Federal University of Rio Grande do Sul
Business School | Postgraduate Program in Administration
Master's Degree in Administration | Area of Concentration: Marketing

Jéssica Lima Miranda

Using small area estimation to identify business opportunities

Porto Alegre, 2024

Jéssica Lima Miranda

Using small area estimation to identify business opportunities

Dissertation, in article format, presented to the Postgraduate Program in Administration of the Federal University of Rio Grande do Sul as a requirement for obtaining the degree of Master in Administration.

Advisor: Prof. Dr. Vinícius Andrade Brei

Defense Committee:

Prof. Dr. **Daniel de Abreu Pereira Uhr** - PPGOM/UFPel

Prof. Dr. **Leonardo Nicolao** – PPGA/UFRGS

Prof. Dr. **Ricardo Limongi França Coelho** - PPGADM/UFG

Porto Alegre, 2024

CIP - Catalogação na Publicação

Miranda, Jéssica Lima
Using small area estimation to identify business opportunities / Jéssica Lima Miranda. -- 2024.
72 f.
Orientador: Vinícius Andrade Brei.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Escola de Administração, Programa de Pós-Graduação em Administração, Porto Alegre, BR-RS, 2024.

1. Small Area Estimation. 2. Spatial Marketing. 3. Business Strategy. I. Brei, Vinícius Andrade, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

ACKNOWLEDGMENTS

First and foremost, I extend my gratitude to my advisor, Professor Vinicius A. Brei, for his invaluable guidance, encouragement and support through these 2 years of journey. His expertise and insights have been instrumental in shaping this research.

I'm also deeply thankful to the professors who have contributed to my master's and life's growth: Professor Leonardo Nicolao, Professor Luiz Antônio Slongo and Professor Fernando Bins Luce. Their dedication and passion for teaching have greatly enriched my learning experience.

To my family, whose support has been my anchor and my everything: my beloved and beautiful sons, Eduardo e Matheus; my powerful mother, Elisabeth; and my incredible husband, Fabricio. I have no words to describe your unconditional love, patience and encouragement. I'm sure it'd be impossible without you all.

I'm also immensely grateful to my colleagues that became friends and made this journey memorable. Thank you, Fernanda and Rafael, for being by my side until the end of it. Your friendship, collaboration, laughter and tears shared made every moment more meaningful.

Finally, I extend my special thanks to my family, friends and colleagues for their support. Your belief in my journey has been a source of inspiration and motivation. Thank you all for being part of this important chapter in my life.

I. INTRODUCTION

Historically, sample surveys have been widely used to produce general estimates of the whole (Gosh & Rao, 1994; Rao & Molina, 2015). As the need for precise and granular quantitative data has increased, so has the innovation of research methods. Detailed and accurate information about small regions and populations has become as crucial as generalized information about the whole in which they are found. Particular knowledge about detailed parameters of smaller regions has become a potent tool for decision-making (Noble, 2010; Rao & Molina, 2015). Moreover, it is often necessary for policy-making in education and health.

Small area estimation (SAE) meets this need. It is a historically used statistical technique that aims to produce reliable estimates of parameters for a small target population in a smaller geographic area (Gosh & Rao, 1994; Rao & Molina, 2015; Gosh, 2020). Over time, several ways of operationalizing a small-area estimation model have emerged, such as regression or synthetic estimates (Noble, 2010; Rao & Molina, 2015). The fact is that SAE transcends the limitations of aggregate-level analyses, providing differentiated insights into specific geographic areas or subpopulations (Gosh & Rao, 1994; Pfeffermann, 2002).

In a business context, identifying and capitalizing on market opportunities is essential to sustaining business growth (Goodman, 1993; Harrison & Pelletier, 2000). As companies strive to make decisions aligned with the characteristics of a given market in a specific and smaller location, more than market data that considers generalization may be required (Noble, 2012; Moura et al., 2017; Gosh, 2020). SAE offers an exciting avenue for companies to delve deeper into localized market dynamics and unlock untapped opportunities (Rao & Molina, 2015; Gosh, 2020).

By leveraging SAE in their analysis, companies, significantly those dependent on the socioeconomic situation of the region in which they are located, can gain a deeper understanding of market potential, consumer behaviors, and demand patterns at the micro level (Moura et al., 2017; Rao & Molina, 2015; Ndofirepi, 2021). This precision allows for creating personalized strategies and appropriate allocation of resources (Ndofirepi, 2021; Delis et al., 2023).

However, despite its potential, the integration of SAE into business decision-making processes remains an area with little explored potential (Gosh, 2020; Delis et al., 2023) compared to public policies. Due to a lack of knowledge and resources, many companies use general market research data at a higher level, ignoring regional variations in micro-regions (Noble, 2010; Rao & Molina, 2015; Gosh, 2020). The true power of SAE lies not only in its ability to provide insights but also in guiding companies in making evidence-based decisions (Pfeffermann, 2013; Rao & Molina, 2015; Gosh, 2020; Ndofirepi, 2021) that reflect the diversity inherent in localized markets.

This paper analyzes and explores the connection between small-area estimation and business opportunities. By investigating the intersections of SAE methodologies, data sources, decision frameworks, and case studies, this paper aims to shed light on the transformative role that SAE can play in equipping companies with the tools they need to seize localized market opportunities. We divided this paper into two parts: the first presents a systematic review of the literature to investigate what has already been studied by academia, within and outside the business perspective, and to identify possible gaps. The second paper aims to test the conceptual model presented for academic and practical investigation of the use of SAE. Based on a chosen market sector, the second paper presents compelling insights that have inspired companies to harness the potential of SAE, driving them towards strategic excellence.

II. FIRST PAPER: LITERATURE REVIEW

Small Area Estimation and Business Opportunities: A Systematic Literature Review and Research Agenda

ABSTRACT

Small area estimation is a subarea of statistics aimed at obtaining detailed and accurate information about the characteristics of a particular geographic region or specific area. Even though the available data is limited, this article reviews the literature on small-area estimation and decision-making from the perspective of identifying local business opportunities. The objective is to understand how this relationship happens and how much it has evolved. In this study, we selected by themes and analyzed 556 articles and book chapters imported from the Scopus and Web of Science databases. Over the past 20 years, there has been significant growth in the number of publications involving the use of small-area estimation during decision-making. However, the most significant focus was public policies and social well-being interventions. The work shows that there needs to be more research on using small-area estimation in integration in marketing studies through reliable granular data. This gap is characterized by the lack of studies exploring potential applications in marketing contexts and the scarcity of methodological frameworks adapted to research needs. As a result, there is a significant opportunity for research investigating how small-area estimation can improve market analysis, consumer behavior modeling, and marketing strategy development, ultimately leading to more targeted and effective initiatives.

Keywords: *Small Area Estimation, Decision Making, Business Opportunities, Market Opportunities, Spatial Marketing.*

1.INTRODUCTION

Making more assertive decisions is both a challenge and a requirement for companies to survive (Goodman, 1993; Harrison & Pelletier, 2000; Vaimann et al., 2012). Given this context, accessing sufficient data to support such decisions can also be challenging and complex (Marchau et al., 2019; Fast & Schroeder, 2020), causing discomfort and tension for marketing managers. In this scenario, small area estimation, or SAE, offers an accurate and detailed approach to analyzing demand and identifying opportunities for creating or growing businesses. By overcoming the limitations of conventional data and providing localized insights into market dynamics (Gosh & Rao, 1994; Rao & Molina, 2015), the use of SAE enables companies to make more informed and effective strategic decisions (Ghosh, 2020), even in competitive and uncertain environments. Therefore, this study focuses on how SAE can assist in managerial decision-making by identifying growth opportunities based on demand analysis.

Statistical data support managerial decision-making; however, they are only sometimes sufficient or available at a more micro-regional level (Rao & Molina, 2015). SAE meets this demand, as it estimates characteristic parameters of a small subpopulation of interest in a small geographic area, such as regions of a municipality (Erciulescu et al., 2021), something that a more extensive survey may not have the strength or resources to do (Gosh, 2020). Academia shows that, in the last twenty years, the use of small-area statistics has gained strength in basically two settings: first, due to the need to formulate assertive public policies that efficiently allocate government resources (Pfefferman, 2002, 2013; Erciulescu et al., 2021; Wang et al., 2012). One example is SAIPE, a US government program to control income and poverty in the population. Second, as support for decision-making in the private sector, especially for those who depend on socioeconomic information (Bickel, 2014; Gosh, 2020), used to explore new local opportunities. However, scholars point out that, historically, the use of SAE has focused more on public policies, government inferences and planning, and resource allocation than on supporting managerial decision-making (Zanutto & Zaslavsky, 2002; Erciulescu et al., 2021; Gosh, 2020). It happens due to the peculiarities of each small region and its specific needs, even though they are part of a whole. One example is the favelas located in Brazilian cities.

