. even though records of production, finances and general data about companies of the German industry is registered, this information is either not openly available to the entire public or is only accessible when paying for a fee. Therefore, the sources shown were not accessed to, limiting the detail of the database devised in this thesis.

Moreover, other researchers and institutions, such as Maximilian Schumacher and Lion Hirth [19] have approached the energy consumption analysis via the combination of different institutions which publish statistical data related to energy consumption. By trying to merge the data, they also evaluate its precision and the variation between the content of the different public sources.

Industry	Problems
Steel and Iron Industry	VDEH Steel Industry Portal [10]: Does not openly present data about companies. They give links to webpages of companies. Only contains a limited number of companies which are part of the VDEH association. Plant Facts [11]: State of the art online platform with company and production information.
Cement Industry	Global Cement Industry Portal [12]: It is not of free access. VDZ Cement Industry Portal [13]: Does not openly present data about companies. They give links to webpages of companies. Only contains a limited number of companies which are part of the VDZ association .
Glass Industry	Glass Global Plant Data [14]: Financials from global players, production data, demoscopic data for countries as well as import and export data. It presents a forecast about future glass consumption in all global regions. Apparently of free access, but requires a registration. Research Germany: Glass and Ceramics [15]:It is not of free access. Contains information about production of glass and ceramics companies in Germany.
Paper and Printing Industry	Fastmarkets Forest Products [16]: Price data, news, and analysis for the forest products market.
Chemical Industry	Euro Chlor Industry Portal [17]: Financials from global players, production data, demoscopic data for countries as well as import and export data. It presents a forecast about future glass consumption in all global regions. Apparently of free access, but requires a registration. Chlorine Industry Review [18]: Contains information about production of chlorine companies in Germany.

Table 2.2 Sources and problems for different industries

During the prior research to the start of the thesis, it was of interest to find an interactive map on Energy consumption in the German industry in 2018, by energy carrier [2] developed by the statistical offices of Germany [2]. This provides a comprehensive analysis of energy consumption patterns in industrial areas across the country and a detailed geographical breakdown of energy usage by administrative district. It permits the understanding of regional variations in industrial energy consumption distinguishing between energy carriers. From the map, it is revealed that natural gas is the predominant energy source for industrial plants, constituting approximately 30% of total energy consumption, followed by electricity at 22%, mineral oils and their derivatives at 16%, and coal at 15%. This resource provides a deeper understanding of regional energy use patterns and informing strategies for optimizing energy consumption within Germany's industrial sector.

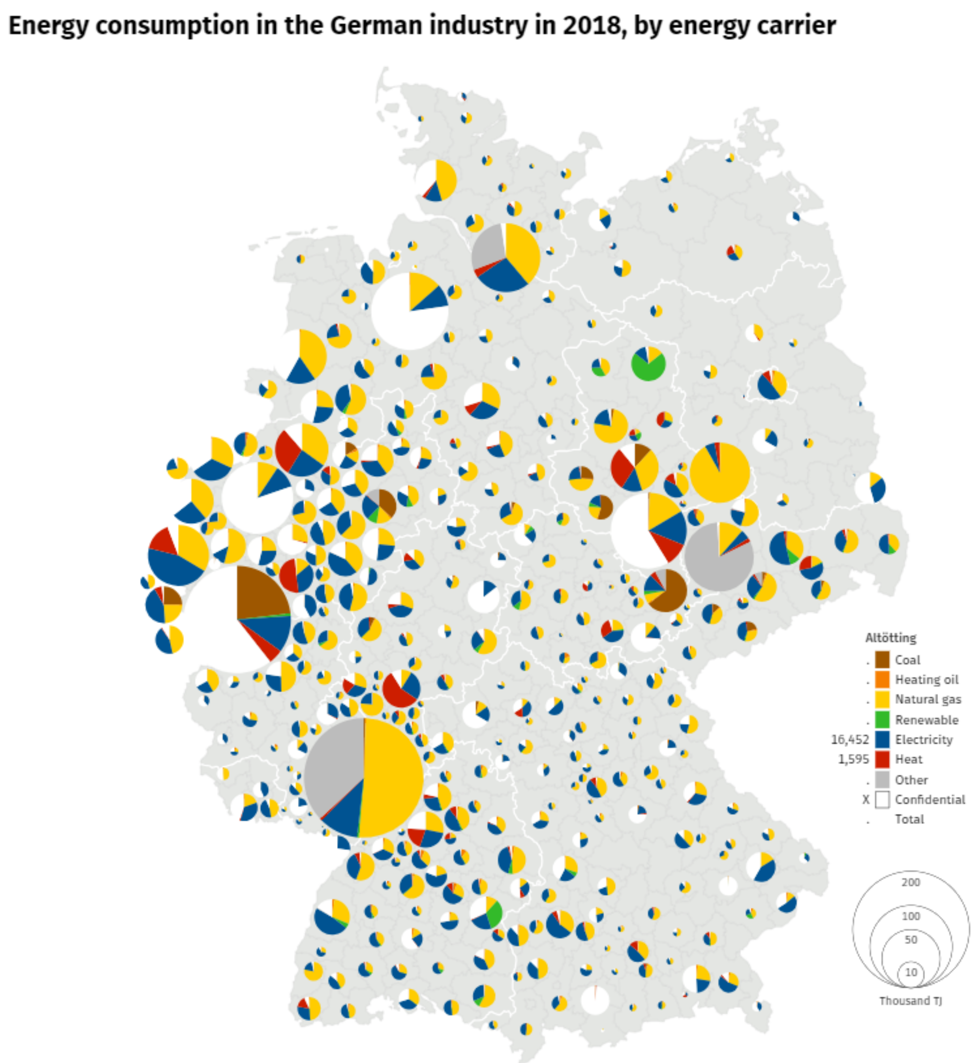


Figure 2.1 Energy consumption in the German industry in 2018, by energy carrier [2]

The main sources considered during the approach of this thesis where two databases which used some of the sources previously mentioned in this section together with top-down and bottom-up approaches, as a means to estimate energy consumption in Germany. This will be hereunder briefly described and further mentioned throughout the thesis. On the one hand, the Fraunhofer database [20] studies the hydrogen potential energy-intensive industries have. The hydrogen potential is assigned to different companies of the industry sector for its use as an energy carrier, as well as a raw material for manufacturing. On the other hand, the second database used analyzes the heat recovery potential of industries with the aim of evaluating its use in district heating applications [21].

Tools and Programs used

During the elaboration of the present thesis, many programs and tools were used, which will also be mentioned in the Methodology.

In the context of this thesis, several advanced programming tools and libraries were utilized to handle data extraction, processing, and visualization tasks. These tools are commonly employed in engineering projects for data-driven research due to their versatility and efficiency. How and when these were used will be further explained in the Methodology section and a brief insight will be given hereunder.

- **Python Programming Language and Libraries**

  Python was the primary programming language used throughout this thesis. Its use is nowadays frequent in data analysis and scientific computing applications due to its simplicity, versatility, and numerous libraries. In this study, Python enabled the integration of various tools necessary for data manipulation, automation, and visualization.

- **Nominatim and Google Maps API**

  Nominatim [22], an open-source geocoding tool based on OpenStreetMap (OSM), was employed to access to detailed information about the locations of the energy-intensive industries in Germany. It enabled to convert addresses into geographic coordinates or obtain these from coordinates. Geocoding is essential in Geographic Information Systems (GIS) and location-based services, making the Nominatim useful for mapping and spatial analysis. In the engineering field, specifically for data analysis, its utility lies on how spatial detailed data is provided, which can be then used in various applications, such as supply chain optimization, facility location planning, and environmental impact studies. Particularly advantageous is how with OSM's open-source nature and rich dataset, Nominatim has been widely adopted in academic and commercial research.

  To supplement the geocoding process, the Google Maps API [23] was used for those cases where Nominatim's data coverage was insufficient. This API proves to be a comprehensive toolset for geospatial analysis, and offers higher accuracy and more extensive global coverage. Its integration in this study allowed for additional assistance in mapping locations, contributing to validating industrial data against spatial references. However, contrary to Nominatim, the requests freely made are limited, reason for it to be used as a supplement rather than as a main source.

- **Selenium for Web Scraping**

  Selenium [24], a tool for automating web browsers, was employed to interact with web pages programmatically. This tool was employed in the extraction of data from the Handelsregister [3], Germany's commercial register. Due to its capability to automate user interactions with web pages, it offers itself as an indispensable tool for web scraping, a technique widely used in engineering and data analysis for data gathering from online sources. Web scraping allows to collect information automatically, for large datasets, helping researchers avoid manually handling repetitive tasks in various web environments.

- **Data Processing with Pandas and Regular Expressions**

  Various programming libraries were used during the data processing. Two of them are to be highlighted due to their extensive use and utility. Being one of these the Pandas library [25], which permitted data manipulation and analysis. It is a powerful Python library that facilitates handling large quatities of information with its different data structures like Series and DataFrames. Between its many possible operations, the most useful ones during the development of this thesis were data cleaning, transformation, merging, and reshaping. Pandas is widely recognized in the scientific community for its robust functionality and integration with other data analysis libraries.

  Additionally, Regular Expressions (regex) [26] were applied to take over tasks such as text manipulation and pattern matching. This tool is convenient for extracting and validating data, particularly when dealing with cases where data is inconsistently formatted. In the context of this thesis, regular expressions allowed for efficient data cleaning and pattern recognition, processes that are critical in ensuring the accuracy and reliability of the dataset.

- **Data Visualization with Plotly Express and Matplotlib**

  For visualizing the final data, the Plotly Express library [27] was used to create an interactive map that displayed the locations and information in label format of energy-intensive industries in Germany. Plotly Express is widely used due to practical and intuitive use and for its ability to generate complex visualizations with minimal code. It provides interactive capabilities which are particularly useful for presenting data in a way that is accessible and engaging to users, moreover allowing for a deeper exploration of spatial relationships and trends.

  Additionally, Matplotlib [28] was employed to create scatter plots, bar graphs and box and whiskers plots for hypothesis validation. Matplotlib is a foundational plotting library in Python, valued for its flexibility and extensive customization options. It is commonly used in engineering research for creating publication-quality plots that can effectively display complex data. In this thesis, Matplotlib facilitated the initial overview of the data obtained, as well as the validation of results by permitting the visualization of correlations between the generated dataset and other references, providing visual means to evaluate the research hypotheses.

The integration of these programming tools in this research not only demonstrates their relevance to the field of engineering but also highlights their effectiveness in helping assist with complex research challenges. Moreover, the adoption of these tools in this thesis is supported by the preparation prior to the start of the thesis with Python programming, acquired through courses and self-directed learning, which provided the necessary expertise to leverage these technologies effectively.

Chapter 3

Methodology

To achieve the objectives of the thesis, the following method was applied:

1. **Data Collection:** The Handelsregister was used as main source, supplemented with data from Fraunhofer [20] and other publicly available sources. Web scraping techniques were used to gather information where necessary.

2. **Data Analysis:** From the different data analysis tools used Python libraries like Pandas and Plotly are to be highlighted. These permitted the processing of the dataset and extract relevant insights on energy consumption patterns.

3. **Geocoding and Mapping:** Geographic Information Systems (GIS) tools such as Nominatim and Google Maps API were used to geographically situate energy-intensive industries, providing a spatial understanding of energy consumption patterns.

4. **Hypothesis Testing:** A correlation between companies' capital values and their energy consumption potential was studied using statistical methods.

5. **Visualization and Interpretation:** The final results were visualized using scatter plots and an interactive map, facilitating the understanding of the data.

3.1 Data sources

The main objective of this project is to create a dataset with financial data of Germany's energy-intensive industries using the Handelsregister's[3] information as a reference and evaluating its correlation with energy consumption. During the literature review done prior to the start of the present thesis, some articles and sources were found which, even though had different objectives as the one exposed in this thesis, can be used as a validation source to evaluate the accuracy of the hypothesis and the results obtained.

3.1.1 Handelsregister [3]

This openly available portal has a record of all the legally registered companies in Germany. It contains information about the location, the register identifier and additionally provides some documents associated to each of the companies which will be of utility for the objective of this thesis. The SI (Structured Information) document is to be used as its XML format is easily readable via Python and key information can be extracted.

The files available in XML format in this collection serve as comprehensive records of company information. They follow a standardized schema that ensures consistency and clarity across different entities. These files are structured, serving certain legal, administrative, and business purposes, and encapsulate various aspects of the company data represented with the Handelsregister, which can be classified as follows:

1. **General company information**: each XML file begins with general information about the company, including its name, legal form, and location. This section typically includes.
   - Official name of the company
   - Legal structure of the company: such as a GmbH (Gesellschaft mit beschränkter Haftung or limited liability company) or an AG (Aktiengesellschaft or public limited company).
   - Location: The registered address of the company, often including the city and country.
2. **Registration and legal data**: this section provides detailed information regarding the company's registration with the relevant authorities.
   - Registration number: A unique identifier assigned by the registration court.
   - EUID: European Unique Identifier, provides the company's identity within the European context.
   - Registered court: The court where the company is registered, along with its address and contact details.
   - Registration date: The official date on which the company was registered.
   - Dates of the last entry and amendment: Dates of the most recent entries and amendments to the company's registration details.
   - Date of the current articles of association
3. **Company purpose and activities**: describes the main purpose of the company, such as asset management or specific business operations.
4. **Capital information**: Gives details regarding the company's financial structure.
   - Capital stock or share capital: The total amount of capital the company is registered with in euros.
   - Duration: The intended duration of the company's activity, often indicated as indefinite.

5. **Representation and management**: Information about the management structure and representation rules.
   - General representation rule: Specifies how the company is legally represented, including conditions under which managers or authorized representatives can act.
   - Special representation rule: Particular details about individual management authority and any special permissions granted.
6. **Key individuals**: Including the following information of key individuals listed below.
   - Names and roles: Full names and roles of individuals such as managing directors.
   - Birth dates and addresses: Necessary for legal identification.
7. **Additional information**:
   - Product information: Information about any software or systems used in the company's registration process, including the product name, manufacturer, and version.
   - Contact details of the registered court: Complete contact information of the registration court, including address and phone numbers.
8. **Technical metadata**: Metadata about the XML file itself, to ensure traceability and version control.
   - Creation timestamp of the file
   - Sender and receiver information: Identifiers of the entities involved in the creation and reception of the file.
   - Product and version information: Details about the software used to create the XML file.

As an initial approach, the Grundkapital, Stammkapital and Hafteinlage value will be read from these files. This represents the shared capital or nominal capital of a company. In other words, the value of the shares that have been issued by the company and presents a fixed amount stated in the company's documented registry.

However, some issues have been encountered while trying to access the information. This has arised the doubt of whether the Handelsregister is actually an openly-available source of data or not. This will be explained in further detail later on in the present thesis, nevertheless the main idea can be summarized with how the request restrictions encountered during the use of this public portal, have become an obstacle during the development of this exploratory project, as well as to other research conducted by other entities.

During the data analysis part of the present thesis, the database produced by Offeneregister.de[29] was of great utility as it records Handelsregister data in various and easily manipulable formats, allowing users to download complete datasets or access the data through APIs. Users can download a line-delimited JSON file or a SQLite database. Moreover, the site includes the OpenCorporates API, which integrates German commercial register data with information from other countries, as well as a SQL API that enables direct SQL queries to the German dataset, returning results in CSV or JSON format. The data is provided under the Creative Commons Attribution License 4.0, requiring users to attribute OpenCorporates when reusing the information. However, it has to be noted that during its use, some issues were encountered due to the lack of a recent update of the data. This involved encountering differences in the registered companies in the Handelsregister[3] in comparison to what appeared in the Offereregister, and how the information of the companies was at certain times different, such as location details or registry data.

3.1.2 Fraunhofer database [20]

This database consists of a record of the 371 most energy-intensive industry sites in Germany, with their corresponding hydrogen consumption potential. This database uses a bottom-up approach to make an estimation of the energy these industries require for production. They take into account the emissions of gases in Germany's different industry sectors, the processes each one of them use, as well as the temperatures to be reached during these. They also apply a top down approach to evaluate energy consumption taking as a reference the information of sectoral datasets. All in all, they combine different data sources and methods, to then evaluate the accuracy of their results, which they represent in an interactive map.

However, as their aim is to evaluate how much hydrogen could be consumed by industries, it not only takes into account the use of hydrogen as an energy carrier. For the chemical industry, refineries and for hydrogen DRI plants in the steel industry hydrogen is also used as feedstock. Therefore when using this database to validate the resulting one from this thesis, this has to be considered.

3.1.3 Heat potential [21]

The heat potential database used as a reference examines the potential for using industrial excess heat to decarbonize district heating systems in the European Union (EU-27) and the UK. In the paper "Decarbonizing District Heating in EU-27 + UK: How Much Excess Heat Is Available from Industrial Sites?", their methodology is explained in detail as well as how unexploited excess heat from energy-intensive industries , such as steel, cement, and chemicals, currently is, leaving it to be released into the environment.

A georeferenced database was developed including data on annual production and excess heat potentials from industrial sites. They estimated excess heat potential by calculating the energy released through exhaust gases during industrial processes. This was evaluated at different temperature levels, considering how much could be captured and used in district heating systems. The industrial sites were then matched to current and potential district heating areas within a 10 km radius. Some of the sources of information used were, similarly to the ones used for the Fraunhofer Hydrogen potential database [20], the European Pollutant Release and Transfer Register (E-PRTR), the EU Emissions Trading System (EU ETS) and industry-specific databases.

3.2 Hypothesis

The approach which this thesis presents is based on addressing a direct correlation between the capital values of companies in energy-intensive industries and their size, production scale, and energy consumption. This is a wide area of research, as understanding the dynamics between a company's financial structure and its operational scale can provide insights into broader economic and environmental impacts. The capital values are reflective of a company's financial foundation, and it is reasonable to associate larger capital reserves with extensive operations, which in turn could lead to higher energy consumption.

After conducting research prior to the selection of this thesis approach, sources have been found which indicate that there is a positive relationship between a company's capital and its energy consumption, suggesting that as companies grow in size and production capacity, their energy needs increase [30]. This is also consistent with the economic principle that energy is a fundamental input for economic activity. As firms expand, their energy requirements typically rise to support increased productivity and profitability. Moreover, investments in energy infrastructure and consumption have been linked to macroeconomic growth, reinforcing the potential validity of the hypothesis employed for the present project. [31] [32] .

However, it's also important to consider the role of technology and human capital in this equation. Technological advancements, optimization of production processes and reducing waste can lead to a decrease in energy intensity, even as a company grows [33].

Taking these considerations into account as future improvements, not only the capital values but also the technological and human capital aspects of these companies could be examined. This multivariable approach could provide a better understanding of the factors influencing energy consumption in energy-intensive industries. However, with the information available in the Handelsregister and in other public sources as a whole, this information cannot be defined, being necessary then, other sources which tend to not be freely available.

For a better understanding of the different capital values associated to the energy-intensive industries in the Handelsregister, this will be defined hereunder.