Different areas and applications address The literature on SAE (Rao & Molina, 2015). Some of the main findings and research areas in the literature on SAE that are most explored include: economic analysis (Rao & Molina, 2015; Ghosh, 2020; Erciulescu, Franco, & Lahiri, 2021), to estimate key economic indicators such as unemployment rates, poverty levels, income distribution, and GDP growth; social sciences (Abotaleb, I. S., & El-adaway, I. H., 2017; Geng, Z., Wang, Z., Peng, C., & Han, Y., 2016; Barua, S., Abedin, Z., Nath, A., & Biswas, C., 2019), to estimate various social indicators such as educational attainment, health outcomes, crime rates, and housing conditions; environmental studies (Feizizadeh, B et al., 2022; Buganova, K., Luskova, M., Kubas, J., Brutovsky, M., & Slepecky, J., 2021), to estimate environmental variables at local scales, such as air quality, water pollution levels, biodiversity, and land use patterns; and public health (Li X., et al., 2022; Rotejanaprasert C., et al., 2020), to estimate health outcomes, disease prevalence, access to health care, and health care utilization rates at the community or neighborhood level.

Compared to the above mentioned areas, SAE can still be explored outside the marketing literature (Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020). Inference on local trust level (Bickel, 2014), local entrepreneurship (Ndofirepi, 2021), preference testing (Demynck, 2015), and negotiation techniques (Delis et al., 2023). SAE can be a significant tool for identifying opportunities for expansion or business creation at a local level. Its importance stems from the need for sufficient data for decision-making (Gosh, 2020). With a more detailed look at a given region, precise statistical data can be the difference in assertive marketing management decision-making (Gosh, 2020). Thus, this article seeks to answer the following research question: How can SAE help identify business opportunities?

This article analyzes the relationship between small-area estimation, decision-making, and business opportunities through a systematic literature review. We adopted the following phases of a systematic literature review (Snyder, 2019): (i) designing the review, with the reason for the work and its research question; (ii) conducting the review, with its operationalization; (iii) analyzing the results; and (iv) writing the review. We subdivided this article into methodology, results, research agenda, and final considerations.

2.METHOD

The systematic literature review emerges as an approach explored and recommended by different scholars (Snyder, 2019; Paul & Criado, 2020; Donthu et al., 2021) to provide a comprehensive overview of the literature related to a given topic. This methodology is applied in fields such as health (Li X. et al., 2022; Rotejanaprasert C. et al., 2020), economics, and business (Pignone et al., 2005; Martínez-Lopéz et al., 2018), stands out for its ability to contribute significantly to the advancement of research, allowing the synthesis of knowledge from broad domains and complex topics (Paul & Criado, 2020).

We reviewed the systematic literature using the Scopus and Web of Science databases. We followed the PRISMA flowchart, Preferred Reporting Items for Systematic Reviews and Meta-Analysis (Snyder, 2019; Page et al., 2021) to transparently document the steps in the identification, screening, and eligibility stages. With this choice, we aim to ensure transparency and reproducibility of the research.

2.1. Identification and Screening

We adopted the following research steps, adapted from works on systematic literature reviews (Booth et al., 2016; Fink, 2019; Snyder, 2019; Page et al., 2021):

1. Definition of a theme and research question, presented in the introduction
2. Definition of keywords
3. Search for keywords in databases

4. Application of initial filters in the databases
5. Export of the databases
6. Insertion, compilation, and exclusion of duplicates
7. Initial exploratory analysis of the results

The search in the databases began with the following criteria: first, keywords in English, based on "title, abstract and keywords": "small area estimation" AND "business opportunity*" OR "market opportunity*" OR "decision making"; second, "topic of interest": business, statistics and decision science; third, "document type": article, conference paper and book chapter; and fourth, "language": English. These four criteria were initial filters for preparing databases suitable for this work. We extracted the databases and compiled them with such preparation, and we removed 102 duplicates, resulting in 556 files. Table 1 summarizes the criteria, identification, and screening steps.

Table 1 - Database search summary

PRISMA Stage	**Criteria**	**Scopus and WOS databases**
Identification	**Search date**	April 15, 2023
	Search by	"title, abstract and keywords"
	Keywords	"small area estimation" AND "business opportunity*" OR "market opportunity*" OR "decision making"
Screening	**Document type filter**	article, conference paper, and book chapter
	Topic of interest filter	business, statistics, and decision science
	Language filter	English
	Total by database	***Scopus*: 514** ***Web of Science*: 135**
TOTAL	**AFTER REMOVAL OF**	**DUPLICATES = 556**

Source: Authors (2023)

2.2. Eligibility

The 556 filtered articles show that the annual average of citations became significant from 1994 onwards, with the first edition of the book "Small Area Estimation" by Gosh and Rao (Gosh & Rao, 1994), as per Appendix I. The average number of annual citations before this date is less than five. Afterward, there was an exciting increase, with significant empirical work from 2012 onwards (Bickel, 2012; Rao & Molina, 2015; Gosh, 2020).

We listed four exclusion criteria while analyzing the 556 previously filtered articles. They are:

(a) area of knowledge: not related to the area of interest (business and management);

(b) relevance: sources of lesser relevance to the research, that is, journals/publishers that do not involve the desired area of knowledge;

(c) methodology: no use of statistics for small areas;

(d) population: without any contact with businesses, entrepreneurs/companies.

Using these criteria, we excluded 309 files due to reason "a," 114 due to reason "b," 60 due to reason "c," and 38 due to reason "d," leaving 35 eligible files. To understand the use of small-area statistics in decision-making involving growth opportunities, we excluded 18 of the 35 files, resulting in a final sample of 17 papers in Table 2. The exclusion of the 18 files was due to the relevance and methodological robustness of each article. We considered the relevance of the study content about the study topic, ensuring that it directly addressed the use of SAE in decision-making related specifically to some opportunity. We also assessed the methodological quality of each study, taking into account the precision of the statistical methods used, the validity of the data used, and the clarity in the presentation of the results. We considered the geographic and sectoral diversity of the studies, seeking to include research that addressed a variety of contexts. It gave us a more comprehensive understanding of using SAE in different organizational settings and environments. At the end of this process, we sought to identify common patterns and trends across the studies and extract meaningful insights that could inform practices and strategies used. Our analysis aims to provide a comprehensive view of how SAE can be applied effectively to drive opportunity identification.

3.RESULTS

3.1. Descriptive analysis

To deepen our analysis, we synthesized relevant aspects of the 17 studies selected from the sample based on criteria established in the literature (Paul & Criado, 2020; Donthu et al., 2021). To achieve this objective, we identified and delimited three significant themes that guided the scope of this work: the relevance of SAE, the type of contribution, and the application and focus on market opportunities. We chose these fundamental themes to provide a clear and comprehensive conceptual framework for our analysis and allow a deeper understanding of the practical and strategic implications of using SAE in identifying and exploiting business growth opportunities.

Table 2 shows the works closest to the focus of exploring the bias of local growth opportunities in conjunction with the use of SAE. The main focus of small-area statistics is to assist in complementing and improving government databases. The most common use of the technique is due to the need for micro-granular information about the regions of the countries. Therefore, the use of SAE mainly helps in government decision-making.

Table 2 - Summary table in chronological order

Work	Relevance (SAE)	Contribution	Application	Business focus
Ren & Bidkhori (2023)	Consistency of statistical estimation techniques within programming	Use of statistical techniques that help complete incomplete databases	public sector	-
Delis, Kazakis & Zopounidis (2023)	Bayesian statistics prove the proportional impact between the two variables	Focus on Bayesian theory, showing how the relationship between negotiations and good management can be influenced	private sector	yes
Ndofirepi (2021)	Analysis of spatial contexts that influence entrepreneurship	Analysis, with a small sample, to identify local factors that influence local entrepreneurship	private sector	yes
Erciulescu, Franco & Lahiri (2021)	Use of SAE in public policies, in practice	Use and importance of using SAE for public sector records	public sector	-
Ghosh (2020)	Advances in the use of SAE	Update on advances in the use of SAE and its main application techniques	both	-
Rao & Molina (2015)	All about SAE, basic book	Explanation of the concept of SAE and its main forms of application, considering different statistical techniques	-	-
Demynck (2015)	Individual preference tests for decision-making	Comparison between statistical inferences, with little focus on SAE	private sector	-
Bickel (2014)	Statistical inference for local confidence level	Using Bayesian statistics to infer the confidence level of a small region	public sector	-
Pfeffermann (2013)	Advances in the use of SAE	Update on advances in the use of SAE and its main application techniques	both	-
Bickel (2012)	Comparison of frequentist and Bayesian regarding the degree of caution	Little application of SAE. The main objective is to show the difference between the use of Bayesian and frequentist statistics	private sector	-
Giang & Shen (2005)	Probabilistic insights to aid decision-making under uncertainty	Using statistical analysis based on the uncertainty of a region to help government decision-making	public sector	-
Petruccie Salvati (2004)	Modeling spatial correlation between small-area random effects	Use and importance of SAE in studies on a lake in the USA (watershed)	public sector	-
Pfeffermann (2002)	Advances in the use of SAE	Update on advances in the use of SAE and its main application techniques	-	-
Zanutto & Zaslavsky	Use of SAE through two data sources: census data and	Use and importance of using SAE for public sector records	public sector	-

(2002)	administrative records			
You & Rao (2002)	Undercover age of the Canadian Census by SAE Data	Use and importance of using SAE for public sector records, focusing on Bayesian application	public sector	-
Baram (1998)	Decision-making based on the indecision of small local areas	Performs a statistical analysis of a small region to define the level of indecision of the population	private sector	yes
Gosh & Rao (1994)	Historical review and contextualization of SAE	Explanation of the concept of SAE and its main forms of application, considering different statistical techniques	-	-

Source: Authors (2023)

3.2. The Importance and Use of SAE

According to Gosh & Rao (1994), small-area statistics began in the 11th century, based on censuses or administrative records aiming at complete enumeration. Small-area estimation arose from the need for reliable estimates for subpopulations or small geographic areas when direct survey data were limited or unavailable (Gosh & Rao, 1994; Erciulescu et al., 2021). Since then, the methods and techniques used in small-area estimation have evolved and expanded to meet challenges and incorporate advances in statistical modeling, data integration, and spatial analysis (Pfeffermann, 2002 and 2013; Gosh, 2020).

Although the exact origin of the term "small area estimation" is not well documented, it gained prominence as a field of study in the statistical literature in the late 1980s and early 1990s (Gosh & Rao, 1994; Gosh, 2020). The book "Small Area Estimation" (Gosh & Rao, 1994) was a landmark in the field, defining concepts and methodological practices later refined by other authors. Over time, researchers and statisticians have recognized the importance of obtaining accurate estimates for small areas to inform decision-making, policy development, and resource allocation at a more localized level.

Gosh (2020) emphasizes that the growing demand for the use of SAE is not only in the public sector but also in the private sector. The field continues to evolve as new challenges and opportunities emerge in estimating parameters or characteristics for small areas using a combination of survey data, ancillary information, and advanced statistical modeling techniques (Pfeffermann, 2013; Gosh, 2020).

To properly understand SAE, it is essential to delineate the differences between SAE and estimates of larger areas, as well as the primary methodologies and barriers of SAE. The differences between SAE and significant area estimates vary according to the context and scale of estimation (Gosh, 2020). Table 3 summarizes the distinction between the concepts (You & Rao, 2002; Gosh & Rao, 1994; Rao & Molina, 2015).

Table 3 - Distinction between small-area and large-area estimates

Considerations	Small Area Estimates (SAE)	Large Area Estimates
Scope of analysis	Smaller areas located within larger regions, such as a neighborhood in a municipality	Larger geographic areas, such as states, countries, or even global regions
Sample size	Specific and limited	Broad and general
Accuracy and	Very accurate and granular	Not very precise and granular

granularity		
Data sources	Use auxiliary data sources, such as information from larger domains	Greater focus on collection via sampling, with comprehensive data
Statistical models	They incorporate models that take into account spatial or temporal dependencies and explore the relationships between auxiliary variables and the target variable	They use more direct models that focus on representative sampling and standard statistical inference

Source: Authors (2023)

The boundaries between SAE and large-area estimation are only sometimes well defined, and there may be overlap or hybrid approaches, depending on the context and specific research objectives (Rao & Molina, 2015). The choice between SAE and large-area estimation depends on the scale, accuracy requirements, available data, and the research or policy questions. However, we can infer that using SAE is more suitable for smaller spatial analyses (Gosh, 2020), which include the collection and improvement of detailed databases.

Estimating characteristic parameters of a small subpopulation of interest in a small geographic area (Rao & Molina, 2015; Gosh, 2020; Erciulescu et al., 2021) has two major fundamental questions (Pfeffermann, 2002): how to produce the research best, and how to assess the estimation error/uncertainties. The first issue concerns the funding side of the research and the types of data we could obtain. Few surveys are large enough to provide subpopulation estimates for all regions of a given country to the point of obtaining reliable data (Rao & Molina, 2015; Ghosh, 2020); for example, data from a demographic census. Depending on the level of information sought to obtain, only some surveys cannot represent data from a population in a granular and total way (Pfeffermann, 2013; Ghosh, 2020). In addition, they are often carried out over a long interval, highlighting significant gaps for intermediate points in such a period (Rao & Molina, 2015; Pfeffermann, 2022).

The second question focuses on the practice of the different methodologies and how much they have evolved (Pfeffermann, 2002; Rao & Molina, 2015; Gosh, 2020).

Regarding the different methodologies, we can divide them according to the type of inference, whether design-based, model-dependent (frequentist or Bayesian), or a combination of both, such as synthetic application (You & Rao, 2002; Pfeffermann, 2002). Both frequentist and Bayesian design methods in SAE seek to improve the accuracy and reliability of SAE by incorporating auxiliary information and optimizing the sample design (Rao & Molina, 2015; Pfeffermann, 2013). In the frequentist approach, design methods involve building unbiased or efficient estimators for small areas based on sample surveys and auxiliary information (Rao &

Molina, 2015; Bickel, 2012). The Bayesian approach combines prior distributions with the likelihood function derived from the sample data to obtain posterior distributions of the parameters for each small area (Rao & Molina, 2015; You & Rao, 2002).

Unlike the frequentist and Bayesian approaches, applying the synthetic control method is the option when direct data for that area are unavailable or limited (Pfeffermann, 2013; Rao & Molina, 2015). It allows researchers to obtain reliable estimates for specific areas or subgroups using data from similar areas. This approach involves constructing a synthetic control area or group by combining information from similar areas or groups with more complete or reliable data (Rao & Molina, 2015). Choosing between these approaches depends on the available data, the nature of the problem, spatial or temporal dependence, and the underlying assumptions and preferences of the research itself and its researchers (Gosh, 2020).

3.3. SAE vs. Traditional Spatial Statistics: Methodological Differences

Due to some peculiar characteristics, SAE is a distinct methodological approach from traditional spatial statistics (Gosh & Rao, 1994; Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020). SAE methods address the unique challenges of estimating parameters from small geographic areas and provide more reliable and detailed insights into local trends and phenomena (Rao & Molina, 2015; Gosh, 2020; Erciulescu et al., 2021). In more depth, this section aims to present the distinctive characteristics that define and differentiate SAE, investigating its distinctive methodologies and highlighting the specific ways in which it diverges from the conventions of traditional spatial statistics. We will also consider the distinction presented in Table 3.

One of the main distinctions is the focus of SAE on smaller geographic units, such as neighborhoods or census tracts (Gosh & Rao, 1994; Pfeffermann, 2002; You & Rao, 2002; Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020)—granularity, or precision at the local scale. While traditional spatial statistics often deal with larger spatial entities (Rao & Molina, 2015; Ghosh, 2020), such as regions, SAE's emphasis on granularity facilitates a thorough examination of localized phenomena, such as neighborhoods or census tracts. This granular focus allows a more detailed and accurate analysis of localized data trends (Ghosh, 2020). For example, when analyzing the income distribution of a particular city, traditional statistics can provide a general overview of the income pattern at the city level. However, SAE allows for a more refined analysis (Gosh & Rao, 1994; Rao & Molina, 2015), identifying discrepancies in income distribution between specific neighborhoods.