- **Grundkapital** Grundkapital, or share capital in English, is a term used in German-speaking countries to refer to the capital that shareholders commit to a company by purchasing shares. It represents the nominal value of all issued shares of a company and is a fixed amount that cannot be altered by business profits or losses. The minimum amount of Grundkapital required for an Aktiengesellschaft (AG), which is a type of public company in Germany, is €50,000. These types of companies use Grundkapital as a means of securing creditor confidence and ensuring that there is an economically safe amount of funds to protect against insolvency. It represents a key component of a company's equity and is listed on the balance sheet as 'gezeichnetes Kapital' or 'subscribed capital'. [34] [35]

- **Stammkapital** Stammkapital, known as share capital in English, is a term used in German corporate law referring to the fixed equity capital of a company, particularly in limited liability companies, or GmbH (Gesellschaft mit beschränkter Haftung). Representing the total contributions made by the shareholders, it serves as a guarantee to creditors for the company's liabilities up to the amount of the capital. The minimum required Stammkapital for a GmbH is €25,000, which must be declared in the company's articles of association and fully paid up to register the company. This capital contribution is crucial and limits the shareholders' liability to the assets within the company, protecting their personal assets. Another legal form which uses Stammkapital is the Unternehmergesellschaft (UG). Also known as a 'Mini GmbH' or '1-Euro GmbH', it has a minimum Stammkapital value of €1. However, it is obligated to allocate a portion of its profits to reserve until the Stammkapital reaches the standard minimum amount for a GmbH. With these regulations it is ensured that the company has sufficient funds to meet its obligations and offers a degree of security for creditors. [36] [37]

- **Hafteinlage** Hafteinlage, translated as "paid-in capital" or "equity capital," refers to the amount of capital that is invested by a partner in a company, where their liability is limited to this capital investment. This term is particularly used in the context of German commercial law, often in relation to the legal company form known as a Kommanditgesellschaft (KG). This is a type of limited partnership in Germany where Hafteinlage represents the maximum amount a limited partner (Kommanditist) is liable for to the external creditors of the company. The registered capital contribution therefore determines the limit of their financial liability and protects the personal assets of the limited partner beyond the agreed Hafteinlage. Typically, this structure is used for family-businesses, startups, and other entities that wish to have a clear distinction between the managing partners and those who contribute capital without participating in the day-to-day management of the company. [38] [39]

The values used during the approach are mainly Grundkapital or Stammkapital for each of the companies provided by the data recorded in the Handelsregister[3]. However for those companies with the Kommanditgesellschaft (KG) legal form, the Hafteinlage value is additionally provided in the resulting dataset attatched to this thesis. This will also be used as a reference and for the mapping.

In order to be able to understand and further validate the hypothesis, the events leading to an increase in capital have been studied. Between the many reasons found from various sources [40] [41] [42], the main ones are described hereunder.

- Raising additional funds: When aspiring for expansion, research and development, acquisitions, or other significant investments a business may increase their capital. This can be achieved by issuing new shares (in the case of an AG) or increasing the contribution commitments of existing or new shareholders (in the case of a GmbH), gaining financial resources without increasing debt.

- Strengthening financial stability: Higher capital can lead to increase the financial stability of the company with the enhancing of its balance sheet. It can reduce the debt-to-equity ratio, making the company reduce its risk appearance to creditors and investors, something which can also potentially improve its credit rating.

- Attracting new investors: An increase in share capital can help attract new investors. In the case of AGs, this is done by issuing new shares through a public offering or a private placement. For a GmbH, it is achieved by having new shareholders invest in the company.

- Mergers and acquisitions: Share capital can be increased when operating mergers or acquisitions, as it can contribute to an easier transaction. For example, the acquiring company may issue new shares to the shareholders of the target company as part of the acquisition deal.

- Regulatory or legal requirements: Sometimes, depending on the company type or the sector it belongs to, these are required, for example, by law, to increase their capital to meet regulatory or legal requirements.

- Conversion of Debt to Equity: As a financial strategy, a company might decide to increase its share capital by converting existing debt into equity. This would reduce the company's debt while improving its financial health.

- Retention of Profits: When profits are generated during a company's activity, they can increase their capital by converting them into share capital. This is typically done to strengthen the company's equity base and is often seen as a sign of confidence in the company's future prospects.

When it comes to raising the Hafteinlage or Haftsumme value, the concept of increasing this value has been found to be slightly different due to the type of legal form associated to these companies [43] [44]. The reasons for an increase of this value will be listed below.

- Additional contributions from existing or new limited partners:
  - The limited partners (Kommanditisten) may agree to increase their existing contributions or the Haftsumme to provide the business with more capital. This situation can be caused by the need for expansion, new investments, or financial restructuring.
  - When new limited partners become part of the KG, they contribute with capital and increase the overall Haftsumme.

- Amendment of the partnership agreement: To formally increase the Haftsumme or introduce additional contributions, a change or amendment to the partnership agreement (Gesellschaftsvertrag) must be made and agreed upon by all partners. This will be registered in the Handelsregister to be legally binding.

- Conversion of debt to equity: Similar to other company types, a KG might also convert debts owed to a creditor into a partnership contribution (Einlage), strategically increasing the capital within the partnership.

3.3 Geocoding

Nominatim is a geocoding software application designed to convert location names, addresses, and other place identifiers into geographic coordinates (latitude and longitude) and vice versa.

In the same way that addresses are searched for in Google Maps, with Nominatim queries where made with the complete or partial addresses extracted from the data sources employed. Using the OpenStreetMap database, it aid in finding the best match for the query. This process is highly efficient and offers rich, user-generated spatial data in OpenStreetMap, which includes streets, cities, landmarks, and various points of interest across the globe.

Additionally a Google Maps API service has been used to obtain more coordinates that were not available from using the Nominatim API, which relies on OpenStreetMap data and may not have access to certain locations or may lack the precision required. However to use the Google Maps API services, it was necessary to obtain an API key, a unique identifier that allows access to Google's services. This ensures that usage is tracked and billed appropriately according to the chosen pricing plan. The API key also provides a layer of security, preventing unauthorized access to the developer's Google Cloud resources.

The application of both API's used for the present thesis can be visualized in the GitHub repository [5]. This python code was used to obtain information during the research of energy intensive companies. Queries were made with the addresses registered in the Handelsregister[3] for the companies. After processing this input the corresponding geographic coordinates along with a complete address were collected in order to insert it into the dataset generated.

Once these coordinates were obtained, these could be compared to the ones appearing in the reference databases used. This was of great utility as when attempting to find coincidential company names between the Handelsregister and the other data sources, the issue of not having recently updated data arised. Some firms were found to no longer exist, had relocated or merely changed their name.

3.4 Selenium

Selenium is a powerful tool for web scraping that allows developers to interact with web pages in an automated manner. It provided a means to automating web browsers across many platforms. Actions such as clicking buttons, filling out forms, navigating through the pages, and extracting data were therefore possible, enabling to accomplish the different interactions between a conventional user and a webpage by running a code.

As it can be seen in the Github repository attatched, via Python coding, Selenium was used to automate interactions with the German commercial register website (handelsregister.de)[3]. The key idea was to search for a list of energy-intensive industries, previously generated by using different sources, and download the XML files where structured information of these companies is recorded. For further understanding, a simple explanation of how the code interacts with web pages will be herunder given.

- **Setting the WebDriver:** Initially, the code sets up a Chrome WebDriver with specific options, including setting a download directory, where the XML files were to be saved, and running in headless mode, which means that while the code was running, the actions which took place in the web browser were not visible.

- **Opening the website:** By introducing the url of the Handelsregister `https://www.handelsregister.de`, specifically to the page were 'Advanced searches' could be done, the WebDriver opened the page with which it is meant to interact.

- **Interacting with elements:** Different locator strategies were used to locate the elements on the page, such as the Name, XPATH or ID. This could be set by previously studying the html of the webpage and inspecting the elements of interest. The strategy used was to first insert the company name, then select the option to make the search looking for the exact company name and then click the search button.

- **Waiting for elements and scrolling:** The tool WebDriverWait was used to improve reliability, by considering that the webpage can take time to load and therefore show the elements of interest to be clickable before interacting with them. Moreover, scrolling was applied, to be sure to have visible those elements subjected to interaction.

- **Handling exceptions:**The common use in Python of try-except blocks was implemented to handle cases where elements might not be found or interactions might fail. When clicking on specific elements addressed with an XPATH, the code triggered file downloads, which were then saved to the specified directory.

- **Closing the browser:** Finally, after making the searches, the WebDriver is closed.

3.5 Elaboration of Dataset

Once all the relevant information was collected from the data sources, an SQLite dataset was to be generated. To do so, Python was used to extract data from the differently-formatted data sources. These presented formats such as .csv, Excel and SQLite, which could be handled by using the different tools and libraries installed in Python. Particularly useful was the pandas library. It provide to handle information in essential data structures such as Series (1-dimensional) and DataFrame (2-dimensional). Between the various applications used from this library, to be highlighted are its handling of structured data through functionalities such as data cleaning, transformation, merging, and reshaping. Additionally, importing/exporting data from the various formats of the data sources could be achieved. Moreover, Pandas were used to integrate scientific libraries like NumPy and Matplotlib, enhancing its capabilities for statistical analysis and data visualization.

Regular expressions were also applied for tasks such as text search and replacement, input validation, and information extraction. Additionally, they have been used for this thesis project to match both simple and complex patterns, requiring in one line what would take many lines of code without their use.

In the GitHub repository [5] attached, the combined use of these tools can be visualized in detail.

Additionally, in order to deal with issues when merging the data from the different data sources, some modifications were necessary. Four simple approaches were used in order to deal with the cases where, through the automated data analysis achieved with Python coding, there where some incongruences.

1. Summing up potentials values: There were some cases where different potential values where associated to a single site registered in the Handelsregister . It was assumed that this was as a consequence of having different production buildings within the same company area. In other cases, the reason lied in the fact that two companies from the reference databases, where associated to a single one in the Handelsregister. [3]

2. Distribution of potential: whenever dealing with the case of having one single company having a match with two or more subsidiaries of the same company in the Handelsregister [3], the approach used was to divide the potential between those. However, it was not considered accurate to do this distribution equally between subsidiaries, as it is probable that each one contributes differently to the potential consumption, when evaluating hydrogen potential, or to the heat recovery potential associated to the site. Therefore, the potential was distributed proportionally to the capital of each subsidiary. The mathematical procedure can be hereunder visualized:

$$\text{Distributed potential} = \frac{\text{Capital of subsidiary} \times \text{Potential of company}}{\text{Sum of capitals of the subsidiaries}} \quad (3.1)$$

3. Manual fitting: Whenever an incongruence in a match, or a match seemed to have been missed out, entries were added or deleted manually from the database. Some cases where this took place were, for example, when a company had recently changed its company name. When this happened, the automated approach may have not found a match, when actually the same company appeared in both the reference database and the Handelsregister [3].

4. Merging administrative subsidiaries: During the automatic search of companies in the Handelsregister's database, administrative subsidiaries belonging to a company dedicated to the energy-intensive sectors defined were found. As these also reflect the size and economic state of the manufacturing company itself, the capital associated to the administrative subsidiaries was added to the capital of the main company. Thus, resulting in a single data entry for a company, reflecting its capital without having data entries with hydrogen potentials or heat recovery potentials associated to companies dedicated to sales or real estate management.

3.6 Mapping companies

Finally, after generating the dataset, an interactive map was used to visually represent where the energy-intentsive industries are located in Germany. By using the Plotly Express library, a high-level interface for creating visualizations quickly and with minimal code, companies were mapped, distinguishing the sectors with different colours. Moreover, the labels were generated so that the circle size was proportional to the size of the capital value of the company represented. Users can easily place their mouse on top of the dots and read the labels where the name, subsector and capital value of the company appears.

3.7 Validation

To evaluate the hypothesis used for the present thesis, the reference data sources were used. This was done by comparing the energy potential values assigned by the Hydrogen potential[20] and Heat Potential[21] databases with the capital value obtained from the Handelsregister[3].

To do so, the library Matplotlib was of great utility. This is a popular plotting library in Python used for creating static, animated, and interactive visualizations. It was employed to generate a variety of plots and charts, and adjusts plot elements like labels, titles, colors, and styles in an easy way. Firstly it was used to generate bar graphs and box and whiskers graphs to give an initial overview of the data. Further on, it was used for generating scatter plots, with their corresponding tendency line and R-squared value. These were created for the different industry sectors from the data present in the generated dataset of the project, as it can be seen in the Github repository attached [5].

Chapter 4

Results

As a result from the information compiled from the different data sources, the final dataset contains those companies which appeared not only in the Handelsregister [3], implying that they are currently registered as active businesses, but also in the dataset provided by the Fraunhofer H2-potential database [20] and the heat excess production database [21].

4.1 Database obtained

The final database in SQLite format is constructed by integrating information from three primary sources: the Handelsregister (Commercial Register) and two publicly available energy consumption databases. This combination provides a comprehensive view of energy-intensive industries. The details of this integration are as follows:

1. **Data from the Handelsregister:**
   - **Location:** Geographic addresses of companies, including street, city, and postal code, which are essential for further geocoding, mapping and spatial analysis.
   - **Company names:** Official names of companies, facilitating the identification and verification of entities within the dataset.
   - **Capital values:** Financial data, in the form of capital data labelled as Grundkapital, Stammkapital or Hafteinlage, providing insights into the economic scale and financial strength of each company.
2. **Energy consumption data from publicly available internet databases:**
   - **Energy usage:** Quantitative data on the energy consumption of various industries or specific companies, estimated from collecting data from a range of sources.
   - **Industry classification:** Categorization of companies based on industry type, which will enable a sector-specific validation of the data.

By merging the data from the Handelsregister with the energy consumption information, the final database offers a detailed and integrated view of energy-intensive industries. The Handelsregister data provides foundational details such as company location, name, and financial status, while the energy consumption data adds specific insights into the energy usage patterns of these industries. This approach will provide a more accurate analysis and evaluation regarding energy consumption and industry impact.

4.2 Overview of the data

Once the resulting dataset was obtained, the reference databases evaluating the heat recovery potential and hydrogen potential of German industries, are to serve once again as reference and will contribute to evaluate results as well as to determine the validity of the hypothesis of this thesis. The number of companies obtained for the different sectors established for each of the databases are represented in 4.1.

Subsector	Hydrogen Potential Database	Heat Recovery Potentials
Metal Processing	19	-
Steel, primary	5	-
Iron and steel	-	18
Chemical Industry	34	52
Non-ferrous metals	2	2
Non-metallic minerals	11	22
Mineral Processing	23	-
Paper and printing	74	84

Table 4.1 Number of Companies Obtained for Hydrogen and Heat Recovery Potential Reference Databases by Sector

The different databases used as references, have a different sectoral distribution, reason for which for some sectors there is a '-' corresponding to the number of companies of the particular sector which was in only one of the databases. On the one hand, the hydrogen potential database [20], will have in the Metal Industry, three different sectors: 'Metal Processing', for processes such as casting and rolling of metals, 'Steel, primary' and 'Non-ferrous metals', to primarily represent the aluminum production. For which, it is notable to add that 'Steel, primary' will not take part in the validation process, for its hydrogen potential having included hydrogen as raw material rather than as an energy carrier. In contrast, the alternative reference database used will only have 'Iron and steel', similar to the 'Metal Processing' sector from the reference previously mentioned, and 'Non-ferrous metals', also for aluminum processing, to represent the Metal Industry.

Moreover, for hydrogen potential evaluation, the Mineral Industry will distinguish the 'Mineral Processing' sector, for glass production, and the 'Non-metallic minerals', for cement and lime, while the heat recovery potential database will encapsulate it in the Non-metallic minerals sector.

In order to understand the different potential and capital orders of magnitude assigned to each sector, maximum and the range of potential and capital values are represented hereunder.

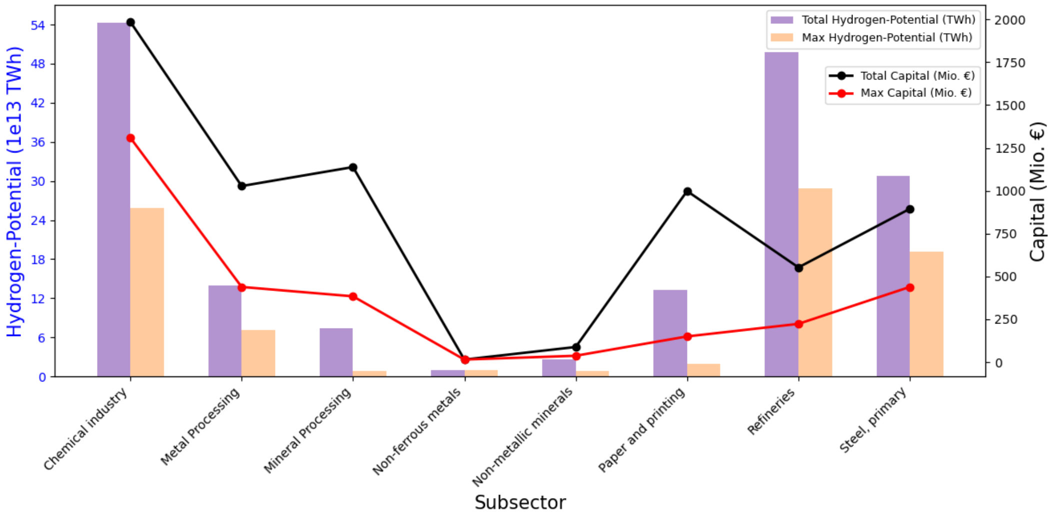


Figure 4.1 Graph Representing the Total and Maximum Values of Hydrogen-Potential and Capital for each Sector.

What can be observed from 4.1 is how the highest total potential corresponds mainly to the Chemical Industry and Refineries sectors. This can give an insight into how, due to the fact that the Fraunhofer database [20], used here as a reference, includes hydrogen potential for feedstock for the manufacture of products as well as an energy carrier, these particular sectors will have a higher assigned potential. Although this will be further analyzed throughout this chapter.

Moreover, from the disparities obtained between capital and potential across the sectors when comparing overall results, the importance of analyzing this relationship isolating the sectors is evidenced.

To further analyze variations, Figure 4.2 represents a Box and Whiskers diagram. What can be observed, the hydrogen potential varies significantly across the different sectors, with Steel, primary and Refineries displaying the highest variability, as shown by the large box sizes and long whiskers. These also present obvious outliers, indicating a few extremely high hydrogen potential values. However, sectors like Non-metallic minerals, Metal processing, and Non-ferrous metals show very low and consistent hydrogen potential, with small boxes and minimal whiskers. Overall, most sectors exhibit low variability in hydrogen potential, except for Steel, primary and Refineries, where the potential is not only higher, but with higher variability.

On the other hand, the capital distribution presents a similar the pattern to the one with the Steel, primary sector in hydrogen potential, showing the highest variability and largest range of capital investment, extending up to 800 million €, along with outliers indicating extreme values. Refineries also have a moderate capital variability, while sectors like Non-ferrous metals, Non-metallic minerals, and Metal

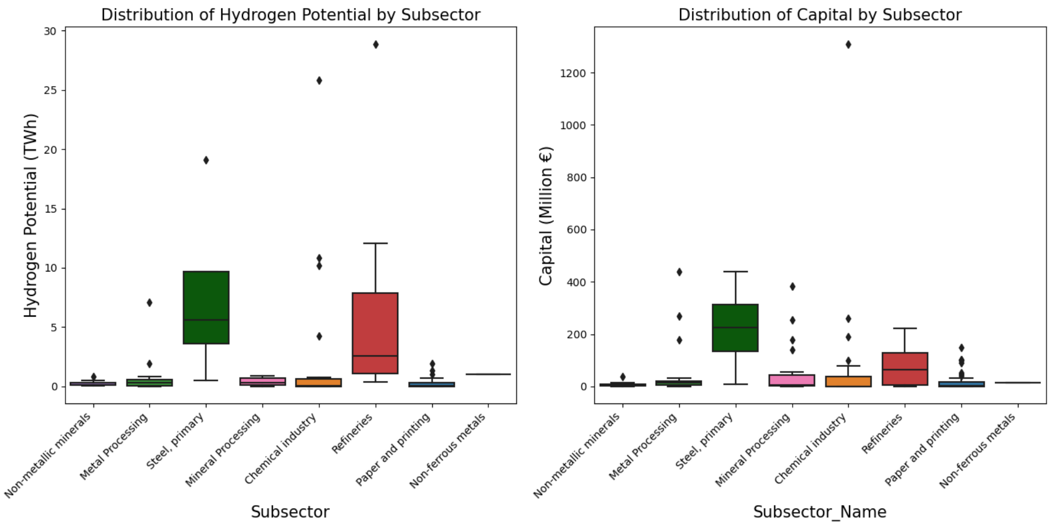


Figure 4.2 Subplots with Box and Whiskers Plots Representing the range of Hydrogen Potential (left subplot) and Capital Values (right subplot) for the different Sectors.

processing show consistently low capital values along with narrow ranges. Outliers are present across various sectors but are particularly prominent in Steel, primary and Refineries.

Regarding the results associated to the heat recovery potential reference database [21], observed in Figure 4.3, it is clear that the Non-metallic Industry and the Refineries are the leading energy-intensive sectors for which the highest heat recovery potential can be assigned. From this plot, conclusions about a relationship between capital and heat recovery potential cannot be obtained, as only the Chemical Industry, Non-metallic Mineral Industry and Refineries seem to have potentials in the scale of representation.

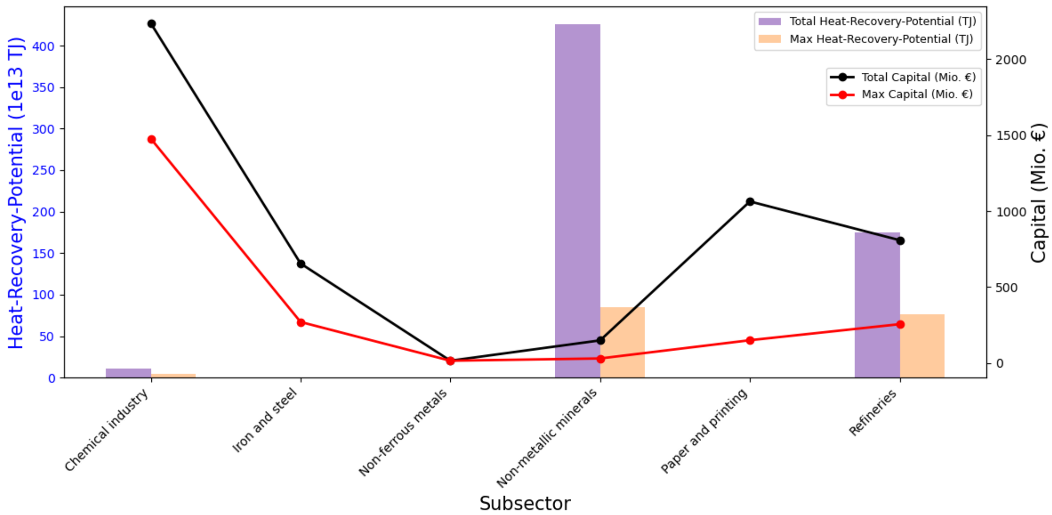


Figure 4.3 Graph Representing the Total and Maximum Values of Heat Recovery Potential and Capital for each Sector

To get a better insight into the remaining sectors, Figure 4.4 represents the Iron and Steel, Non-ferrous metals and Paper and Printing Sectors.

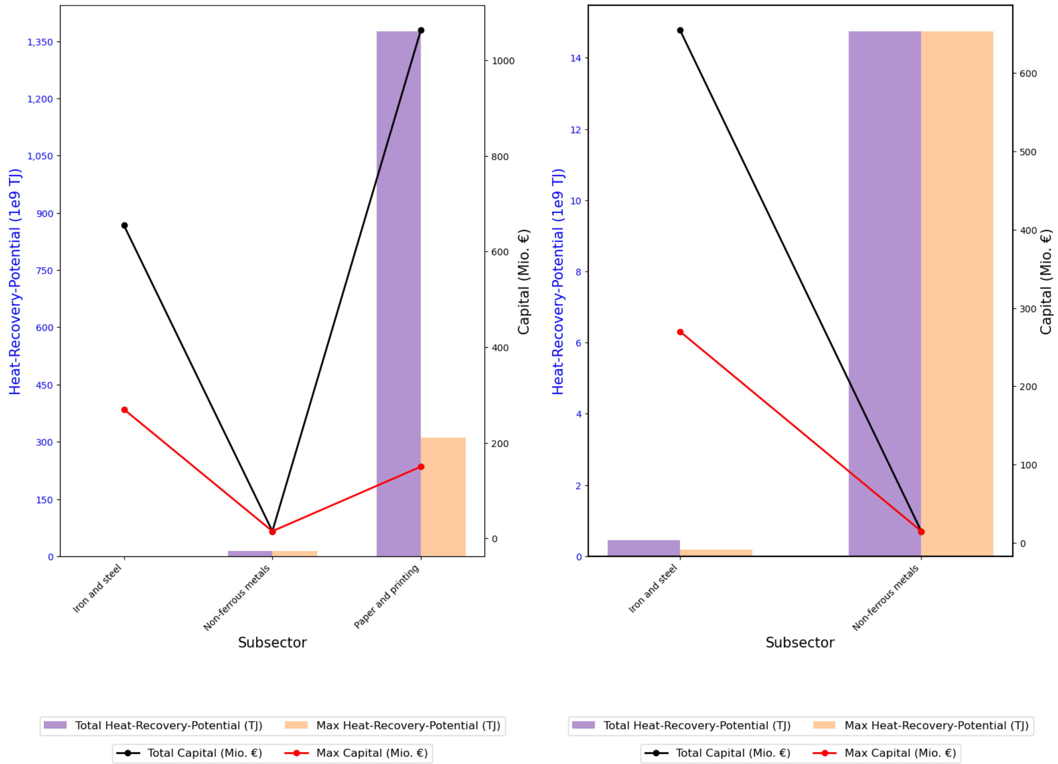


Figure 4.4 Subplots with Graphs Representing the Total and Maximum Values of Heat Recovery Potential and Capital for the Lower Potential Industry Sectors : In the first subplot (in the left side) the sectors represented are Paper and Printing, Iron and Steel and Non-Ferrous metals. In the second subplot (in the right side) the sectors represented are Iron and Steel and Non-Ferrous metals.

As it can be observed from Figure 4.4, the Iron and Steel sector still remains to have a proportionally low potential, compared to the other sectors. This can be mainly caused by the fact that most of the processes it presents generate less heat to be recovered, as, while coke ovens will account for relatively high potentials, others like the use of blast, oxygen and electric arc furnaces will comparatively provide not even half the potential. However, capital will be high for these industries, maybe corresponding to the high inversion required for the processing machinery.

From 4.5 the reader can initially observe that the distribution of heat recovery potential shows significant variability, especially in the Non-metallic minerals subsector, highlighted with the size of the box and the extent of the whiskers beyond 80 TWh, indicating a wide range of values and some outliers, something also evident from 4.3. Variation can also be observed in a smaller scale for the Iron and steel sector, with a moderately sized box and whiskers stretching up to around 20 TWh, along with a few outliers. As seen from the previous graphs, the remaining sectors, including Chemical industry, Refineries, Paper and printing, and Non-ferrous metals, have consistently low heat recovery potential, reflected by their small boxes and minimal whiskers, indicating very limited variability.

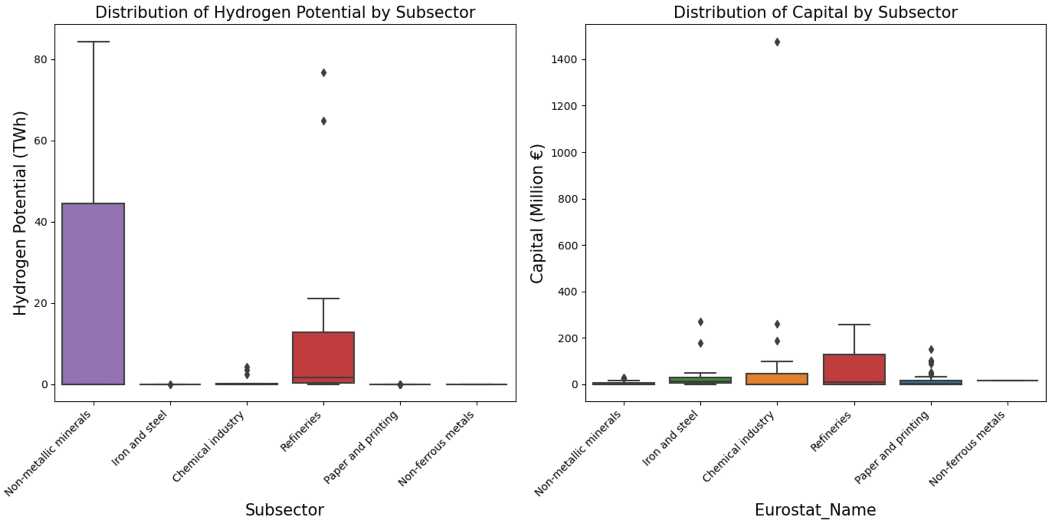


Figure 4.5 Subplots with Box and Whiskers Plots Representing the range of Heat Recovery Potential (left subplot) and Capital Values (right subplot) for the different Sectors

Capital distribution presents smaller ranges for most sectors, with Paper and printing showing the largest box, evidencing a moderate variability in capital for these companies, of up to approximately 600 million €. Outliers are present across multiple sectors, with the Chemical industry having a notable outlier of around 1,200 million €. Sectors presenting smaller-sized boxes with low capital values are Non-metallic minerals, Iron and steel, Refineries, and Non-ferrous metals. This suggests minimal variability and consistently low investment levels. The presence of outliers in sectors like Iron and steel and Refineries suggests occasional high capital investments, though the overall range remains limited.

The concepts extracted from this initial overview will be analyzed with more depth in the following section.

4.3 Plots for Validation of the Hypothesis

After the initial overview of the dataset, the sectors will be hereunder analyzed in further detail. To do so, the value of the hydrogen potential from the hydrogen potential database was compared with the capital value for each of the companies. For the heat potential database, from the different heat recuperation potential levels it offers, only the one associated with 'Level 3' was used, as generally, the different levels presented similar tendencies. Level 3 shows the heat potential when having 95°C as the reference temperature and at what in the report of the database is referred to as the 'Current situation'. This essentially means that the excess heat generated in the processes is not used for actions such as preheating, which would result in a lower potential as part of the waste heat would be internally used. [21]

To visually represent the results, scatter plots were used, associating a different color to each of the different sectors. Companies are represented with dots, situated according to the capital in euros and to either the hydrogen potential values in Watt-hour or the heat recovery potential in Joules. Depending on the company type, and its respective capital value, the colour of the dots varies. The Stammkapital, for GmbH companies, is represented by the base colour, the Grundkapital, for AG companies, is represented by a darker tone of the base colour, and the Hafteinlage, for K.G. companies, is represented by a lighter tone of the base colour. The colour scheme selected for the different subsectors is presented below:

- Red: Refineries.
- Blue: Paper and printing Industry.
- Green: Metal Industry sectors, such as Iron and Steel, Metal Processing and Non-ferrous metals.
- Orange: Chemical Industry.
- Purple: Mineral Industry sectors, such as Non-metallic minerals and Mineral processing.

Moreover, a regression line is presented below each of the plots, as well as the R-squared or coefficient of determination value. These elements will help identify whether a relationship between capital and energy is present.

4.3.1 Refineries

For refineries the Fraunhofer database [20] assigns hydrogen potential values exclusively according to how hydrogen can be used as feedstock, something which while showing a different manner of hydrogen consumption compared to the other sectors, can also give insight into the production and energy consumed for this production. On the other hand, from the second reference database, refineries are projected as a highly profitable heat recovery source, due to the combustion of large quantities of fuel gas.

- **Hydrogen potential database**

  Leaving out the outliers, the data reflected in Figure 4.6 appears to follow a trend that aligns with the proposed hypothesis. However, it is to be considered that the sector under examination comprises a limited number of companies, which may affect the robustness of the conclusions drawn. Additionally, it is important to note that in the reference database, the authors of the database, when aspiring to maximize hydrogen's future potential, propose the use of it as feedstock for crude oil refining and electrolysis. This distinction may have implications for the interpretation of the results and their applicability to broader contexts.[20]

- **Heat potential database**:

  Refineries are characterized for producing heat mainly through the combustion of refinery fuel gas. The exhaust gases are at medium temperatures (150℃ to 260℃), leading to a considerable potential for heat recovery as large amounts of fuel are used. [21]

  When taking a general overview of the data in Figure 4.7, having removed obvious outliers, a positive trend is observed, though the fit is quite weak.

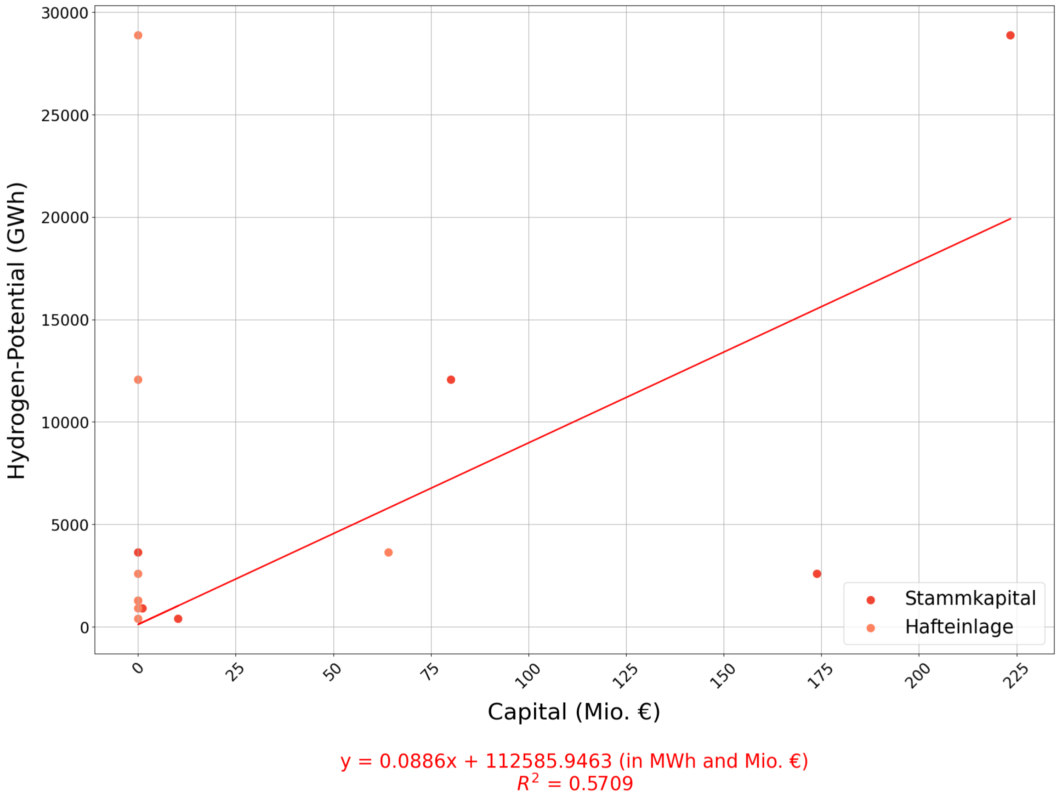


Figure 4.6 Results of the Refineries: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

This is not only because of the highest potential company of the plot, but also because of the fact that most companies of the sample seem to lie in the area of lower capitals and potentials. Therefore, it seems sensible to look into these in more detail.

However ,as it can be observed in Figure 4.8, this approach did not result in a clear trend, presenting a very weak fit to a negative relationship between capital and potential. Most of these companies seem to have their capital close to the minimum required for their specific legal form, therefore suggesting that they may invest their profits in research rather than in raising their capital, something which aligns with the global trend of companies to investigate and shift to new and environmentally friendlier fuels for transport.