It is particularly relevant in urban areas, where socioeconomic characteristics can vary significantly from one neighborhood to another. Furthermore, the local-scale accuracy offered by SAE can be crucial for planning and decision-making at the municipal or regional level (Zanutto & Zaslavsky, 2002; Demynck, 2015; Erciulescu et al., 2021). For example, government authorities can use SAE estimates to identify areas with the greatest need for investment in infrastructure or public services (Gosh & Rao, 1994; Rao & Molina, 2015; Erciulescu et al., 2021). Likewise, companies can use this information to segment markets more precisely, adapting their marketing and distribution strategies to meet the specific needs of each area (Rao & Molina, 2015; Ghosh, 2020).

Incorporating auxiliary information is another essential feature of SAE (Gosh & Rao, 1994; Pfeffermann, 2002; Rao & Molina, 2015), addressing the challenge of limited direct survey data in small areas. Traditional spatial statistics often face the constraints of sparse data, but SAE strategically integrates supplementary information, whether from detailed demographic data, administrative records, or even satellite imagery (Gosh & Rao, 1994; Rao & Molina, 2015). These supplementary data help overcome problems associated with small sample sizes, contributing to more accurate estimates for areas with sparse survey data, especially in areas where direct survey data are scarce or nonexistent. For example, when estimating the unemployment rate in an urban neighborhood, SAE can supplement household survey data with detailed demographic information, such as population composition by age group and educational level, obtained from government sources.

Integrating satellite imagery can provide valuable insights into geographic and environmental characteristics (Rao & Molina, 2015) that influence phenomena of interest, such as population density or land use. By incorporating this ancillary information into SAE models, researchers can obtain more accurate and comprehensive estimates, even in areas where direct survey data are scarce. Combining survey data with ancillary information is one of the main advantages of SAE, allowing a more complete and reliable analysis of small geographic areas (Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020). By overcoming the limitations of traditional survey data, SAE offers a more detailed and accurate perspective of local dynamics, making it an exciting tool for various research and planning applications.

Another critical characteristic of SAE that distinguishes it from others is hierarchical modeling (Rao & Molina, 2015). SAE recognizes and adapts to the hierarchy inherent in spatial data, in which observations can go up into larger spatial units (Gosh & Rao, 1994; Rao & Molina, 2015). Hierarchical models allow borrowing strength from neighboring areas, contributing to more excellent reliability in parameter estimation for small areas (Rao & Molina, 2015). This hierarchical approach is essential to capture the data's spatial structure and incorporate the spatial dependence between observations. By recognizing the existence of geographic clusters and the mutual influence between neighboring areas, hierarchical models enable a more accurate analysis of local trends and a better representation of spatial variations in the phenomena of interest (Gosh & Rao, 1994; Rao & Molina, 2015).

Finally, unlike traditional models, SAE accommodates the unique challenges of small-area estimation. This peculiarity allows SAE to capture the fine details of local trends (Rao & Molina, 2015), incorporating factors such as spatial autocorrelation and temporal dynamics. The complexity of these models is essential to produce accurate and reliable estimates in the context of small-area estimation. Furthermore, the flexibility of SAE models allows the incorporation of different sources of uncertainty, such as sampling error and modeling error, increasing the robustness of the estimates (Rao & Molina, 2015). This integrated approach to dealing with multiple sources of uncertainty is crucial to providing a comprehensive assessment of the reliability of SAE estimates (Gosh & Rao, 1994; Baram, 1998; Pfeffermann, 2013;) and adequately informed decisions based on these results.

In summary, we can infer that the distinctive features of SAE, including its granularity, incorporation of auxiliary information, use of hierarchical models, and development of custom estimation models, collectively contribute to its effectiveness in providing differentiated insights into local trends and phenomena (Gosh & Rao, 1994; Pfeffermann, 2002; Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020). Understanding these methodological distinctions is essential for researchers and practitioners seeking to employ SAE in contexts where traditional spatial statistics may be insufficient. Through this exploration, researchers and practitioners gain insight into SAE's unique capabilities, with a different approach to spatial data analysis in contexts where traditional methods are insufficient.

3.4. Identifying Market Opportunities and Making Decisions from Small Data

Pfeffermann (2002) exemplifies the latent growth in demand for SAE by citing that many statistical agencies have focused on improving this methodology, mainly due to countries seeking to base their future censuses on registration systems. An example is the SAIPE program of the US Census, whose objective is to map the income and poverty levels in the country's counties (SAIPE, 2023). Given the goal of complementing their databases, these countries recognize the flaws and inaccuracy of their current administrative data, making it necessary to collect information at a smaller census level (Gosh, 2020). Policymakers increasingly demand access to reliable data from smaller areas, such as age group, gender, race/ethnicity, poverty, and education status, for use in political decision-making (Rao & Molina, 2015). Politicians and government officials need data on school districts, health service areas, non-governmental organizations, and private organizations (Rao & Molina, 2015; Ghosh, 2020).

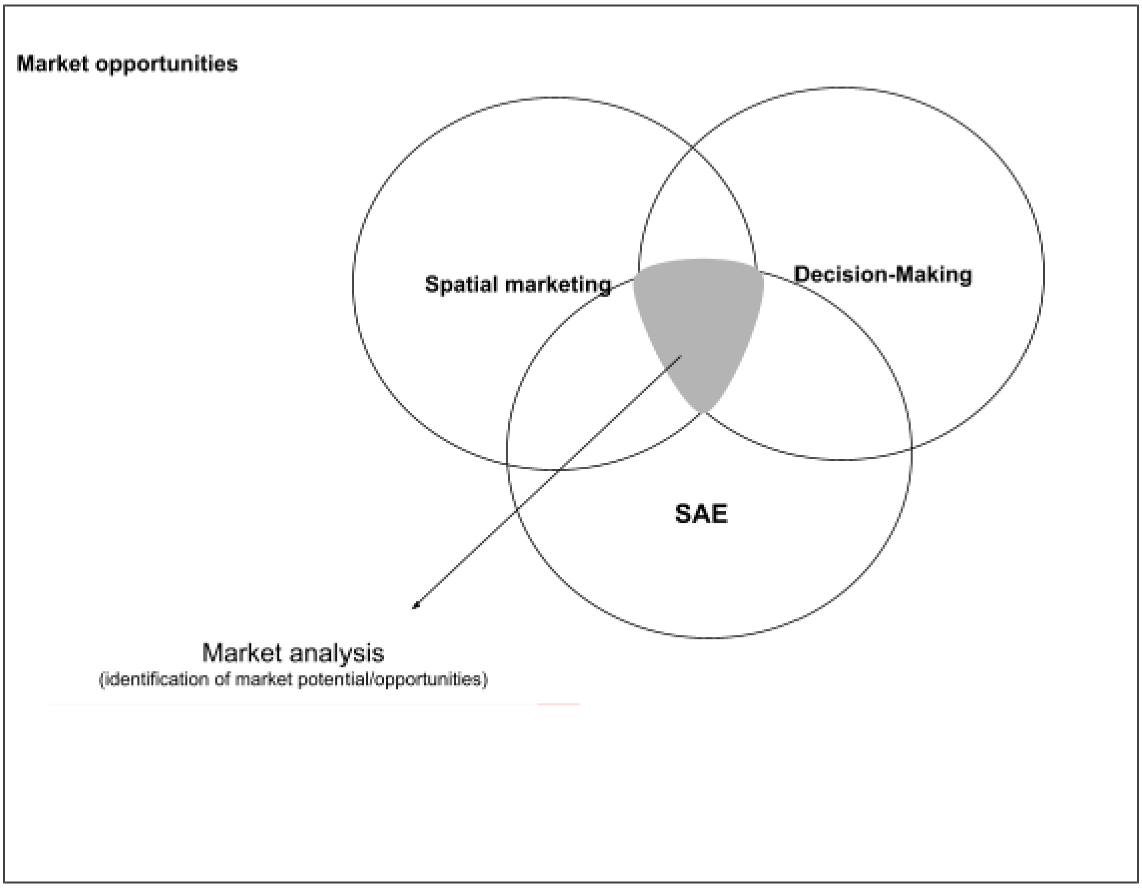
In the public sector, concerns about the distribution, equity, and disparity of subgroups/sub-regions in various areas, such as income, education, and socio-environmental conditions, make the demand for the use of such a practice even more latent (Gosh & Rao, 1994; You & Rao, 2002; Petrucci & Salvati, 2004). We understand that the benefits of its use for the implementation of public policies stimulate social well-being, one of the objectives of any government. There is an excellent exploration of small-area estimation methods amid government policy decisions (Pfeffermann, 2002; Rao & Molina, 2015) due to the need for effective population reach by public policies that aim at social well-being. However, SAE is less explored in a more market-oriented context, as we saw in Table 2, although it may still be as relevant as it is for the public sector.

In the private sector, SAE is highly interested in specific characteristics or behaviors on a small geographic scale, in which we can insert the target audience and the organization. This approach allows a more precise understanding of consumer needs and preferences at a local level, enabling companies to adjust their marketing strategies and develop product/service offerings that are more aligned with market demands. In addition, SAE estimates influence not only potential customers' decisions regarding the adoption of the organization's products/services (Baram, 1998; Demuynck, 2015) but also internal management decisions, such as policy formulation, expansion and business strategies (Ndofirepi, 2021; Delis et al., 2023).

In this way, spatial marketing strategies can be tailored more informedly, adapting to the particularities of specific geographic areas and maximizing their market reach and effectiveness (Gosh, 2020). For example, a company can identify regions with high potential demand for its products and focus its marketing efforts on these areas, thus optimizing the return on marketing investment. Furthermore, the interconnection between decision-making, SAE, and marketing (Rao & Molina, 2015; Gosh, 2020; Delis et al., 2023) enables a more accurate and detailed assessment of business growth opportunities. This granular market analysis allows companies to identify unserved market niches, understand competitive dynamics in different regions, and adapt their business strategies according to the specific needs of each location.

This direct and relevant link with spatial marketing stands out as one of the critical points of this study, highlighting the potential of SAE estimates to inform and improve companies' marketing and decision-making strategies in the current competitive environment. The search for market opportunities is a central problem for organizations due to the importance of adapting to the complex and dynamic market scenario (Gruber et al., 2008). Assertive decision-making in the face of promising growth opportunities is the ideal scenario. At the same time, it also causes great tension due to its essential weight in the survival of companies (Goodman, 1993; Harrison & Pelletier, 2000; Vaimann et al., 2012). It is only possible to remain in the market by innovating and embracing meaningful opportunities presented in a market scenario (Fast & Schroeder, 2020).

Figure 1 - Connection between fields of study



Source: Authors (2023)

Knowing where to get information from and how to operationalize this search for relevant data to identify market opportunities is necessary to make assertive decisions (Gruber et al., 2008; Miller et al., 2007). SAE can be essential in assessing and understanding market opportunities by providing granular information on demographic characteristics, potential demand, and consumer behavior (Ghosh, 2020; Delis et al., 2023).

Identifying market niches is one way to detect opportunities (Abell, 1980; Gruber et al., 2008). By estimating consumers' characteristics, preferences, and needs in different small areas, companies can discover specific niches that may have unique requirements, allowing them to develop specialized products, services, or marketing approaches to serve these markets (Rao & Molina, 1994 and 2015). Additionally, market potential and demand estimates at the small-area level assist companies in identifying ideal locations for opening or expanding their businesses. It helps companies strategically position themselves to capitalize on market opportunities in specific geographic areas (Ghosh, 2020; Ndofirepi, 2021).

SAE can also assist in identifying areas suitable for geographic expansion (Rao & Molina, 2015) and market entry and expansion strategies. Companies can assess market potential and the feasibility of expanding into specific geographic regions by analyzing small-area estimates. It includes assessing factors such as market size, competitive landscape, consumer preferences, unmet demands, and purchasing power (Gruber et al., 2008) at the small-area level. SAE assists in making data-driven decisions about which areas offer the most promising opportunities for growth and expansion. It aids in strategic planning and entry or expansion for maximum impact (Ndofirepi, 2021). Such planning will directly affect assessing business ideas' viability, securing financing, and developing realistic growth projections.

Regarding market segmentation, using SAE allows obtaining information about different small areas or subpopulations with distinct characteristics (Rao & Molina, 2015). Using SAE, companies can identify segments with unique needs, preferences, or behaviors. This knowledge allows the customization of marketing strategies, product offerings, and messages to better align with the specific requirements of each tiny area or subpopulation (Gruber et al., 2008; Rao & Molina, 2015).

Another important point in using SAE to identify opportunities is those that infer about the future. Identify small areas' emerging trends and demographic changes (Ndofirepi, 2021). By analyzing small area estimates over time, companies can detect changes in consumer behavior, demands, market dynamics, or demographic composition in specific geographic regions. It allows companies to respond promptly to evolving market conditions and capitalize on emerging opportunities (Delis et al., 2023; Ndofirepi, 2021).

In short, small-area estimation provides valuable information for companies to identify and evaluate growth opportunities at a localized level. By understanding consumers' characteristics, preferences, and behaviors in different small areas (Ndofirepi, 2021), companies can develop targeted strategies and make informed decisions based on location, in addition to allocating resources effectively to capitalize on market opportunities and drive business growth (Ghosh, 2020; Delis et al., 2023).

4. RESEARCH AGENDA

The previous sections have highlighted the untapped potential for companies to gain competitive advantage through localized insights. We know that as the field of SAE continues to be explored, some key questions deserve further investigation

4.1. Theory

There are unanswered theoretical questions about the use of SAE, specifically in the context of marketing and business opportunity identification:

- What other theoretical frameworks can explain how SAE-based insights lead to better strategic decisions and improved business performance?
- What are the critical success factors and best practices for companies leveraging SAE to identify market niches and gain a competitive advantage in their respective industries?
- What role can government policies and initiatives play in promoting using SAE to identify unmet needs?

4.2. Method

Adapting and testing SAE methodologies, explicitly focusing on unmet needs and opportunity identification. Below are some questions that we believe can be answered:

- How can companies ensure the accuracy and reliability of small-area estimates when available data may be limited or subject to measurement errors at the local level?
- How do we know the most assertive methodology to implement SAE in identifying unmet needs?

- How can companies effectively communicate the benefits and potential of SAE to stakeholders (investors and customers) to demonstrate the value of using localized insights in identifying unmet needs?
- How can SAE techniques be adapted to capture the heterogeneity of localized market behaviors and preferences?
- What innovative statistical models can increase the accuracy of small-area estimates for identifying market niches?
- How can SAE methods be integrated with machine learning techniques to increase predictive power in assessing market potential?

4.3. Application

Apply SAE, in addition to the areas widely explored, as per Table 2, with a focus on presenting the benefits and feasibility of such methodology, seeking to answer the following questions:

- How can companies effectively integrate small-area estimates into their decision-making processes to complement their market analysis?
- What are the specific challenges and barriers companies face when adopting SAE to identify unmet needs in a small area?
- How costly is the SAE implementation process for companies/businesses?
- How can companies collaborate with research institutions, statistical agencies, or data experts to gain access to the knowledge and resources needed to implement SAE in their decision-making processes
- How can companies effectively communicate the benefits and potential of SAE to stakeholders (investors and customers) to demonstrate the value of using localized insights in identifying unmet needs?
- How can we increase awareness and understanding of SAE among entrepreneurs and small businesses to encourage greater adoption and use of the technique?
- How can public-private partnerships facilitate the integration of SAE into business decision-making?

5. FINAL CONSIDERATIONS

Our research highlights an attractive opportunity to leverage SAE to identify business opportunities. SAE offers a precise and essential framework for this purpose (Rao & Molina, 2015; Delis et al., 2023; Ndofirepi, 2021). Despite its potential, it is clear that there is a significant gap in the widespread adoption and integration of SAE methodologies in the context of business strategy and marketing (Ghosh, 2020). This gap represents both a challenge and an opportunity for academia, industry practitioners, and policymakers to delve deeper into the applications of SAE and its implications.