In order verify the validity of this assumption, some outliers have been researched. With high potential and low capital HOLBORN Europa Raffinerie GmbH stands out the most. In line with what was previously assumed, they seem to have recently invested in energy-efficiency measures, such as green diesel and sustainable aviation fuels [45]. MiRO Mineraloelraffinerie Oberrhein GmbH & Co. KG, presents a slightly higher potential, and a significantly higher capital. The reason for it having this higher capital lies in the fact that it

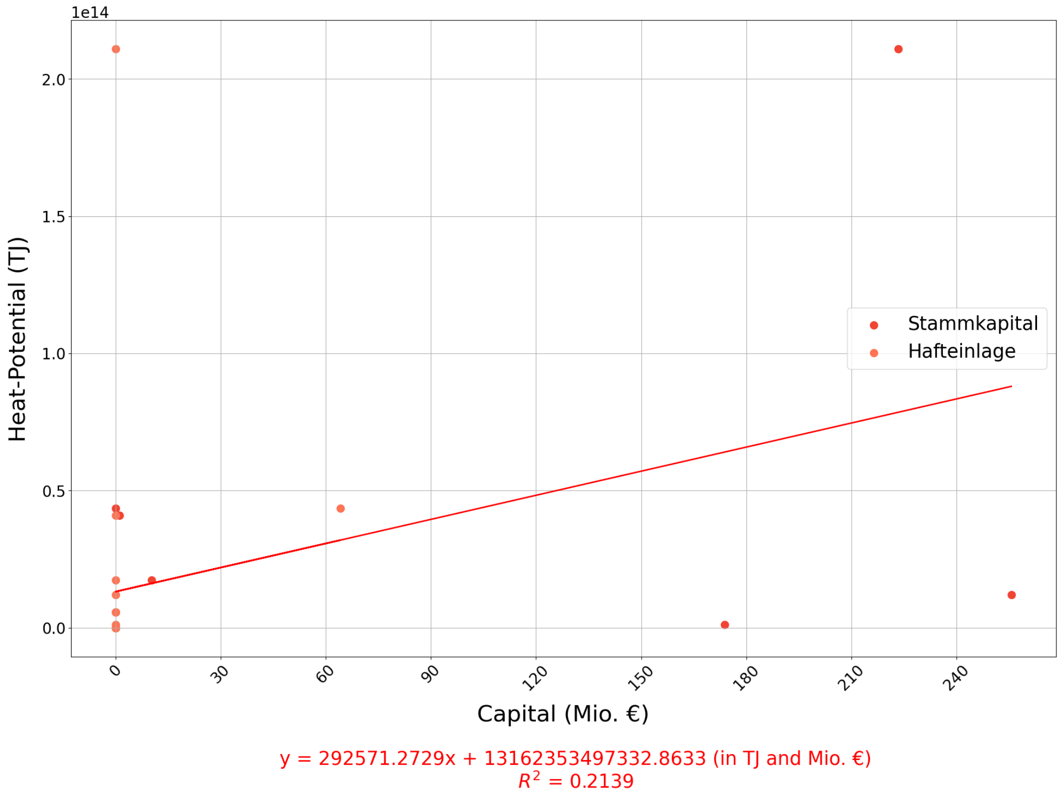


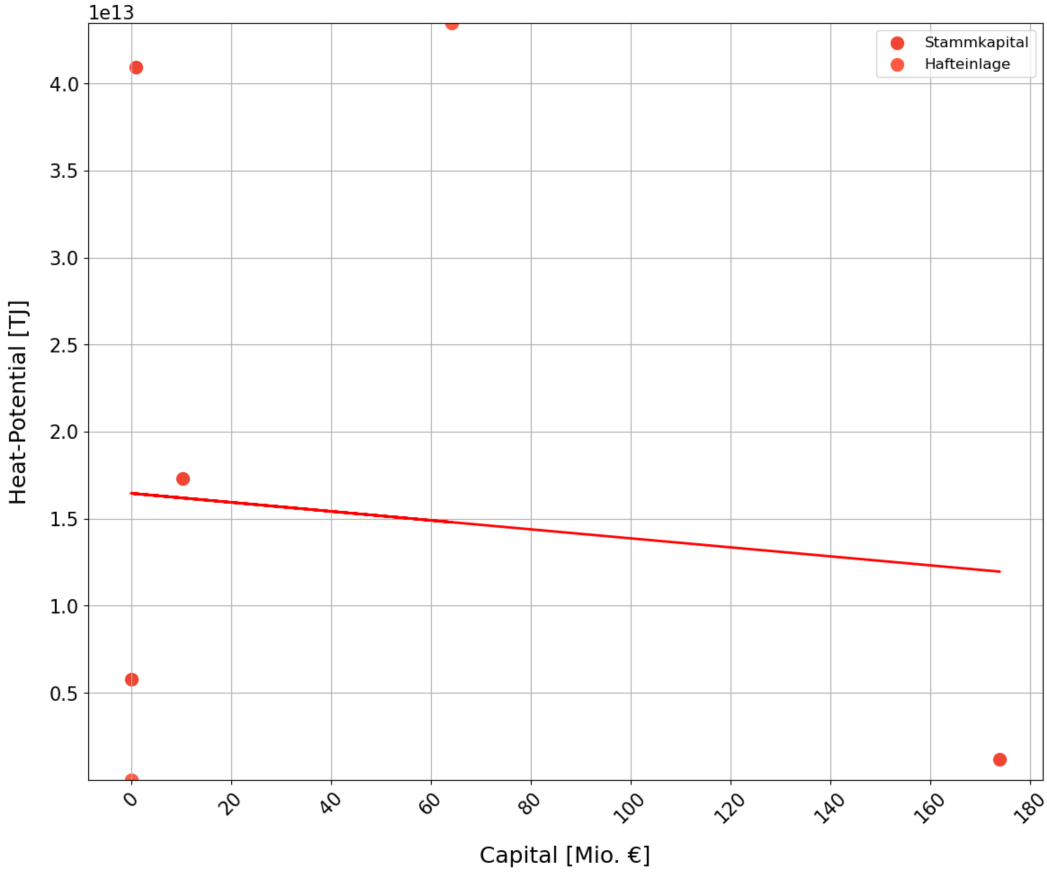
Figure 4.7 Results of the Refineries: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

is a much bigger company, with numerous shareholders, higher capacity and has products such as propylene and bitumen which HOLBORN does not [46].

Bayernoil, much like MiRO Mineraloelraffinerie Oberrhein GmbH & Co. KG however, also follows the hypothesis, in a lower scale. Paired with having lower production figures, it presents a lower capital [47].

Looking into companies with high capitals but low potentials, PCK Raffinerie has been researched and found to be one of the largest refineries in Germany, accounting 11% of Germany's total capacity. This contrasts with the fact that it presents a quite low heat recuperation potential compared to the previously mentioned companies, whose production is significantly lower. When looking at the reference database, this company does seem to provide more potential at 'Level 1', when the reference temperature is 25°C. This is probably due to the processes it undertakes, and therefore is not an adequate reference for this plot. When looking into this level in Figure 4.9, a weak trend can be seen, therefore it does not provide a more reliable validation than when looking at level 3 potential.

Therefore, from these companies, the hypothesis of the thesis seems valid for those companies which are not applying a strategy of investing in research



y = -25868.7431x + 16450382982724.1582 (in TJ and Mio. €)
R^2 = 0.0077

Figure 4.8 Results of the Refineries: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 5e13 TJ. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

and investigation, something which while it might benefit the future of the company economically, will reduce their assets and therefore their capital.

4.3.2 Paper and Printing

For both of the reference databases, this sector presents a weaker energetic hotspot, being the process undertaken mainly steam generation for paper drying and chemical pulp production. It will as a consequence account for lower potentials regarding individual sites. However, when considering the high number of industrial sites, it will considerably contribute to a total high potential.

- **Hydrogen potential database**

  The data sample of the paper industry from Figure 4.10, having removed the biggest outliers, presents a weak however positive trend. As expected, most

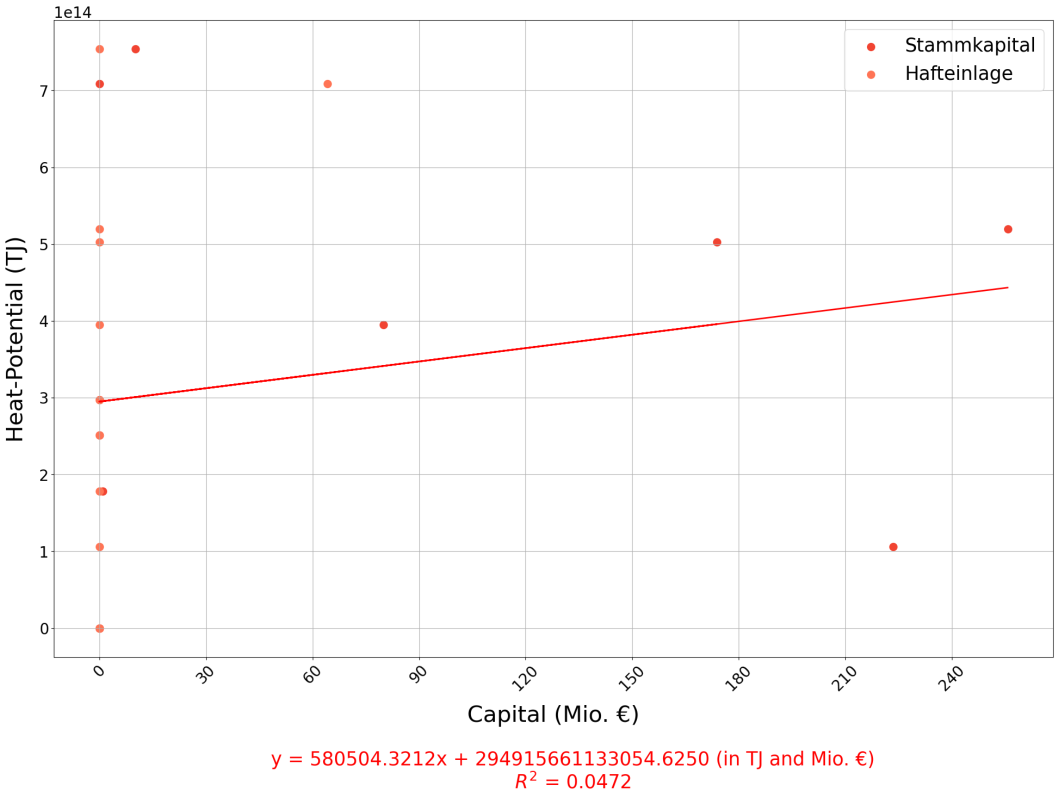


Figure 4.9 Results of the Refineries: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. The heat recovery potential is represented by a reference temperature of 25°C. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

companies are in the lower potential and capital region, reason for which this one will be hereunder analysed.

Analysing companies from figure 4.10 which contribute to a weak fit to the trend, an important outlier to be detected regarding high potential and low capital is Papierfabrik Adolf Jass Schwarza GmbH. This company is one of the leading paper recycling industries in Germany and is part of Jass Schwarza. Initially, the group started as a family-owned business in Gronau an der Leine, expanding to Fulda and later to the Schwarza facility. The fact that its Stammkapital or shared capital is 25000€, the minimum for this type of legal form, suggests it has maintained its initial shared capital, corresponding maybe to the fact that it is not the main facility and does not require therefore to increase its equity, as long as the main company has an economically stable state. They are also involved in R&D projects and may invest its benefits in these projects to improve sustainability and promote circular economy rather than in increasing capital. Moreover, being a family-owned business, it may be more adequate to maintain a low shared capital for their tax or inheritance planning purposes. Notable to be mentioned is the fact that the

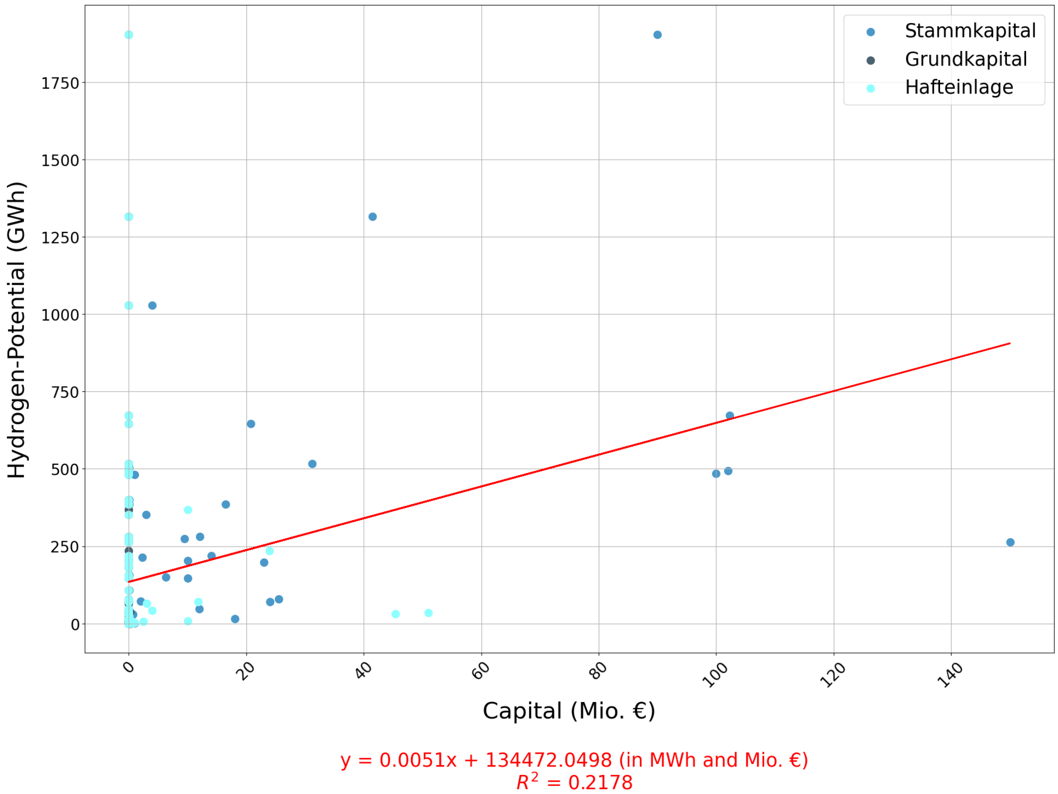


Figure 4.10 Results of the Paper and Printing industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

bigger Fulda facility is not registered in the Handelsregister, though it would have been interesting to compare the capitals of both facilities to evaluate the assumptions. [48] [49]

Observing the Smurfit Kappa Zülpich outlier, similar conclusions can be drawn. As it also a GmbH and is part of a larger group, its financial resources can be allocated and managed at the group level, allowing individual facilities to operate with lower capital. This suggests how these kind of companies invest in production and profit distribution between group members, therefore indicating that their share capital may not be the best indicator of their size or energy consumption.

Driving attention to the lower capital and lower potential companies, a large number of companies is situated in the region close to the values of lower capital and lower potential. These have been observed to be those which form part of a larger group and are primarily smaller size subsidiaries, not take such an important role in production. An example would be some WEPA group subsidiaries like WEPA GREEN GmbH and Wepa Kraftwerk GmbH.

A company with particularly high capital for its relatively low potential is KANZAN Spezialpapiere GmbH. Compared to Papierfabrik Adolf Jass Schwarza

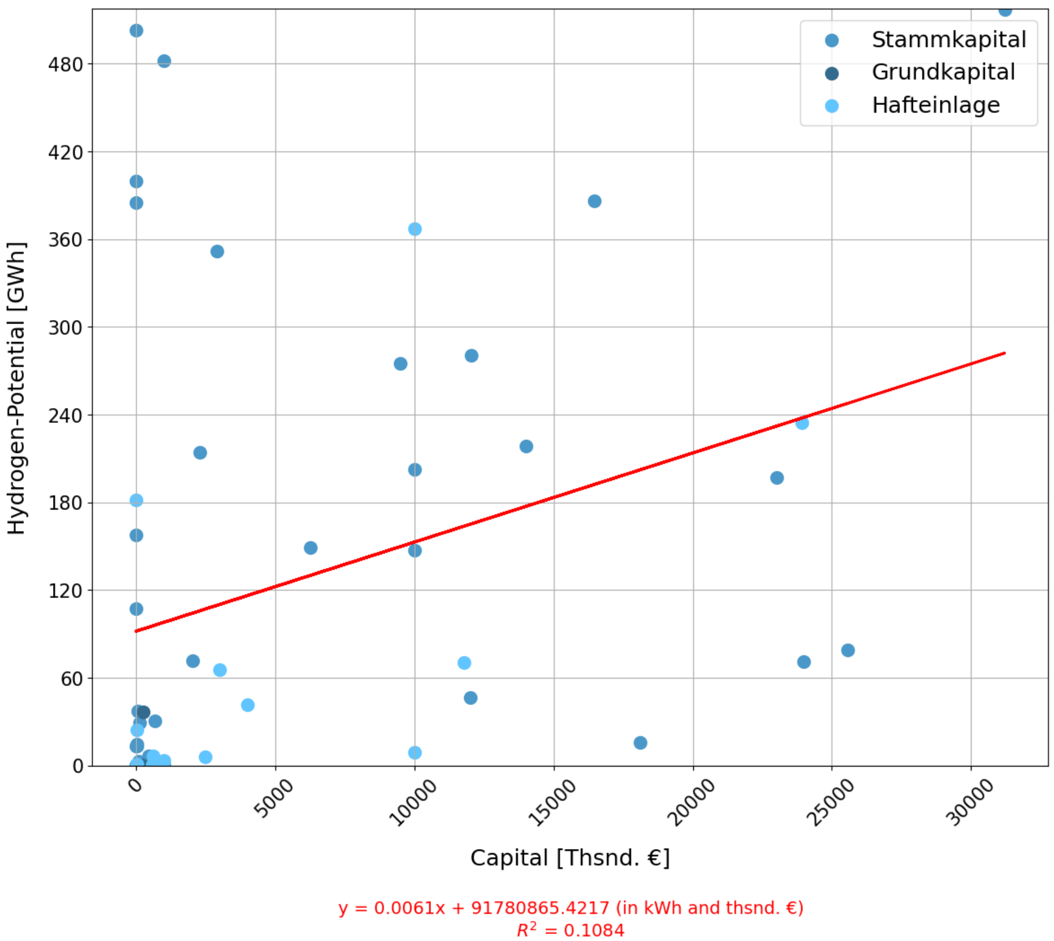


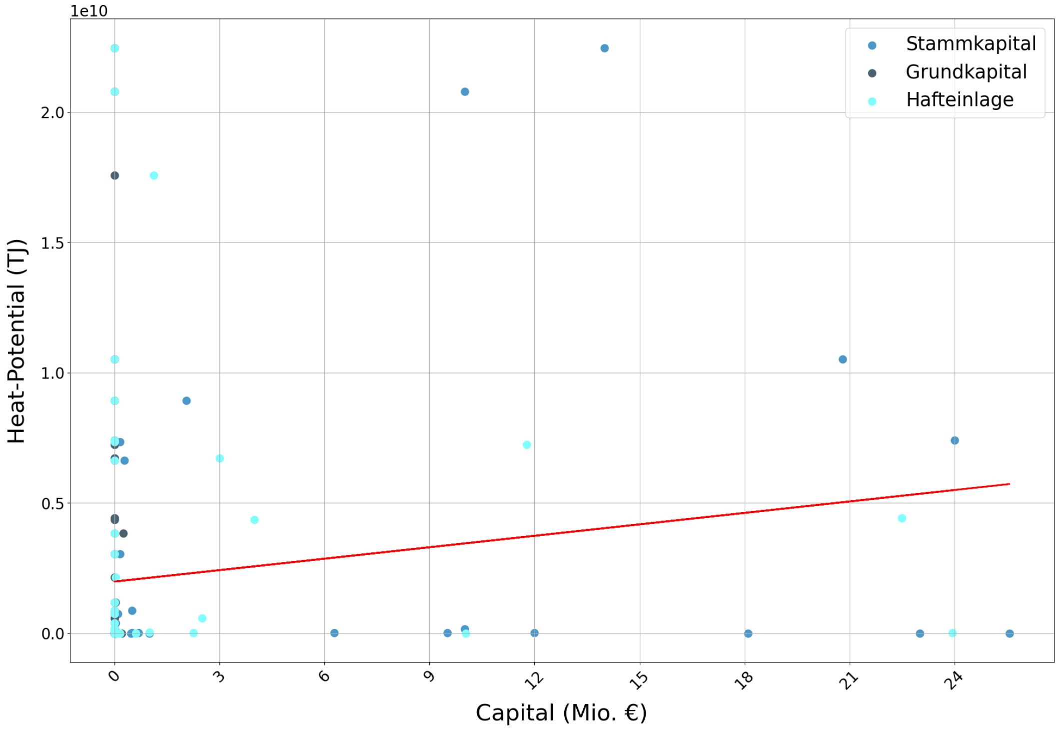
Figure 4.11 Results of the Paper and Printing industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 1 TWh. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

GmbH, it has a significantly lower paper production, corresponding to a lower potential. The reason for it having a higher capital lies in how this is the main facility from the group formed originally by a joint venture with German Zanders Feinpapiere AG, and Japanese trading company Marubeni. Therefore, it does require a higher equity. [50]

Despite the above-mentioned exceptions, a positive trend supporting the hypothesis can be observed in Figure 4.11, when not taking into account the smaller subsidiaries and other outliers. However, due to the low R-squared value, the trend is of a rather weak fit.