Through a systematic literature review, we outline the fundamental advances in SAE, its main applications, and its intersection with marketing (Pfeffermann, 2013; Rao & Molina, 2015; Gosh, 2020), managerial decision-making, and, most importantly, market opportunity identification (Delis et al., 2023; Ndofirepi, 2021). SAE provides localized insights into market dynamics, consumer behavior, and market potential at a granular level (Rao & Molina, 2015; Ghosh, 2020; Erciulescu et al., 2021). However, adopting SAE techniques and integrating small-area estimation in business opportunity identification still needs to be explored in marketing. Naturally, academics and marketing practitioners need to familiarize themselves with the concept of SAE and its potential to provide valuable insights for marketing endeavors.

SAE promises to guide market entry strategies, resource allocation, and risk management (Ghosh, 2020; Ndofirepi, 2021). At the same time, companies that adopt SAE to identify business opportunities can gain a significant competitive advantage in their market strategies. However, one obstacle to its adoption lies in the financial resources and expertise required to implement its methodological techniques (Pfeffermann, 2013).

While SAE offers valuable insights into localized trends and phenomena, it has limitations. One significant area for improvement is its reliance on available data (Gosh & Rao, 1994; Pfeffermann, 2013; Rao & Molina, 2015), which can be scarce or of variable quality, especially at the subnational level. Limited access to reliable data can undermine the accuracy and reliability of SAE estimates (Pfeffermann, 2013; Ghosh, 2020), potentially leading to biased or inaccurate results. Furthermore, SAE models often assume homogeneity within small areas, ignoring the spatial heterogeneity within these regions (Rao & Molina, 2015).This oversimplification may mask significant local variations and nuances, limiting SAE's applicability and resulting in highly diverse or complex geographic settings. Furthermore, SAE techniques require careful consideration and expertise in statistical modeling, which can pose challenges for researchers and practitioners without specialized training (You & Rao, 2002; Rao & Molina, 2015; Ghosh, 2020). These limitations underscore the need for cautious interpretation and validation of SAE results, as well as continued efforts to improve data quality and refine modeling approaches to increase the robustness of SAE methodologies.

Despite this study's comprehensive nature, several limitations should be acknowledged. First, excluding grey literature may have omitted relevant studies that were not indexed in traditional databases but could have provided valuable insights, data, and perspectives on various topics. Furthermore, restricting the search to only two databases (Scopus and Web of Science) and articles written in English may have limited the scope of the literature review, potentially overlooking valuable information published in other languages or specialized repositories. Although we tried to include a wide range of studies, the possibility of publication bias cannot be completely ruled out. Future research could address these limitations by conducting more extensive searches across multiple databases, including grey literature sources, and considering studies published in languages other than English to ensure a more comprehensive literature review.

In summary, although the widespread use of SAE in business opportunities has yet to be realized, companies have substantial potential to benefit from its application. Collaboration between researchers, statisticians, and companies is critical to bridging the gap between SAE and business opportunities. Raising awareness, addressing data availability challenges, streamlining methodologies, and fostering collaboration can help bridge the gap and empower companies to leverage SAE to identify and pursue promising business opportunities effectively.

III. SECOND PAPER: EMPIRICAL PAPER

After preparing the systematic literature review in the first paper, it was possible to identify the theoretical framework for contextualizing and understanding the main concepts and paradigms related to SAE and an evident gap in the use in business opportunities. Thus, the following study explores the methodology with an empirical test involving SAE, unmet needs, and opportunity identification.

Using Small Area Estimation to Predict Demand

ABSTRACT

Small area estimation provides detailed data on geographic microregions; although the available data is limited, combining several databases is necessary. This study investigates the application of small area estimation as a strategic tool for business expansion or creation, identifying opportunities in Brazilian microregions in the electrical products sector. Using data from the Household Budget Survey (POF) and other macroeconomic sources, the analysis revealed distinct demands for different economic classes in São Paulo city. Small area estimation demonstrated its ability to make inferences about consumption patterns, even without a complete and single database. The results highlighted the importance of aligning marketing strategies according to the socioeconomic context of the microregion, validating small area estimation as an efficient tool for identifying market opportunities, and contributing to even more assertive and informed decision-making. This was converted into a demand index by microregion of the city and by socioeconomic class. Furthermore, the approach presented can be replicated in other sectors addressed by POF, expanding the impact of strategies based on localized data.

Keywords: *Small Area Estimation, Microdata, Market Opportunities, Business Strategy, Spatial Marketing.*

1. INTRODUCTION

The ability to identify and seize localized market opportunities is essential for the growth and success of companies in the face of competitiveness and market dynamism. To this end, strategic analyses using sufficient data become even more necessary (Noble, 2012; Rao & Molina, 2015; Moura, Neves, & Silva, 2017). Unlike aggregated data, which provide summary statistics for entire populations or groups, microdata contains detailed information about individual units, such as individuals, households, firms, or other entities (Rao & Molina, 2015; Gosh, 2020; Ndofirepi, 2021). They play a vital role in understanding various aspects of the economy and provide valuable insights (Pfeffermann, 2013; Gosh, 2020; Erciulescu, Franco, & Lahiri, 2021; Ndofirepi, 2021), allowing researchers to analyze economic trends and understand individual and household behavior. These possibilities help study business dynamics and forecasting studies of business dynamics and forecasts of demand for goods and services (Gosh, 2020; Delis, Kazakis & Zopounidis, 2023).

Small area estimation (SAE) has emerged as a powerful microdata statistical tool that enables researchers and practitioners to gain insights into specific geographic regions or subpopulations (Rao & Molina, 2015; Gosh, 2020). Through it, it is possible to make evidence-based decisions, such as the demand for electrical products, a growing market niche, and technological transformation due to changes in consumer preferences towards cleaner and more sustainable solutions (Rodrigues, 2007; Baran, 2012; Fast & Schroeder, 2020; Costa, Conceição, Silva & Conceição, 2021). SAE transcends the limitations of aggregate-level analyses, offering a granular perspective with the potential to reshape strategic decision-making and resource allocation strategies (Pfeffermann, 2013; Gosh, 2020). Despite the various types of SAE applications in the literature, such as in the healthcare area (Li X. et al., 2022; Rotejanaprasert C. et al., 2020), in mobility and engineering (Abotaleb et al., & El-adaway, I. H., 2017; Geng et al., Y., 2016; Barua et al.; C., 2019) and in sustainability and agribusiness (Feizizadeh, B et al., 2022; Buganova et al.; J., 2021), little is known about the use of small area estimation applied to business strategies.

Using tools such as SAE, microdata analytics usually focus on empirical research and policymaking (Rao & Molina, 2015; Gosh, 2020), providing detailed information about individuals and entities. This application focuses on helping researchers and decision-makers better understand complex social, economic, and demographic phenomena. Although the literature acknowledges that SAE may have potential advantages there may be potential advantages of SAE in guiding market entry decisions, identifying market niches, and optimizing resource allocation, its use is not yet widely explored in this area (Rao & Molina, 2015; Ndofirepi, 2021). Therefore, there is a need for empirical research that evaluates the applicability and effectiveness of SAE-based approaches in business contexts.

This empirical paper aims to fill this gap by testing the importance of using small area estimation in business growth and expansion opportunities and answering the following research question: Does microdata from SAE enable the identification of market opportunities? Therefore, the main objective of this empirical study is to investigate such a relationship through public databases. These connection points aim to complement and boost the use of localized information for strategic excellence and informed decision-making (Fast & Schroeder, 2020; Costa et al., 2021; Erciulescu et al., 2021; Ndofirepi, 2021). More specifically, the study seeks to: a) assess the extent to which localized market data derived from SAE influence the identification and pursuit of sales growth opportunities (demand); b) examine how the integration of SAE-based insights can impact strategic decision-making processes in organizations (marketing + SAE); c) to test the potential of the SAE method in estimating Brazilian demand for electrical products in microregions of the biggest Brazilian city, São Paulo. Furthermore, the findings of this study can inform policymakers, researchers, and practitioners about the value of incorporating SAE techniques into the business toolkit and fostering entrepreneurship in the country.