- **Heat potential database**

  In the paper and pulp industry sector, recuperation of heat will primarily come from boilers and kilns. The potential for recovering this heat for this process is lower due to lower working temperatures. [21]



y = 146.4564x + 1979474243.2192 (in TJ and Mio. €)
R^2 = 0.0492

Figure 4.12 Results of the Paper and Printing industry: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

As an overview of the paper and printing industry data sample, a greater number of companies are situated in the area of lower potentials, reason for which a potential threshold is set to select the sample to consider when evaluating the hypothesis.

From the plot, even though the trend seems to be positive, the R-squared value is rather low, indicating a poor fit and a high variation concerning the trend. Therefore, this sample will not present a suitable reference for the validation of the thesis' hypothesis.

It can be observed that many of the small-scale paper companies seem to have the minimum capital value when presenting the GmbH legal form. These seem to be part of groups of companies, such as Glatfelter, WEPA and Papierfabrik GmbH. Moreover, many of the companies also present the KG legal form, with which capital values presented are the equity capital values, composed by the liability amount of each of the partners involved in the business.

The reason for these results may lie in the paper industry sector. The industry can be considered to be mature and is susceptible to cyclical demand patterns and ever-increasing environmental regulations, therefore contributing to

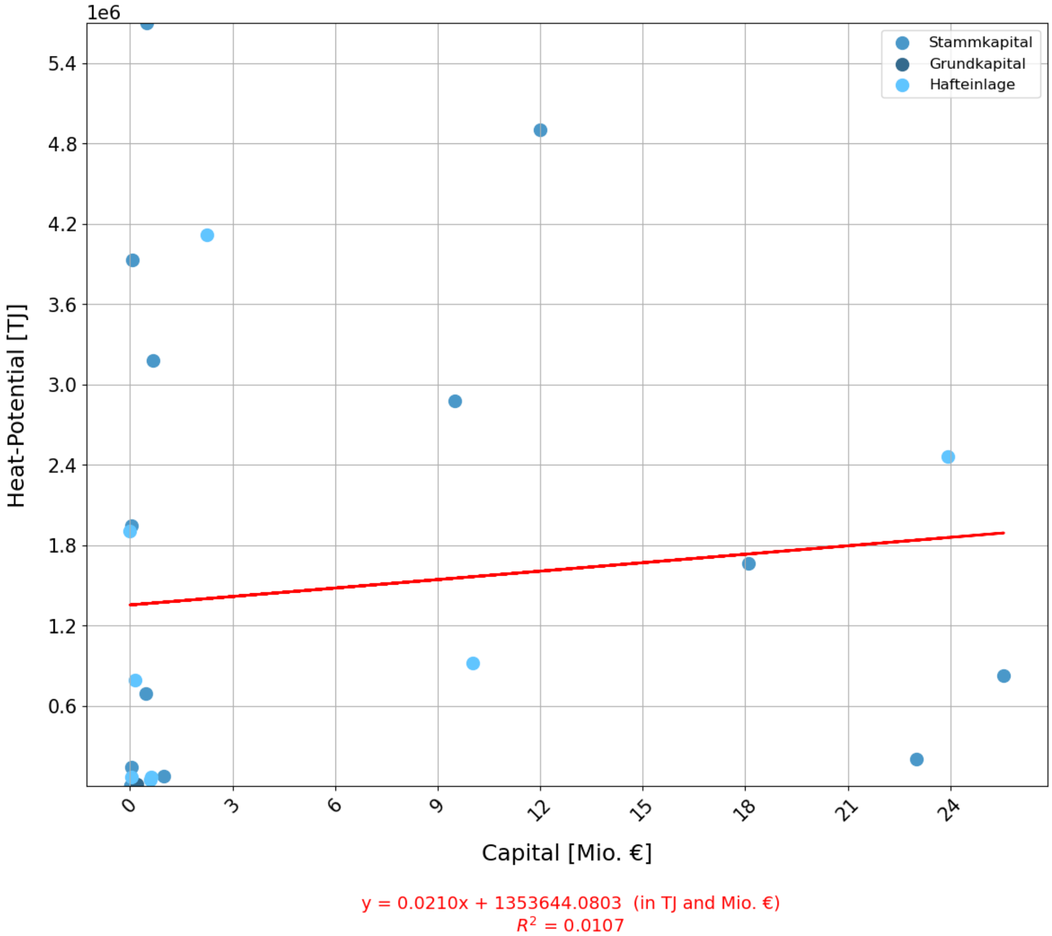


Figure 4.13 Results of the Paper and Printing industry: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 6e6 TJ. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

a complex valuation landscape that does not always reflect the substantial capital investments made by companies in this sector.

Moreover, when entering the paper business, substantial investments in machinery, technology, and environmental compliance measures are required. These investments, while crucial for operational efficiency, do not necessarily translate into commensurate increases in market value. This paired with the commodity status of many paper products, subjected to price volatility and margin pressures can difficult for paper companies to be able to increase their equity. Additionally, these may have an asset-light business model, not making further material acquisitions past the initial operational phase, and rather investing in growth to create new subsidiaries, something aligned with the fact that a high number of companies in this industry seem to be part of a group

[51] [52].

To take into the account is also the fact that the paper industry faces a transition in the market. This is caused by shifts in demand patterns driven by digitalization and changing consumer preferences. The same way that some segments, such as packaging, have grown due to the rise of e-commerce, others, like newsprint and writing paper, have seen declining demand. This additionally to the modest growth prospects, contributes to a valuation environment where capital investments may not directly correlate with perceived market value [51].

4.3.3 Chemical Industry

Once again, hydrogen potential in this sector will be primarily assigned to how hydrogen can be used as feedstock for the production of chemicals, and additionally as an energy carrier. Regarding, its recovery of the heat potential, it will be considered for different processes which will range from low to high required temperatures.

- **Hydrogen potential database**

  The observed data in Figure 4.14 demonstrates a favorable trend that lends support to the proposed hypothesis, though with a rather low fit, attributed to some outliers to be analysed. However, several important factors require consideration. For one side, the dual-use of hydrogen in this sector has to be considered, as hydrogen serves a dual purpose as both an energy fuel and a feedstock, similarly to the refinery sector. For the chemical industry hydrogen is evaluated so as to use it for steam reforming in the production of methanol and ammonia, for obtaining olefins from methanol, for electrolysis using methanol and ammonia as raw materials. This multifaceted utilization may influence the interpretation of results.

  Basell Polyolefine GmbH and INEOS Manufacturing Deutschland are the most notable outliers from what can be seen in the Figure 4.14, representing those dots with a low capital in comparison to their high hydrogen potential consumption. Taking into account that these companies manufacture oleofin-based products, it is logical that a high potential is assigned in the reference database. This is due to how producing oleofins is one of the highest hydrogen-consuming processes addressed in their approach. as hydrogen is meant to be used as feedstock to produce it from methanol.[53][54] [20]

  For the second outlier mentioned, INEOS Manufacturing Deutschland, the low-capital issue may lie in the fact that the branch addressed in Fraunhofer's database is no longer the main site for production of the company in Germany. In the Handelsresgister, the company INEOS Köln GmbH was found, which has a higher capital value, suggesting that this branch may be the one with higher activity and equity. This addresses the issue of the fact that the information from the reference database may not be regularly updated.

  When observing the outliers with high capital but lower hydrogen potential, the company BASF SE is a clear example. Investigating its activities, the conclusion drawn is initially how, being a major chemical company, its activity has broadened to a wide variety of production processes. Therefore, being

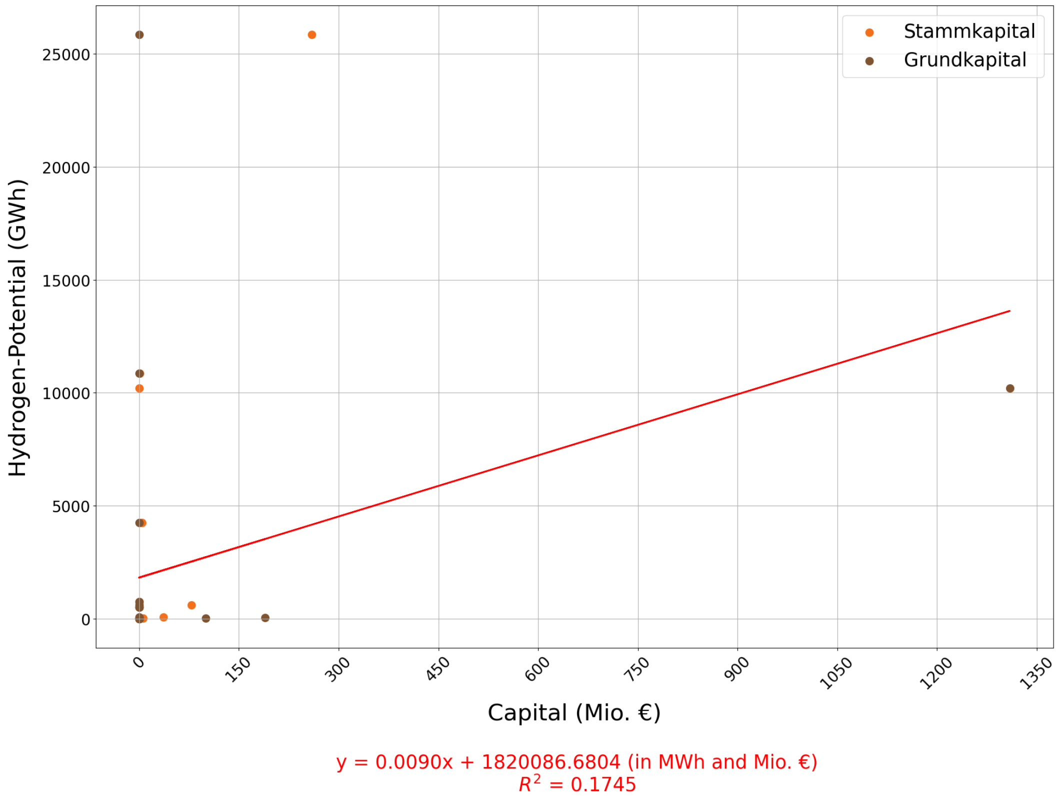


Figure 4.14 Results of the Chemical Industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

assigned a potential mainly according to its chemical processes, particularly the production of chlorine, ammonia, methanol and olefins, other activities, such as material processing, offering industrial solutions and surface technologies, may not be taken into account, even though they highly contribute to their economy.

- **Heat potential database**

  This sector seems to have a higher number of companies which have been assigned a relatively low potential, reason for which initially, a potential threshold is to be set at 1e12 TJ, in order to analyze outliers and understand their situation in the plot.

  When observing the sample of companies of the chemical industry in Figure 4.15, having set a less restrictive potential threshold, the outlier CABB GmbH can be observed in the area of higher capitals and represents the dot with the lowest potential with this. When researching this company, this has been found to be part of a small group of companies, and is focused on specialized chemicals, which can produce higher economic margins than basic chemicals, like the ones manufactured in other high potential companies such as

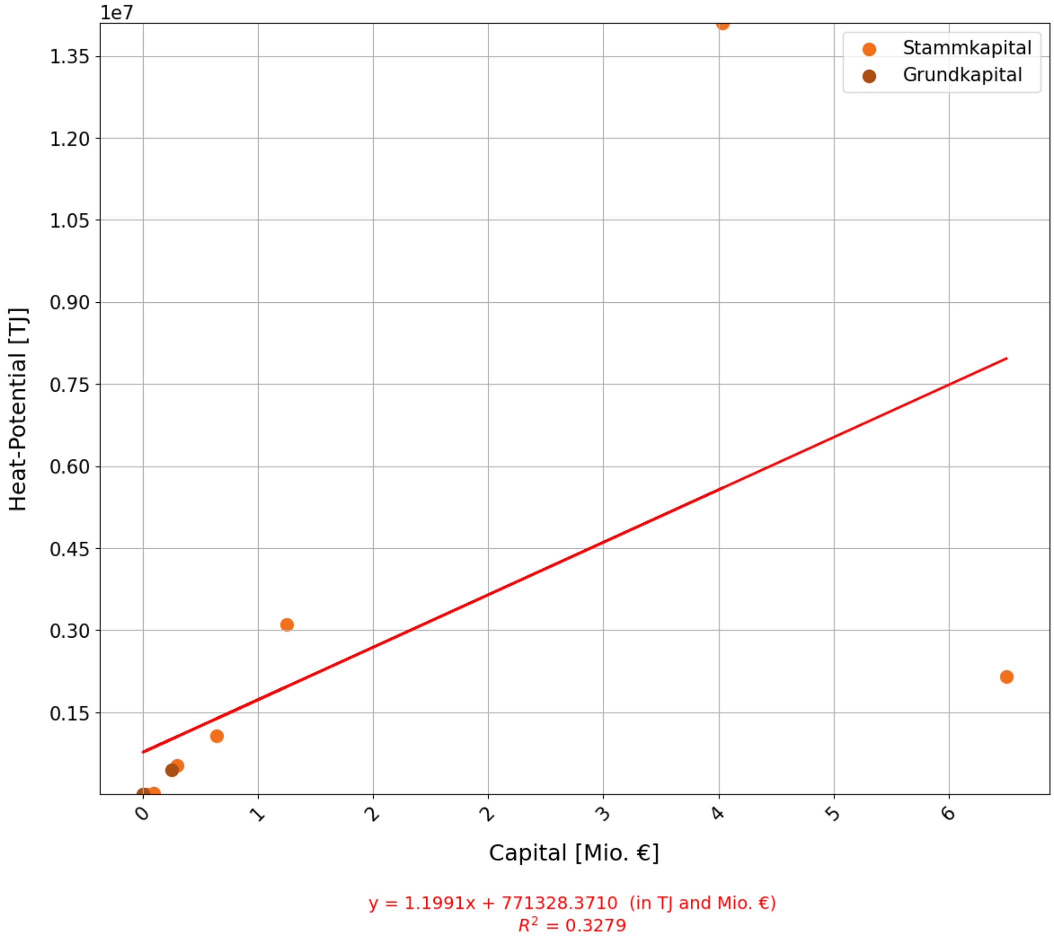


Figure 4.15 Results of the Chemical Industry: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

INEOS. Moreover, from the paper associated to the reference database, the products produced mainly from chlorine of this company, are associated to lower potentials than the assigned to the production of other basic chemicals such as ethylene and ammonia. Thus explaining, the lower assigned capital.

Many companies seem to lie in the area of even lower potentials, so a more restrictive threshold can permit a better comprehension of the results, as applied in Figure 4.16.

From Figure 4.16, a positive trend and a high R-squared value can be identified, suggesting an outstanding fit to the trend. However, it has been noted that these companies, even though not grouped with the approach used when dealing with duplicates, do belong to only 3 different larger groups of companies and another self-sustained company, something which in other cases has led to having confronted with the hypothesis. However, researching about their products, it has been noticed that they seem to commercialize with sim-

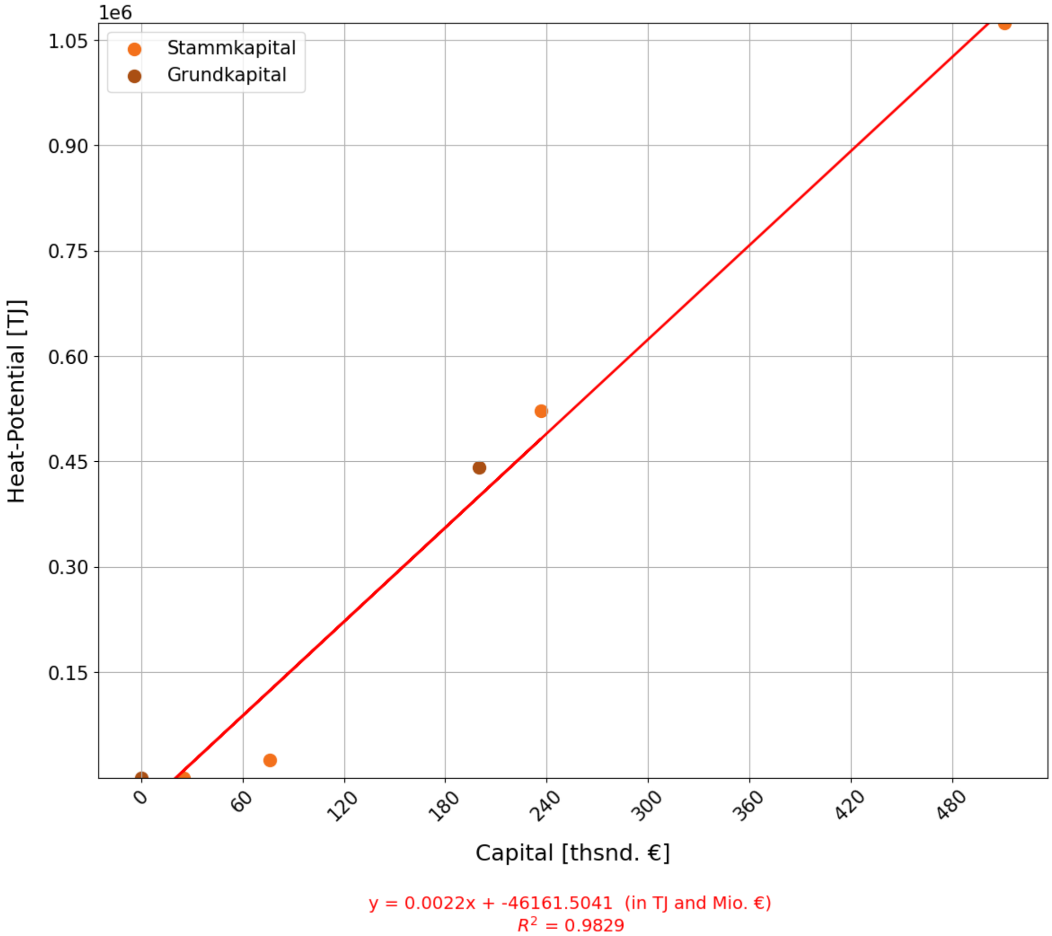


Figure 4.16 Results of the Chemical Industry: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 1.2e6 TJ. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

ilar products, so that implies that comparing them can provide accurate results.

4.3.4 Metal Processing

The processing of metals is linked to high operating temperatures, which will result sometimes in high potentials associated when certain processes are undertaken. For example, when using coke ovens, high temperature from the exhaust gases will be generated and natural gas burners will result in high quantities of hydrogen to be consumed. However, other forms of metal processing will be associated to much lower potentials. Between the processes considered, they both employ blast furnaces and hot rolling for the manufacture of products. Additionally, the Direct Reduced Iron (DRI) process fed with hydrogen is considered for steel production in the hydrogen potential database, something to be considered when evaluating the hypothesis for this particular subsector.

- **Hydrogen potential database**

  When observing Figure 4.17 with all the metal processing companies, a positive trend with a quite strong fit, evidenced by the high R-squared value, can be noticed.

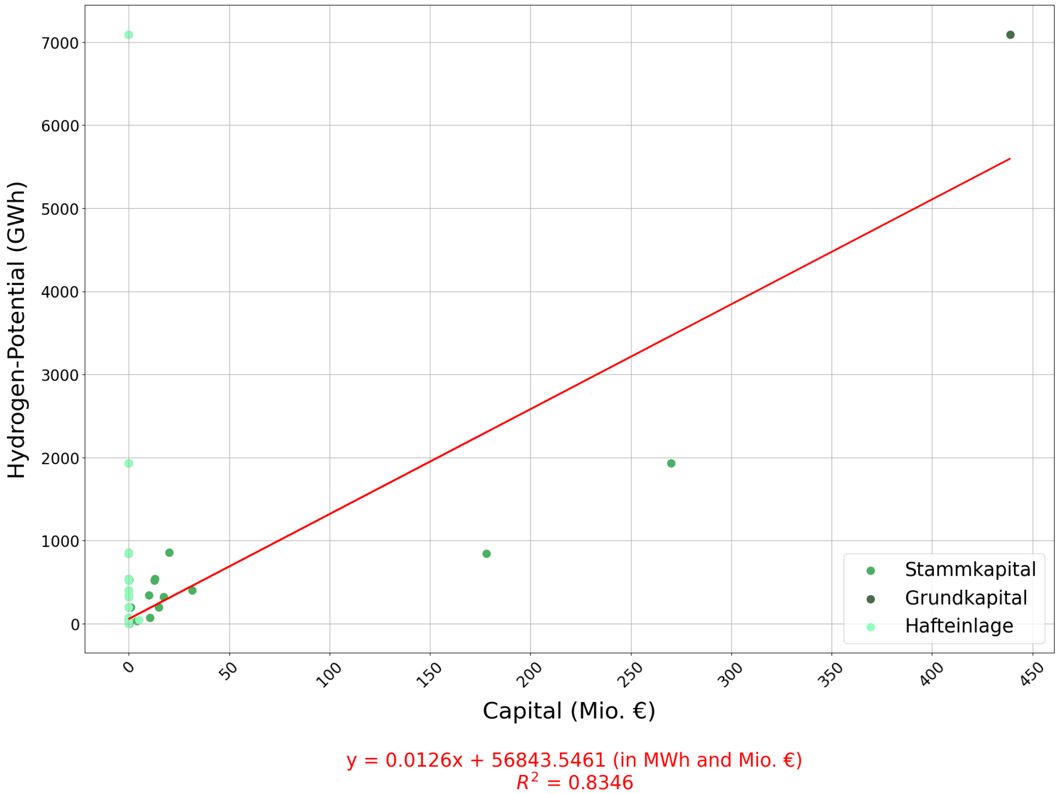
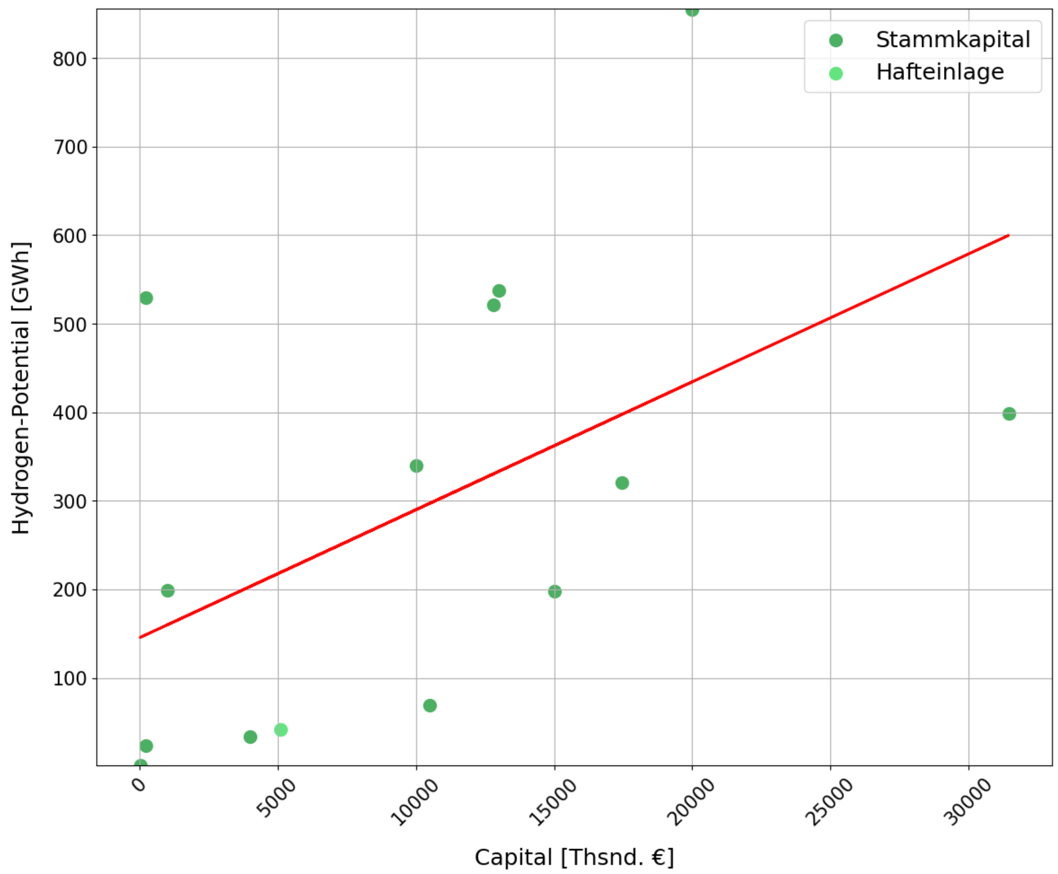


Figure 4.17 Results of the Metal Processing industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

  However, due to the high number of companies found in the lower capital and lower potential area, this area can be of greater interest to analyze.

  After applying a potential threshold of 1 TWh, Figure 4.18 still projects a positive trend, although with a weaker fit. Compromising the fit to the trend, the highly noticeable outlier with a high potential associated to a low capital is the company ArcelorMittal Duisburg GmbH is to be analysed. This company is part of the ArcelorMittal group, from which subsidiary companies are represented in the plot with higher potentials. The fact that this one presents such low capital may suggest that the group strategy may have led to money not being raised to their tangible assets, and has been invested in other facilities. However, due to the scarce information available, more detail about their strategy cannot be presented. [55]



y = 0.0145x + 145189832.6661 (in kWh and thsnd. €)
R^2 = 0.2663

Figure 4.18 Results of the Metal Processing industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 1 TWh. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

For the subsector of Steel, primary, as can be observed in Figure 4.19, the relationship is not only positive but presents a strong fit to the trend. As previously mentioned, hydrogen is mainly considered as feedstock for DRI processing. However, this is also a valid way of projecting energy consumption, as the higher the production, the higher the amount of hydrogen, as raw material, will be necessary.

- **Heat potential database** For the reference database in this sector, it is to be noted that steel is included within the metal processing sector, which is mainly dedicated to iron and steel production. It involves processes like coke ovens, blast furnaces, and basic oxygen furnaces. These generate high-temperature heat, making this sector a potential contributor to the recovery of excess heat. For example, blast furnaces operate at temperatures between 130 ℃ and 250 ℃ for exhaust gases. [21]

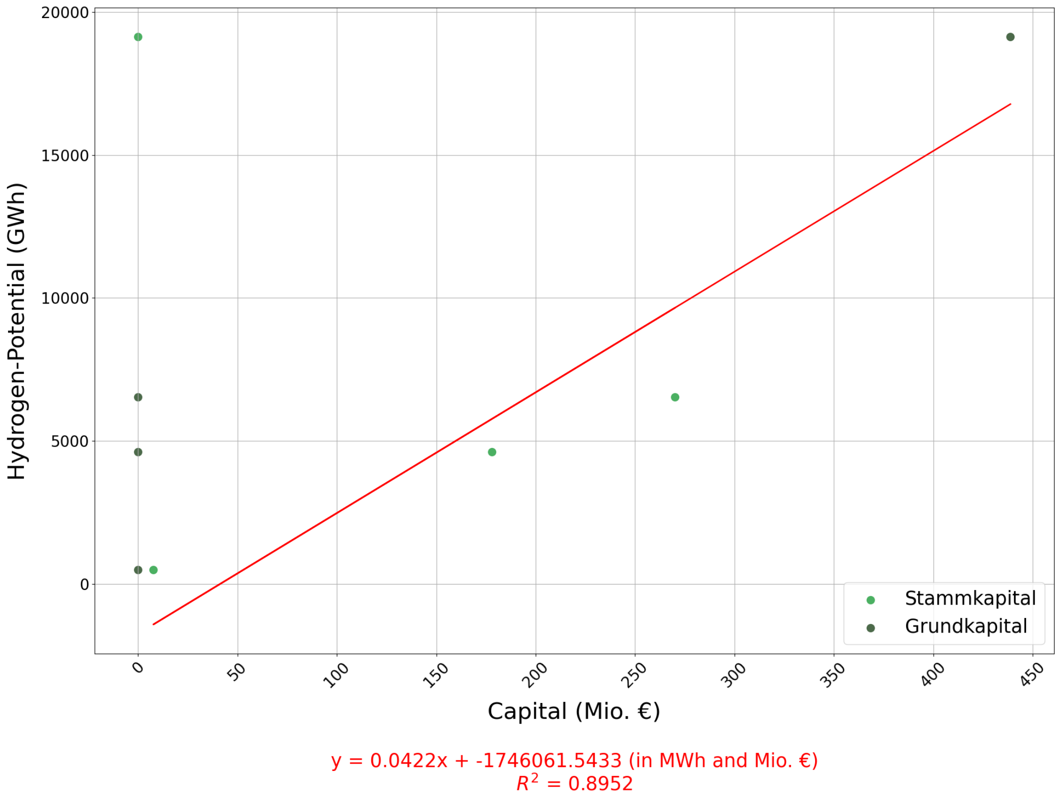


Figure 4.19 Results of the Steel, primary sector: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

From what can be seen from the 4.20, there seems to be a negative though not very clear relationship between the heat potential and the capital in this industry, evidenced by how the R-squared value is rather low.

To attempt to understand this lack of tendency, the companies from the Figure 4.20 are to be hereunder analysed. When researching products, most of them seem to be dedicated to steel production, presenting different modalities such as recycling steel and producing from basic to more specialized steel. Moreover, highly influencing in heat potential is the processes utilized for production. Most of these use electric furnaces, which have higher operational costs, due to the electricity required, but present a considerably lower potential and initial capital cost than coke ovens, which are still used by companies like ArcelorMittal Duisburg. This variability of processes and how these can affect the companies' assets and capital, can explain a wide variety of capitals associated with a similar energy potential. [55]

4.3.5 Non-Ferrous Metals

The analysis of the sector in question was constrained by significant limitations in the available data. Notably, the sample size was exceptionally small, comprising

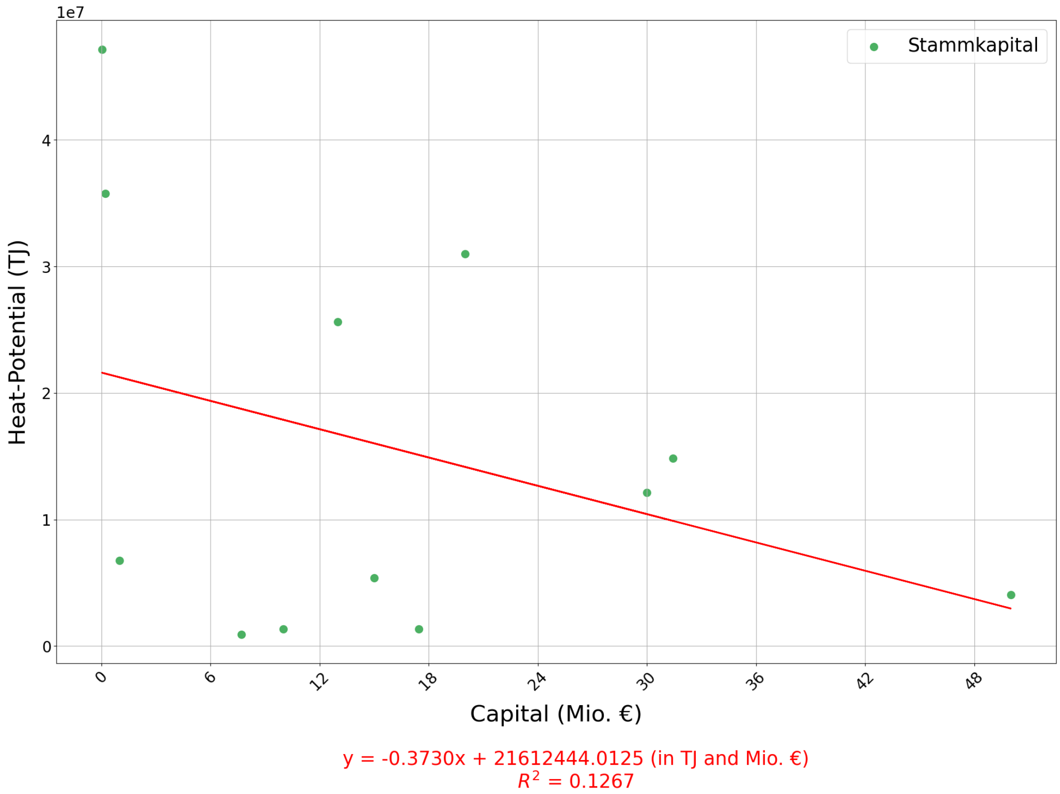


Figure 4.20 Results of the Metal Processing industry: The scatter plot visualizes the relationship between capital and heat recovery potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

only two different companies, with one of these being represented through its various branches. Therefore, establishing a reliable trend or pattern would prove to be challenging, and would not adequately evaluate the hypothesis proposed in the thesis.

4.3.6 Non-Metallic Minerals

This sector encompasses the production of glass, ceramics, cement and lime. These, compared to the rest of the sectors, will present medium-to-high scale potentials, regarding the usable heat potential for district heating. However, when evaluating their potential to apply hydrogen as their energy source, they prove to be of high interest due to the high temperatures required for processes, such as glass melting where a temperature of up to 1600°C is reached.

- **Hydrogen potential database**

  The hydrogen potential reference database distinguishes between 2 sectors when evaluating the potential of the mineral sector. On the one hand, it analyzes the Non-metallic mineral sector, to which glass manufacturing is attributed, comprehending the container and flat glass processes. And, on the other hand, the mineral processing sector is distinguished, including in it the burning of lime and the production of cement and clinker. [20]

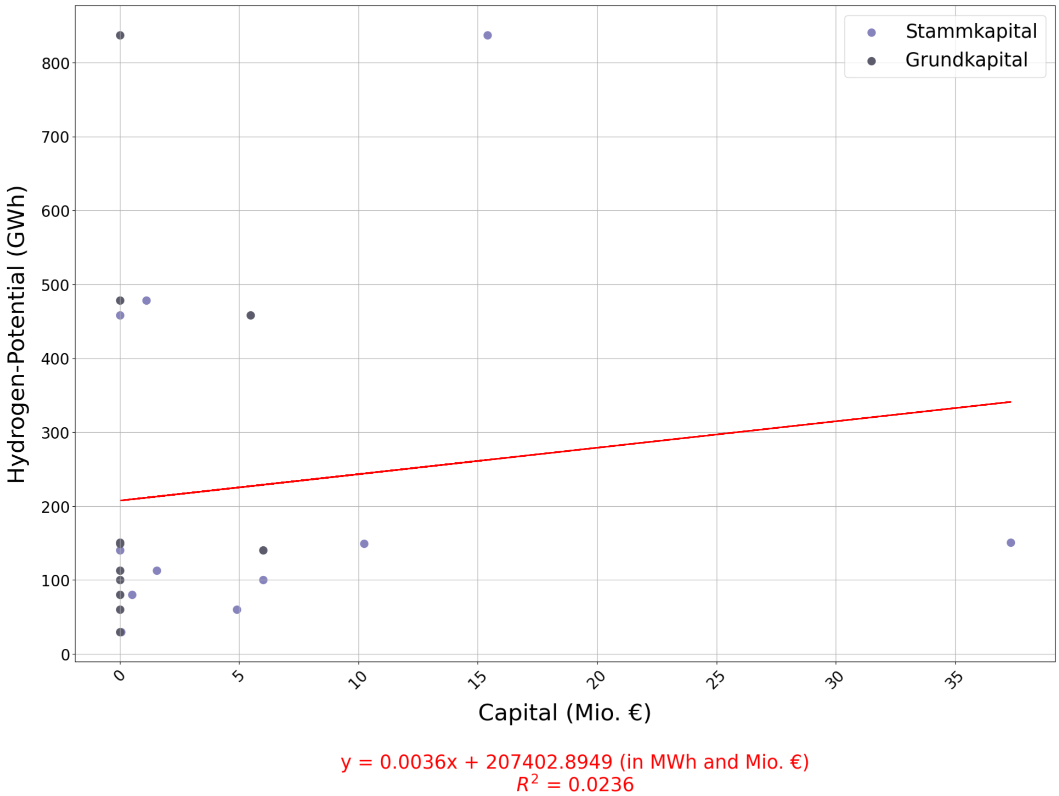


Figure 4.21 Results of the Non-metallic industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

As an overview of the sample of Non-metallic minerals in Figure 4.21, when analyzing the hydrogen potentials associated, a greater number of companies lie in the lower potential area of the plot. Consequently, the approach selected is to analyze the lower potential data samples to evaluate the correlation between capital and energy consumption.

Evaluating the companies to which a potential smaller than 200 GWh has been assigned, a positive trend can be observed, with a moderate fit, suggested by the R-squared value.