The choice of the electrical products sector to test SAE was motivated by strategic factors. The availability of data from the Household Budget Survey (POF), which details the consumption of electrical products in Brazilian households, is a central source for this research and facilitates the application of the SAE method, allowing more accurate analysis of purchasing behavior in specific areas. In macroeconomic terms, the electric power sector, which includes products mentioned in the POF, had a positive contribution to Brazil's GDP in the third quarter of 2023, according to data from IBGE and the Ministry of Mines and Energy (IBGE, 2023; MME, 2023). In the accumulated total of 2023, the electricity, gas, water, sewage, and waste management activities sector showed growth of 5.8% (IBGE, 2023).

Furthermore, there is a wide distribution of electrical product consumption in different regions of Brazil, making the use of SAE relevant since this method captures the nuances of demand in microregions. These data show that the electrical materials sector is expanding, which reinforces the importance of adequately estimating its regional demand to identify market opportunities. The Economic Panorama of ABINEE (2024), which highlights the fundamental role of the electrical sector in economic development, also reinforces the relevance of this study. By using SAE, it is possible to deepen this analysis, mapping demand in small areas and adjusting business strategies according to regional specificities. With varied demand in different geographic areas, this market represents a promising field for using methodologies that estimate specific behaviors in locations with little or no information available.

This article aims to empirically validate the integration of small area estimation in business strategies. A theoretical review of the application of SAE in market segmentation strategies, spatial analysis of business opportunities, and data-driven marketing strategies underpins the methodology developed with theoretical concepts and limits. After the theoretical framework, the methodological section describes the empirical process employed to test these theories with accurate data. The methodology adopted describes the structured steps that begin with data acquisition and preparation and continue through to analysis and interpretation of results. Next, we present the results found, discussing the usefulness of SAE as a tool for strategic decision-making in marketing and business planning.

2. THEORETICAL FRAMEWORK

2.1. Small Area Estimation (SAE) and Market Segmentation

Although the exact origin of the term “small area estimation” is not well documented, it gained prominence as a field of study in the statistical literature in the late 1980s and early 1990s (Gosh & Rao, 1994; Gosh, 2020). Small area estimation arose from the need for reliable estimates for subpopulations or small geographic areas when direct survey data were limited or unavailable (Gosh & Rao, 1994; Erciulescu et al., 2021). Since then, the methods and techniques used in small-area estimation have evolved and expanded to address various challenges and incorporate advances in statistical modeling, data integration, and spatial analysis (Pfeffermann, 2002 and 2013; Gosh, 2020). The book “Small Area Estimation” (Gosh & Rao, 1994) was a landmark in the field, delimiting concepts and methodological practices later refined by other authors. Over time, researchers and statisticians have recognized the importance of obtaining accurate estimates for small areas to inform decision-making, policy development, and resource allocation at a more localized level.

The use of small-area estimation is relevant in the public and private sectors (Rao & Molina, 2015; Ghosh, 2020). The field continues to evolve as new challenges and opportunities arise in estimating parameters or characteristics for small areas. Using a combination of survey data, ancillary information, and advanced statistical modeling techniques (Pfeffermann, 2013; Gosh, 2020), SAE allows for more accurate estimates for specific subgroups or regions. Furthermore, the ability to integrate ancillary data from different sources, such as censuses, administrative records, and large commercial databases, broadens the scope of application of SAE in the private sector, particularly in areas such as marketing, resource allocation, and strategic planning (Ghosh, 2020). This potential makes SAE increasingly relevant in a dynamic business environment, where companies seek to adapt quickly to regional and consumer behavior changes (Rao & Molina, 2015).

The differences between small-area estimation and large-area estimates vary according to the context and scale of estimation (Gosh & Rao, 1994; Gosh, 2020). However, we can outline here five converging considerations by the authors on the distinction between the concepts (You & Rao, 2002; Gosh & Rao, 1994; Rao & Molina, 2015): (a) scope of analysis; (b) sample size; (c) precision and granularity; (d) data source; (e) statistical models. About the first two distinctions, SAE has a smaller scope. The areas to be analyzed are smaller, as is the sample size. Therefore, precision will be more accurate and granular in SAE analyses than large-area estimates. About the fourth distinction, SAE uses significant data sources, with auxiliary data for complementation, even if the analysis is micro. Large-area estimates focused more on sampling, with comprehensive data and without supplementary data. Finally, regarding statistical models, SAE incorporates models considering spatial or temporal dependencies and exploring the relationships between auxiliary variables and the target variable. In contrast, large-area estimates use more direct models that focus on representative sampling and standard statistical inference.

The boundaries between SAE and large-area estimation are only sometimes well defined, and there may be overlap or hybrid approaches, depending on the context and specific research objectives (Rao & Molina, 2015). The choice between SAE and large-area estimation depends on the scale, accuracy requirements, available data, and the research or policy questions. However, we can infer that SAE is best suited for limited spatial analyses (Chandra et al., 2012; Ghosh, 2020), which include the collection and improvement of detailed databases.

Estimating characteristic parameters of a small subpopulation of interest in a small geographic area (Rao & Molina, 2015; Ghosh, 2020; Erciulescu et al., 2021) has two major fundamental questions (Pfeffermann, 2002): "How to produce the best way," and "how to evaluate the estimation error/uncertainties.". The first question focuses mainly on the research funding side and the types of data that one wants to obtain. Few surveys are large enough to provide subpopulation estimates for all regions of a given country, for example, to the point of obtaining reliable data (Rao & Molina, 2015; Ghosh, 2020).

The second question focuses on the practice of different methodologies and how much they have evolved, which is the research motivation of some authors in the field (Pfeffermann, 2002; Rao & Molina, 2015; Gosh, 2020). As for the most explored methods, we can divide them according to the type of inference, whether design-based, model-dependent (divided, once again, into frequentist and Bayesian approaches), or a combination of the two (You & Rao, 2002; Pfeffermann, 2002). Design-based inference is derived directly from the sampling plan and is advantageous due to its robustness (Rao & Molina, 2015), as it makes few assumptions about the data. However, although its effectiveness may be limited in areas with few observations, on the frequentist approach, on the other hand, the frequentist approach uses statistical models that associate the variable of interest with auxiliary variables, being efficient but dependent on the adequacy of the model (Pfeffermann, 2013; Rao & Molina, 2015). Besides relying on models, the Bayesian approach incorporates prior distributions to reflect uncertainty in the parameters, which is helpful, especially when there are few observations (Rao & Molina, 2015; Ghosh, 2020). However, it can be sensitive to the choice of these distributions. Finally, combinations of the frequentist and Bayesian approaches emerge as promising alternatives, allowing the integration of the benefits of both techniques to generate more accurate and robust estimates in areas with low data availability (Rao & Molina, 2015).

In market segmentation, SAE plays an essential role by allowing companies to tailor their strategies to specific subpopulations or geographic regions. Traditional market segmentation methods often rely on broad categories that can ignore significant variations in smaller regions (Rao & Molina, 2015; Ghosh, 2020). By using SAE, companies can gain a more granular understanding of their target markets, allowing them to accurately identify micro-segments (Fabrizi et al., 2018; Franco & Lahiri, 2021). This approach increases the effectiveness of marketing campaigns by focusing resources on areas with the highest potential return on investment (ROI), for example.

SAE allows companies to reach beyond broad demographic categories (Chandra et al., 2012; Fabrizi et al., 2018; Ghosh, 2020) that may share similar purchasing behaviors, socioeconomic status, or cultural preferences. For example, a company might use SAE to identify neighborhoods in a city with a higher concentration of high-income (Class A) households, who may be more receptive to premium products or services. Conversely, areas dominated by low-income (Class E) households may be targeted for cost-effective solutions or products that cater to budget-conscious consumers.

Integrating SAE with modern data analytics techniques further amplifies its usefulness in market segmentation (Rao & Molina, 2015; Gosh, 2020). Companies can create detailed market profiles by combining SAE with data from consumer surveys, transaction data, and social media analytics. These profiles can be used to develop personalized marketing strategies that resonate with specific audiences, leading to increased customer engagement and loyalty.

Furthermore, the application of SAE in market segmentation is not limited to consumer markets (You & Rao, 2002; Pfeffermann, 2002; Rao & Molina, 2015). B2B companies can also benefit from this approach by identifying potential business customers in specific regions that match their ideal customer profile. This level of precision in segmentation can lead to a more efficient allocation (Fabrizi et al., 2018; Ghosh, 2020) of sales and marketing resources, reducing costs and improving the overall effectiveness of business development efforts.