An important company group found in this low capital and potential subgroup is Gerresheimer. Having the recorded subsidiaries of this group, Gerresheimer Lohr seems to stand out for having a relatively low capital for a high potential, maybe due to the fact that its production is not only related to the glass industry and may not generate and requires financial assets as high as other subsidiaries, such as Gerresheimer Tettau. Much like the other companies before mentioned which were part of a group, the financial assets

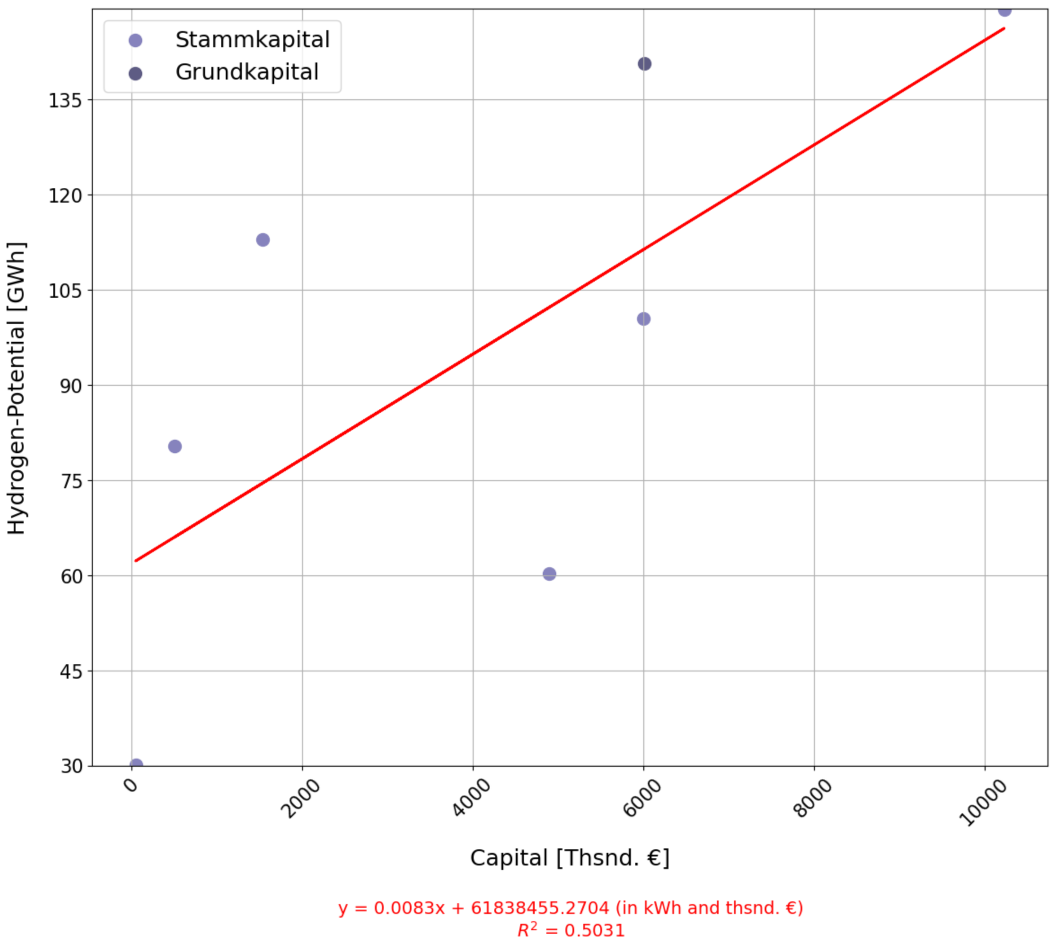


Figure 4.22 Results of the Non-metallic industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 200 GWh. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

may be divided by the different subsidiaries in a strategic manner, which may or may not be related to the intensity of their production. [56]

For the Mineral processing sector, where cement, lime and clinker companies are represented, two points can be made from what Figure 4.23 portrays.

Firstly, the light purple dots which seem to follow a positive trend in the central and right area of the plot, correspond to subsidiaries of a group which is associated to one single company of the reference database. Therefore, due to the approach explained in the Methodology section regarding the distribution of potentials, these are not to be taken into account. And, consequently, the area of low capital is left to evaluate the hypothesis.

From this detailed view of the data a positive trend with a moderate fit is observed, validating the hypothesis set in this thesis. Analyzing outliers, the company Portlandzementwerk Wittekind Hugo Miebach Söhne KG is to be

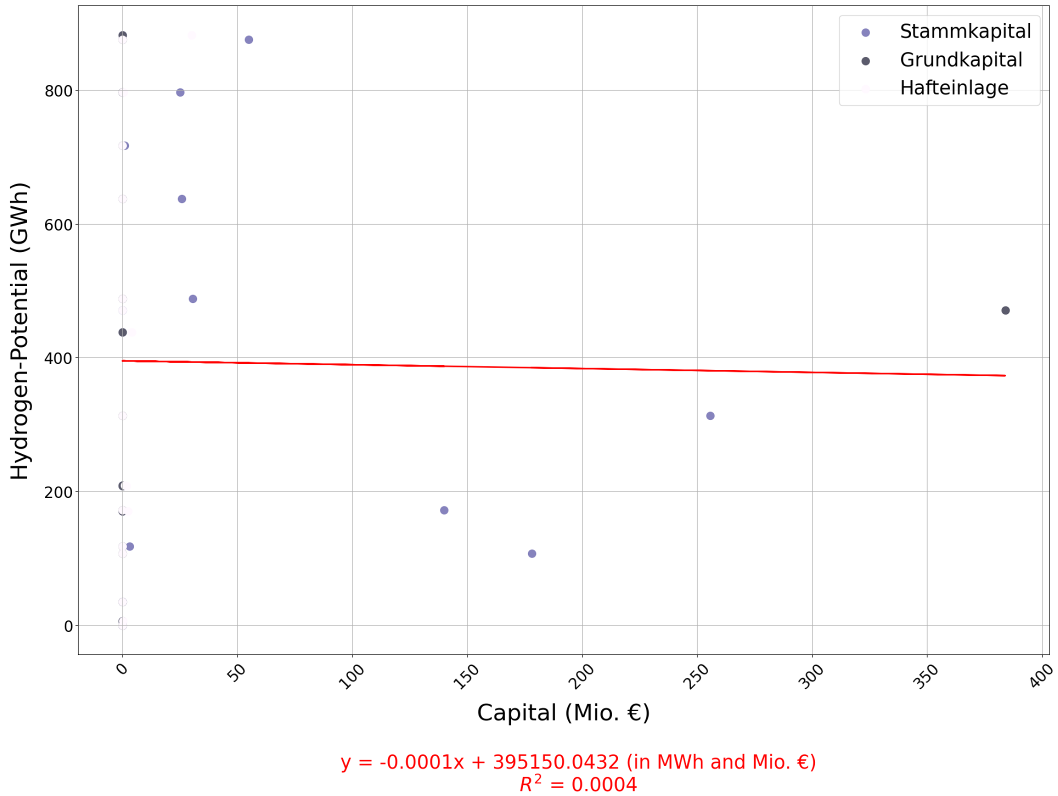


Figure 4.23 Results of the Mineral Processing Industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

noted with a high potential associated with a relatively low capital. This is a medium-sized, family-owned company in the Cement sector in Germany which, having the same potential as Märker Kalk GmbH, has a lower capital. This can be due to the fact that Märker Kalk GmbH is diversified between the production of lime and cement, which may suggest that due to its legal structure and higher production capacity for both Märker Kalk GmbH and Märker Klinker GmbH, requires of a higher capital. Moreover, they recently went through a modernization project, which requires the build up of capital. [57] [58]

- **Heat potential database**

  This databases unites the entire mineral sector into a single group, offering in that way a wider sample to evaluate. When removing obvious outliers, companies with lower heat potentials seem to be more abundant, reason for which it can seem convenient to take a close look at these.

  As it can be observed on the Figure 4.25, even though the regression line presents a positive relationship, the low R-squared value demonstrates that the fit to this tendency is quite poor. The companies which seem to contribute

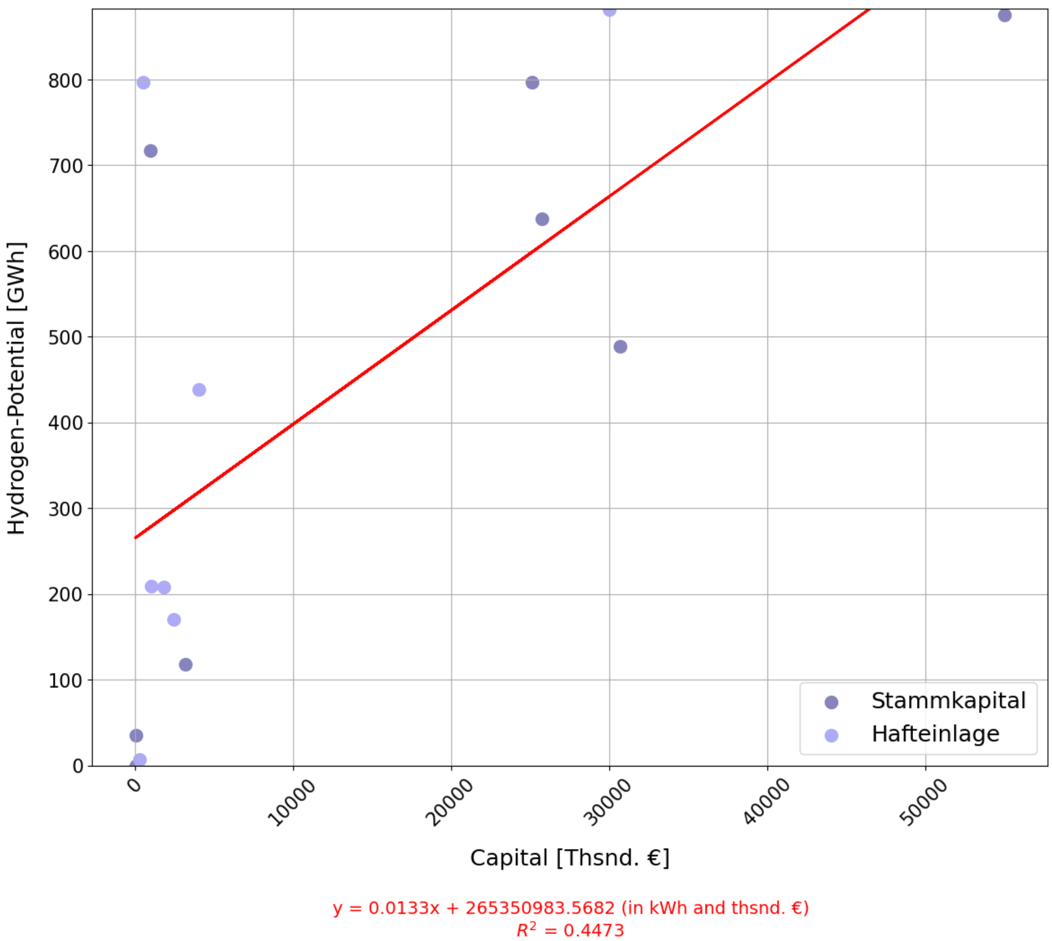


Figure 4.24 Results of the Mineral Processing Industry: The scatter plot visualizes the relationship between capital and hydrogen potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 1 TWh. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

to this slim R-squared value are Gerresheimer Essen GmbH, with a low capital corresponding to a high potential, and SGD Kipfenberg GmbH with a low potential for a high capital.

When looking into Gerresheimer Essen GmbH, as previously commented, it belongs to a larger group, which encompasses other subsidiaries which do follow the trend supporting the thesis' hypothesis. A reason for this particular subsidiary to present a high potential compared to its low capital can be due to it being a subsidiary of the main company Gerresheimer AG, settled in Düsseldorf. The main company is probably the one owning the highest capital in order to represent stability and equity. Moreover, the group covers the production of other plastic products, while Gerresheimer Essen GmbH produces container glass, which is associated with higher potential. These are mere assumptions as specific production data of all the subsidiaries is not publicly

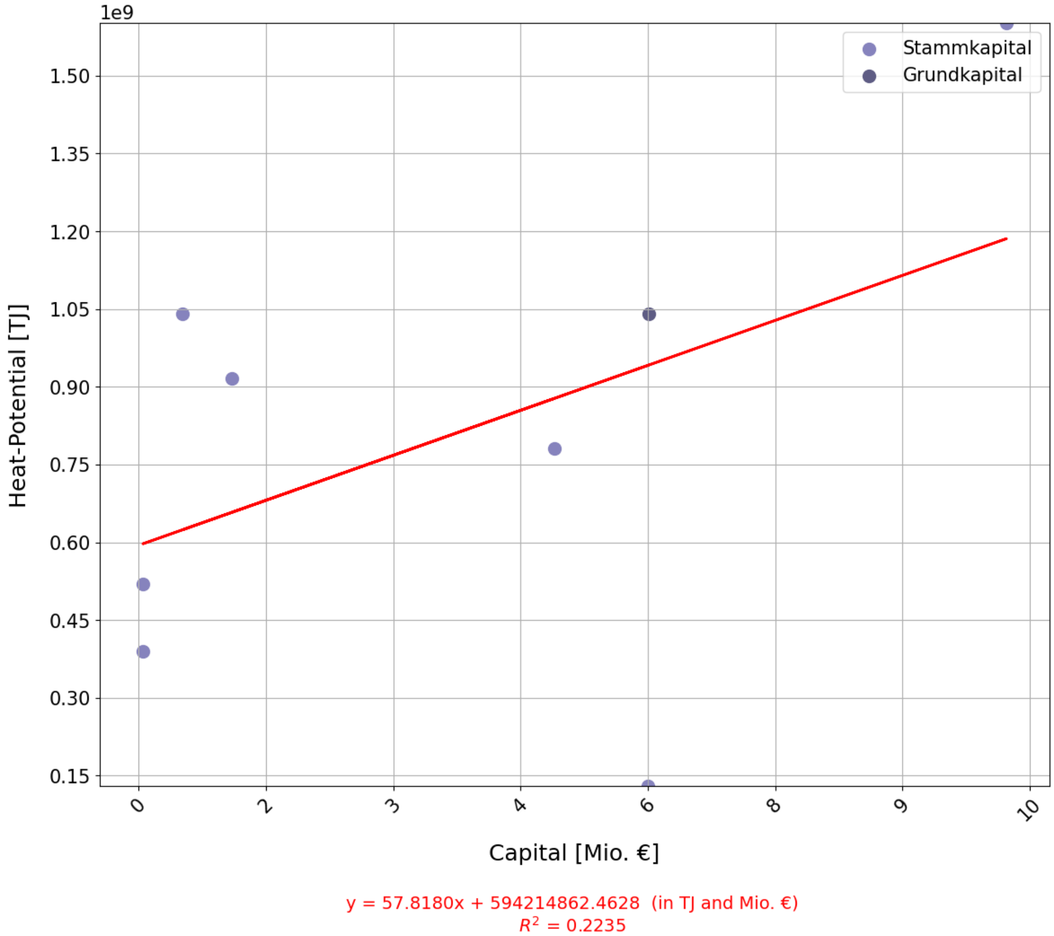


Figure 4.25 Results of the Non-metallic Minerals Industry: The scatter plot visualizes the relationship between capital and heat potential for companies and distinguishes between types of capital contributions: Stammkapital, Grundkapital, and Hafteinlage. A threshold is set to represent the companies with a potential smaller than 1.5e10 TJ. Each dot represents a company and the red line represents the linear regression, and its equation and R-squared value are shown below the plot.

available. [56]

SGD Kipfenberg GmbH, on the other hand, presents a lower heat recuperation potential for its high capital value. This company's activity also has a long history and is also dedicated to the production of pharmaceutical glass. However, no production figures can be found. What can explain the differences with Gerresheimer Essen GmbH is the fact that SGD Kipfenberg GmbH does not belong to a group, so more of its revenues can have been self-invested in their capital, rather than in expansion. Moreover, not having expanded to other subsidiaries also suggests that their production volumes may be lower, thus providing a lower heat recuperation potential. [59]

4.4 Map obtained

In order to visualize the results and be able to locate the companies addressed in the database easily, an interactive map was generated and attached to the GitHub repository associated to this thesis [5]. The colour of the dots of the map are associated to the sector the company belongs to, something which can be seen in the legend. Some of the sectors encapsulate multiple subsectors, for the following cases: the mineral industry sector includes non-metallic minerals and mineral processing subsectors, and the metal industry sector includes iron and steel, metal processing, non-ferrous metals and steel, primary subsectors. Moreover, the size of the dot indicates the magnitude of the capital, thus providing a graphic overview of the information obtained. Whenever the cursor is placed over a dot, a label with the name, sector, capital, hydrogen potential and heat recovery potential is shown.

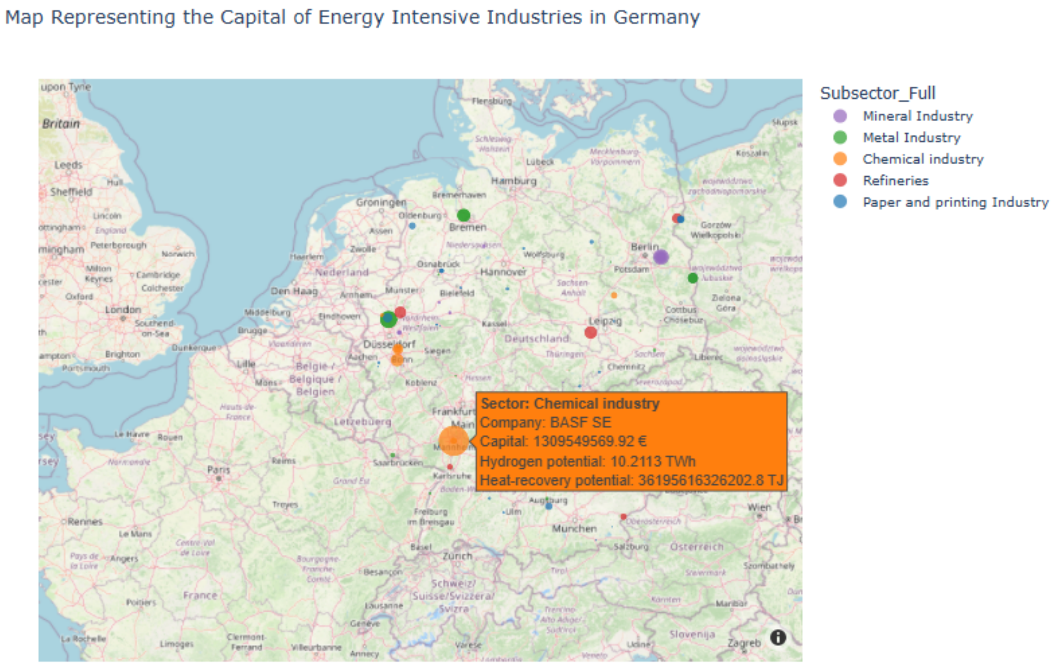


Figure 4.26 Map Representing the Capital of Energy Intensive Industries in Germany. Sectors are distinguished with different colours. The dots represent companies and their size represent the magnitude of the capital value associated. Each company has a label with its name, subsector, capital value, hydrogen potential and heat recovery potential.

The map permits to interactively zoom in and out of the map to explore the different locations of the dataset's industries.

Chapter 5

Discussion and Conclusion

Following the analysis and detailed examination of the Results, several conclusions were extracted regarding the correlation between a company's capital and heat recovery or hydrogen potential across various sectors. These will be hereunder discussed while highlighting ideas for possible implementation or further research that could result from this project.

5.1 Correlation between Capital and Energy Potential

In general, it can be stated that there is a general and positive, however sometimes weak, correlation between company capital and energy consumed or generated and to be recovered, as the hydrogen potential database [20] and heat recovery potential database [21] helped respectively verify, particularly when obvious outliers are excluded. This suggests that, in some cases, capital serves to give a sense of the energy consumption or recovery capacity. However, this trend is not consistent across all sectors, and notable deviations were identified.

The analysis highlights several factors that led to the mentioned deviations between capital and both hydrogen potential and heat recovery potential, which compromised the strength of the trends observed across sectors. A persistent issue is the presence of financial strategies that are decoupled from direct capital increases, often seen in sectors where companies are part of larger groups. This is particularly evident in the paper industry, also affected by the industry's cyclical demand and ongoing transition due to digitalization, further weakening the correlation between capital and energy potentials. This issue also affected the mineral processing sector, and weakened the fit between capital and energy potential, especially for heat recovery.

Another factor which led to the appearance of outliers in the trend where the presence of diversified companies whose capital is spread across different activities. An example of this is the chemical sector, where outliers like BASF have diversified operations, making it difficult to pinpoint a clear correlation between their capital and the hydrogen potential used for producing basic chemicals. Similarly, the issue arises in the metal processing sector, where the variability of processes—ranging from electric furnaces to coke ovens—creates a wide range of capital investments that diffuse the correlation when evaluating the heat recovery potential.

Sectors such as refineries and steel production show stronger correlations, where capital investments more directly align with hydrogen potential. However, even in these sectors, weak correlations for heat recovery potential suggest that capital investments may be used for innovation and to research for cleaner energy sources rather than directly increasing heat recovery efficiency.

5.2 Limitations of the Handelsregister Database and Data Availability [3]

An important issue to be addressed is how the *Handelsregister*, while being a public online database, was challenging to use as a means to fulfill the research. A cause of this was the mechanisms which limited its access, perhaps as a means of protection from cyber-attacks, even though it is meant to be free and openly available. The blocking of IP addresses after multiple queries made within a short time frame was the main issue. This technical limitation made it difficult to obtain the large datasets necessitated when attempting to get company information with the automated method used, as explained in the Selenium section. This led to more time being spent and the need to look for alternative locations with unrestricted IP addresses, and as a consequence compromised the extension of the data collection.

Moreover, the scarcity of freely and publicly available company data online has resulted in the commercialization of such information. As this information can prove to be of great utility, many businesses have emerged that collect and sell company information of different sectors, exploiting the lack of a comprehensive, open-source platform. Had more data been accessible, the number of companies included in the resulting dataset for the present thesis could have been larger. Consequently, the combination of the factors explained considerably influenced the scope of this research, as some companies were not included simply because obtaining their information required additional costs or time, also affecting the robustness of the validation in the Results.

5.3 Sectors with a Positive Correlation

After analyzing the different sectors, the ones listed below seemed to project a correlation between company capital and energy consumption or heat recovery potential.

- **Non-Metallic Minerals**: A moderate positive correlation was found between capital and hydrogen potential for the companies dedicated to glass and cement production. This was particularly so when analyzing companies with lower capital and energy potential. The non-metallic mineral sector consistently showed alignment with the proposed hypothesis, however, the size of the low potential company sample was rather low.

- **Metal Processing**: This sector displayed a relatively strong correlation between capital and hydrogen potential, with a high R^2 value. Steel production represented this in particular, as energy consumption is a significant factor, leading to higher energy potentials. For the heat potential dataset, the same conclusion could not be extracted, though this is intuited to be a cause of the encapsulation of very different processes in this database's particular sector, where, unlike in the hydrogen potential database, the iron and steel manufacturing and processing was put into a single sector. [21]

- **Chemical Industry**: While the overall fit was weak, a positive correlation was observed in companies involved in the chemical sector. A moderate fit to the trend was obtained during the validation with the heat potential database, the more accurate reference for this particular sector due to how the hydrogen potential database also included hydrogen as a raw material for manufacturing, apart from as a source of energy. While both give insight into production and thus energy consumption, the yield of hydrogen for the different applications of this element can be different. [20]

- **Refineries**: The refinery sector showed a positive and moderate correlation when evaluating capital against hydrogen potential. Even though this reference database also considered the use of hydrogen as feedstock, as its potential was exclusively evaluated in this manner, it still gave an insight into its production, and therefore its potential energy consumption [20]. However, a positive but weak correlation when evaluating its heat potential was seen, something which can be attributed to the financial strategies of these companies. Companies that heavily invest in research and development, to promote the transition to sustainable fuel alternatives such as green diesel or sustainable aviation fuel projects, often see a reduction in their capital without an immediate increase in their assets, affecting the relationship between capital and energy.

5.4 Sectors with Weak or No Correlation

Other sectors presented either weaker or no correlation at all between energy consumption or heat recovery potential:

- **Paper and Printing**: The paper industry, being the sector with the highest data sample, exhibited a rather weak correlation between capital and energy. This was highlighted in the cases where companies were part of larger corporate groups. A common case observed was for subsidiaries showing low capital despite their contributions to energy potential. The causes of this can be attributed to financial strategies at the group level. The paper industry appeared to invest in production rather than increasing capital, weakening the correlation.

- **Non-Ferrous Metals**: This sector was the most affected by data access limitations, reason for which it was not possible to establish any reliable trend between capital and hydrogen potential. Thus, no significant conclusions could be drawn.

5.5 Hypothesis Validation

All in all, the hypothesis of this exploratory research suggesting a proportionality between capital and energy potential was found to be sector-dependent. Moderate to strong correlations were validated, sometimes with only one of the reference databases, with the metal processing, refineries, chemical and minerals sectors. Nevertheless, others such as in the heat recovery potential evaluation with refineries, and the paper industry, even more so whenever high research investment was present, portrayed weaker fits to the trend. This gives a sense of the importance of sector-specific approaches when analyzing the relationship between capital and energy consumption.

5.6 Value of the Created Dataset for Future Research

The same way as the reference databases have been used, the dataset created during this thesis can serve as a valuable reference source for future research projects exploring the correlation between capital and energy consumption or recovery potential as well as other data-based analyses. After compiling a freely and easily accessible list of companies across various industries, the dataset offers a foundation for validating similar hypotheses and analyzing energy use patterns in large industrial sectors. Future researchers with broader access to information sources can make use of the information collected in this dataset for validation, as well as to expand its content, perhaps with more detailed or sector-specific data. This source is published in a GitHub repository, openly available to the public and with an access link in the present thesis document [5].

5.7 The Role of Publicly Available Energy Data in the Energy Transition

Access to more detailed and publicly available information on energy consumption by large industries can prove to be of considerable utility. It would contribute to the implementation of strategies promoting the integration of renewable energy sources, as it would aid in the placement of renewable energy plants in regions with high industrial energy demand. Thus, achieving a decentralized and optimized energy supply chain. Moreover, unused products from industries could be taken advantage of, such as, as outlined in the heat recovery potential database [21], excess heat which could be recovered for efficient district heating networks. Energy waste would substantially be reduced and a circular economy could be consequently promoted, benefiting both industry and communities.

Eventually, a successful and urgently needed energy transition can be fulfilled with access to energy consumption data. Integration of ever-increasing renewable energy sources and investing in the optimization of energy recovery from industrial processes can contribute to a cleaner and more sustainable energy future.

5.8 Data Limitations for Validation

The assessment of the validity of the hypothesis, would be significantly improved with additional and more comprehensive data. As previously discussed, the lack of freely available company data online significantly resulted in a smaller company data sample and sectors being underrepresented. As future objectives, collecting more open and updated datasets would be crucial to provide deeper insights, more so when addressing sectors critical for the energy transition.

5.9 Outlook and Future research

Taking in all the conclusions previously mentioned, it can be stated that while financial data alone cannot precisely determine energy consumption, this exploratory analysis has demonstrated its potential use as a proxy for estimating energy demand patterns within industry sectors. As future improvements and research continuation, the aim would be to address the limitations highlighted in this study by integrating additional data sources, particularly from large sectoral associations or industry reports. This way, the resulting dataset obtained and the correlations observed would have more robustness.

An important aspect which would be of utility for future research as a whole, is to ensure that databases are consistently updated. During this research, discrepancies were detected as a result of outdated data across different sources, thus limiting the accuracy of the analysis. Taking the effort to keep the information up to date can substantially improve the reliability of energy consumption estimates and make the analysis more relevant to real-time industrial trends.

Moreover, the research could pursue the design of advanced models that incorporate real-time energy consumption data and could refine the relationship between financial metrics and energy usage. These regions with high energy demand could be located, allowing for the strategic placement of renewable energy plants. This integration would support the development of a Decentralized Energy System, promoting an efficient and collaborative management of energy generation and consumption.

In conclusion, expanding the dataset and obtaining financial support to access larger pools of information will not only facilitate the estimation of energy consumption but will also help foster more energetically sustainable industrial practices, contributing to the much-needed energy transition.

Bibliography

[1] *EUROSTAT*. Accessed: 2024-04-02. 2024. URL: https://ec.europa.eu/eurostat/de/data/database (cit. on pp. 13, 15, 16).

[2] *DESTATIS Statistische Bundesamt*. Accessed: 2024-08-02. 2024. URL: https://www.destatis.de/EN/Press/2020/04/PE20_152_435.html (cit. on pp. 13, 15–18).

[3] *Handelsregister: The German Commercial Register*. Accessed: 2024-06-19. 2024. URL: https://www.handelsregister.de/ (cit. on pp. 13, 14, 19, 21, 24, 26, 28, 30, 31, 33, 64).

[4] John Conti et al. "International Energy Outlook 2016". In: *U.S. Energy Information Administration* (2016). Accessed: 2024-09-11 (cit. on p. 13).

[5] María Martín Escudero. *Master Thesis: Exploratory Data Analysis of the Handelsregister to Evaluate the Energy Consumptions of Germany's Industry Sectors*. 2024. URL: https://github.com/marmartinesc/Master-thesis (cit. on pp. 14, 28, 30, 31, 61, 66).

[6] *AG Energiebilanzen*. Accessed: 2024-04-02. 2024. URL: https://ag-energiebilanzen.de/ (cit. on p. 16).

[7] *BDEW Strom-Kennzahlen*. Accessed: 2024-04-02. 2024. URL: https://www.bdew.de/service/daten-und-grafiken/ (cit. on p. 16).

[8] *ENTSO-E Transparency Platform*. Accessed: 2024-04-02. 2024. URL: https://transparency.entsoe.eu/ (cit. on p. 16).

[9] *IEA Monthly Electricity Statistics*. Accessed: 2024-04-02. 2024. URL: https://www.iea.org/data-and-statistics/data-product/monthly-electricity-statistics (cit. on p. 16).

[10] VDEH Association. *VDEH Steel Industry Portal*. Accessed: 2024-09-05. 2024. URL: https://www.vdeh.de (cit. on p. 17).

[11] Boston Consulting Group. *Plant Facts*. Accessed: 2024-09-05. 2024. URL: https://plantfacts.bcg.com (cit. on p. 17).

[12] Global Cement. *Global Cement Industry Portal*. Accessed: 2024-09-05. 2024. URL: https://www.globalcement.com (cit. on p. 17).

[13] VDZ. *VDZ Cement Industry Portal*. Accessed: 2024-09-05. 2024. URL: https://www.vdz-online.de (cit. on p. 17).

[14] Glass Global. *Glass Global Plant Data*. Accessed: 2024-09-05. 2024. URL: https://plants.glassglobal.com (cit. on p. 17).

[15] Research Germany. *Research Germany: Glass and Ceramics*. Accessed: 2024-09-05. 2024. URL: https://www.researchgermany.com (cit. on p. 17).

[16] Fastmarkets. *Fastmarkets Forest Products*. Accessed: 2024-09-05. 2024. URL: https://www.fastmarkets.com (cit. on p. 17).

[17] Euro Chlor. *Euro Chlor Industry Portal*. Accessed: 2024-09-05. 2024. URL: https://www.eurochlor.org (cit. on p. 17).

[18] Chlorine Industry Review. *Chlorine Industry Review*. Accessed: 2024-09-05. 2024. URL: https://www.chlorineindustryreview.com (cit. on p. 17).

[19] Maximilian Schumacher and Lion Hirth. "How Much Electricity Do We Consume? A Guide to German and European Electricity Consumption and Generation Data". In: *SSRN Journal* (2015). Accessed: 2024-06-19 (cit. on p. 16).

[20] Marius Neurwith et al. "The Future Potential Hydrogen Demand in Energy-Intensive Industries - a Site-Specific Approach Applied to Germany". In: *Energy Conversion and Management* (2022). URL: https://isi.pages.fraunhofer.de/pshp/ (cit. on pp. 18, 21, 24, 31, 33–35, 39, 48, 56, 63, 65).

[21] Pia Manz et al. "Decarbonizing District Heating in EU-27 + UK: How Much Excess Heat Is Available from Industrial Sites?" In: *Energy Conversion and Management* (1-09-2021). Maximilian Schumacher is affiliated with Neon Neue Energieökonomik GmbH (Neon), Germany. Lion Hirth is affiliated with Neon Neue Energieökonomik GmbH (Neon), Germany, and the Potsdam Institute for Climate Impact Research (PIK), Germany. Jane Doe is affiliated with Institute of Energy Studies, Germany. John Smith is affiliated with University of Renewable Energy, Germany. (cit. on pp. 19, 24, 31, 33, 36, 38, 39, 46, 53, 63, 65, 66).

[22] *Nominatim*. Accessed: 2024-08-20. 2024. URL: https://nominatim.org/release-docs/latest/ (cit. on p. 19).

[23] *Google Maps API*. Accessed: 2024-08-20. 2024. URL: https://mapsplatform.google.com/ (cit. on p. 19).

[24] *Selenium*. Accessed: 2024-08-20. 2024. URL: https://selenium-python.readthedocs.io/getting-started.html#simple-usage (cit. on p. 19).

[25] *Pandas*. Accessed: 2024-08-20. 2024. URL: https://pandas.pydata.org/ (cit. on p. 20).

[26] *Regular Expressions*. Accessed: 2024-08-20. 2024. URL: https://www.regular-expressions.info/ (cit. on p. 20).

[27] *Plotly Express*. Accessed: 2024-08-20. 2024. URL: https://plotly.com/python/plotly-express/ (cit. on p. 20).

[28] *Matplotlib*. Accessed: 2024-08-20. 2024. URL: https://matplotlib.org/ (cit. on p. 20).

[29] *Offene Register*. Accessed: 2024-08-29. 2024. URL: https://offeneregister.de/ (cit. on p. 23).

[30] Łukasz Topolewski. "Relationship between Energy Consumption and Economic Growth in European Countries: Evidence from Dynamic Panel Data Analysis". In: (2021). Accessed: 2024-08-15 (cit. on p. 25).

[31] Grażyna Szustak et al. “The Relationship between Energy Production and GDP: Evidence from Selected European Economies”. In: (2022). Accessed: 2024-08-15 (cit. on p. 25).

[32] Tod Moss and Jacob Kincer. “How does energy impact economic growth? An overview of the evidence”. In: *Energy Conversion and Management* (2023). Accessed: 2024-08-15 (cit. on p. 25).

[33] Uwe Deichmann et al. “Relationship between Energy Intensity and Economic Growth”. In: (2018). Accessed: 2024-08-15 (cit. on p. 25).

[34] Federal Ministry of Justice and Consumer Protection. *Aktiengesetz (AktG) § 7 Grundkapital*. [Accessed on 2024-09-05]. 2023. URL: `https://www.gesetze-im-internet.de/aktg/__7.html` (cit. on p. 25).

[35] Theodor Baums and Klaus J. Hopt. *Aktiengesetz Kommentar*. 5th. Munich: C.H. Beck, 2021 (cit. on p. 25).

[36] Gerd Dörner and Wolfgang Schwäbe. *GmbH-Recht*. Munich: C.H. Beck, 2022 (cit. on p. 26).

[37] Haufe Group. *Stammkapital – Definition, Höhe und Einzahlung*. Accessed on 2024-09-05. URL: `https://www.haufe.de/recht` (cit. on p. 26).

[38] Federal Ministry of Justice and Consumer Protection. *Handelsgesetzbuch (HGB) § 171 Hafteinlage*. [Accessed on 2024-09-05]. 2023. URL: `https://www.gesetze-im-internet.de/hgb/__171.html` (cit. on p. 26).

[39] Karsten Schmidt and Marcus Lutter. *Gesellschaftsrecht*. 20th. Munich: C.H. Beck, 2023 (cit. on p. 26).

[40] Finances Time. “Why Do Companies Issue Shares? 5 Simple Reasons Explained”. In: (April 2023). https://financestime.com/blog/why-do-companies-issue-shares-5-simple-reasons-explained/, Accessed: 2024-08-15 (cit. on p. 26).

[41] MA Community. “Capital raising for growth and innovation”. In: (October 2023). https://mnacommunity.com/insights/capital-raising-for-growth-and-innovation/, Accessed: 2024-08-15 (cit. on p. 26).

[42] Shrishti Bhardwaj. “What is Share Capital?” In: (December 2023). `https://www.smallcase.com/learn/what-is-share-capital/`, Accessed: 2024-08-15 (cit. on p. 26).

[43] Lawyers Germany. “The Limited Partnership in Germany”. In: (January 2023). Accessed: 2024-08-15 (cit. on p. 27).

[44] *Establishing a KG (limited partnership)*. Accessed: 2024-08-15. January 2023. URL: `https://www.bildt-law.de/en/kg-gruenden` (cit. on p. 27).

[45] *HOLBORN Europa Raffinerie*. Accessed: 2024-09-06. 2024. URL: `https://www.holborn.de/` (cit. on p. 40).

[46] *Azubiyo*. Accessed: 2024-09-06. 2024. URL: `https://www.azubiyo.de/ausbildungsbetriebe/miro-mineraloelraffinerie-oberrhein/` (cit. on p. 41).

[47] *Branchenanalyse Mineralölindustrie*. Accessed: 2024-09-06. 2024. URL: `https://www.boeckler.de/fpdf/HBS-007177/p_fofoe_WP_133_2019.pdf` (cit. on p. 41).

[48] Uwe Deichmann et al. “Success Story Adolf Jass Schwarz”. In: (2024). Accessed: 2024-09-07 (cit. on p. 44).

[49] *Jass Group*. Accessed: 2024-09-07. 2024. URL: https://jass.com/ (cit. on p. 44).

[50] *Kanzan*. Accessed: 2024-09-07. 2024. URL: https://www.kanzan.de/ (cit. on p. 45).

[51] *The packaging, pulp and paper industry in the next decade | McKinsey*. Accessed: 2024-09-07. 2019. URL: https://www.mckinsey.com/industries/packaging-and-paper/our-insights/pulp-paper-and-packaging-in-the-next-decade-transformational-change (cit. on p. 48).

[52] *World Paper Mill*. Accessed: 2024-09-07. 2020. URL: https://worldpapermill.com (cit. on p. 48).

[53] *Basell:Polyolefins*. Accessed: 2024-09-07. 2024. URL: https://en.wikipedia.org/wiki/Basell_Polyolefins (cit. on p. 48).

[54] *INEOS*. Accessed: 2024-09-07. 2024. URL: https://www.ineos.com/businesses/ineos-olefins-polymers-europe/sites/ (cit. on p. 48).

[55] *ArcelorMittal*. Accessed: 2024-09-07. 2024. URL: https://barsandrods.arcelormittal.com/mill-duisburg (cit. on pp. 52, 54).

[56] *Gerresheimer*. Accessed: 2024-09-07. 2024. URL: https://www.gerresheimer.com/en/home (cit. on pp. 57, 60).

[57] *Orcan Energy*. Accessed: 2024-09-07. 2024. URL: https://www.orcan-energy.com/de/details/zementwerk-wittekind-beabsichtigt-deutliche-minderung-seiner-co2-emissionen-dank-abw%C3%A4rmel%C3%B6sungen-von-orcan-energy.html (cit. on p. 58).

[58] *Märker group*. Accessed: 2024-09-07. 2024. URL: https://www.maerker-gruppe.de/ (cit. on p. 58).

[59] *SGD Pharma*. Accessed: 2024-09-07. 2024. URL: https://www.sgd-pharma.com/ (cit. on p. 60).