In short, using SAE in market segmentation offers a powerful tool for companies looking to optimize their marketing strategies. By providing accurate estimates for small geographic areas, SAE allows companies to identify and target specific market segments with greater precision, ultimately leading to more successful and cost-effective marketing efforts.

2.2. Spatial Analysis of Business Opportunities and Geomarketing

One of the main challenges in this endeavor is understanding the nuances of consumer behavior and market demand at a localized level. Traditional market research methods often must capture these subtleties, leading to missed opportunities and suboptimal decision-making. In this scenario, SAE becomes a strategic tool for geomarketing (Ghosh, 2020). By incorporating detailed spatial data and robust statistical techniques, SAE allows for identifying demand patterns that vary across regions (Rao & Molina, 2025; Ghosh, 2020), providing valuable insights into where and how to invest. It allows companies to optimize product distribution, adjust marketing campaigns to local needs, and seize business opportunities that could be more evident in a more generalized analysis.

According to Rao & Molina (2015) and Gosh (2020), SAE in the private sector is particularly valuable when analyzing specific characteristics or behaviors in small geographic areas. These localized insights can have a profound impact both on a potential customer’s decision to adopt an organization’s products or services (Baram, 1998; Demuynck, 2015; Fabrizi et al., 2018) and on the organization’s internal managerial decisions, such as the expansion of internal policies, strategies, and business opportunities (Fabrizi et al., 2018; Ndofirepi, 2021; Delis et al., 2023). However, knowing the best way to operationalize it can be challenging for managers.

Spatial analysis is a methodological approach that examines geographic patterns to uncover relationships, trends, and correlations in data, particularly those related to specific locations (Cliquet, 2013). This technique allows companies to understand insights into how geographic location impacts consumer behavior, market demand, and competitive dynamics. By employing spatial analysis, companies can effectively map consumer preferences (Cliquet, 2013; Melnyk & Nyzhnyk, 2018), identify areas with high market potential, and optimize the positioning of their products or services.

A significant advantage of spatial analysis is its ability to reveal hidden patterns and relationships that may not be apparent through traditional data analysis methods (Cliquet, 2021). For example, through spatial analysis, a company may discover that certain neighborhoods experience disproportionately high demand for specific products due to demographic factors such as age, income, or lifestyle. Leveraging these insights allows companies to tailor their marketing strategies better to meet the needs and preferences of these target areas better to meet the needs and preferences of these target areas.

In market analysis, spatial segmentation can help decision-makers use small-area estimates to identify areas with high market potential, unmet consumer needs, growth opportunities, and competitive analysis (Fabrizi et al., 2018; Ghosh, 2020). This approach guides decision-making in market expansion, such as selecting optimal locations for new stores, franchises, distribution centers, and other strategic differentiators. The search for market opportunities is a central concern for organizations due to the importance of adapting to the complex and dynamic market landscape (Gruber et al., 2008). Making assertive decisions in response to promising market opportunities is the ideal scenario, but it is also a significant source of stress due to its crucial role in business survival (Goodman, 1993; Harrison & Pelletier, 2000; Vaimann et al., 2012). Innovation and adoption of critical opportunities presented by the market scenario are essential to remain competitive (Fast & Schroeder, 2020).

Geomarketing extends spatial analysis capabilities by integrating geographic data directly into marketing strategies (Cliquet, 2013). This approach uses location-based data to enhance marketing campaigns, refine customer segmentation, and optimize distribution networks. Geomarketing enables companies to create highly targeted and personalized marketing efforts based on the geographic location of their customers.

For example, a retail chain might use geomarketing to analyze the distribution of its customer base across different regions. Through this analysis, the company can identify underserved areas where new stores can be established or adjust its marketing messages to align with specific customer preferences in different locations. This precision in targeting increases the effectiveness of marketing campaigns and contributes to higher conversion rates.

SAE is a statistical tool that can be integrated with geomarketing strategies to increase the accuracy and granularity of market analysis (You & Rao, 2002; Rao & Molina, 2015; Ghosh, 2020). By incorporating SAE into geomarketing efforts, companies can obtain more accurate estimates of market potential at a localized level, enabling more informed decision-making (Cliquet, 2013; Rao & Molina, 2015). For example, a company seeking to expand its presence in a new city can use SAE to estimate the potential demand for its products in various neighborhoods, even without direct sales data for those areas. This approach helps minimize risk and maximize return on investment by focusing on areas with the highest growth potential.

Despite the significant benefits of integration offered by integrating spatial analytics, geomarketing, and SAE, several challenges still need to be solved. A key challenge lies in the availability and quality of geographic data (Fabrizi et al., 2018; Ghosh, 2020). Accurate and up-to-date data are essential for effective spatial analytics and geomarketing (Cliquet, 2021). However, acquiring such data can be particularly challenging in regions with limited infrastructure. Another challenge is integrating multiple data sources, such as demographic, sales, and geographic information (Rao & Molina, 2015). Companies must have the technical expertise and tools necessary to manage and analyze these diverse data sets effectively to manage and analyze these diverse data sets effectively.

Integrating spatial analytics, geomarketing, and SAE represents an approach to identifying and capitalizing on business opportunities. By leveraging these techniques, companies can gain a competitive advantage through a deeper understanding of the unique characteristics of their markets, precise targeting of the right customers, and optimized operations to meet local demand (Ghosh, 2020; Cliquet, 2021). As the business landscape continues to evolve, harnessing the power of location-based data will become increasingly vital to achieving sustainable success.

3. METHOD

3.1. Data-Driven Marketing Strategies and POF (Brazilian Market)

Data-driven marketing strategies have become essential for companies (Gruber et al., 2008; Rao & Molina, 2015) seeking to understand their target audiences, optimize their marketing efforts, and gain competitive advantage. Today's vast amount of data allows companies to make informed decisions based on empirical evidence rather than just intuition. In Brazil, the Household Budget Survey (POF) offers a wealth of data that can be leveraged for this purpose, providing detailed information on household consumption patterns, income distribution, and consumer behavior across various demographic segments (Bazotti et al., 2016; IBGE, 2018).

Data-driven marketing involves using data analytics to guide the creation, execution, and optimization of marketing strategies. By analyzing consumer data, companies can gain a deeper understanding of their customers' needs (Gruber et al., 2008; Fabrizi et al., 2018; Ghosh, 2020), preferences, and purchasing behaviors, allowing them to tailor their products, services, and marketing messages accordingly (de Almeida, 2011; Carvalho & Pereira, 2012). This approach increases the effectiveness of marketing campaigns and improves customer satisfaction and loyalty by providing more relevant and personalized experiences.

In the Brazilian context, data-driven marketing is precious due to the country's diverse and complex market landscape (IBGE, 2018). With significant variations in income, consumption habits, and cultural preferences across regions, a one-size-fits-all marketing approach is unlikely to be effective. Instead, companies should adopt a more granular and targeted strategy that considers the unique characteristics of each market segment. The Household Budget Survey (POF) is a comprehensive household survey conducted by the Brazilian Institute of Geography and Statistics (IBGE). It collects detailed information on the income, expenses, and consumption habits of Brazilian families, making it an essential resource for understanding the socioeconomic dynamics of the Brazilian Market (Silva, 2014; Bazotti et al., 2016; IBGE, 2018). The POF is also crucial for understanding the socioeconomic profile of the Brazilian population and provides granular data on household spending across several categories, such as food, housing, education, and transportation (Silva, 2014; Bazotti et al., 2016). POF data serve as a critical resource for policymakers, researchers, and businesses, offering detailed insights into how different population segments allocate their income. For businesses, POF data are precious and particularly valuable for developing targeted marketing strategies, designing products, and setting prices that align with specific consumer groups the purchasing power and preferences of specific consumer groups (Pintos-Payeras, 2009; Silva, 2014; Vaz & Hoffmann, 2020).

Survey data are often used with other economic indicators to create a detailed picture of consumer behavior in Brazil, allowing for the identification of trends, regional variations, and potential market opportunities (IBGE, 2018). By leveraging POF data, companies can make informed decisions that meet the diverse needs of the Brazilian Market, ensuring that their products and services are relevant and accessible to their target audience. The results of this survey have enabled the creation of inflation indices measured by IBGE, which are extremely important for the country's economy (Silva, 2014), such as the National Consumer Price Index (INPC) and other derived indices, such as the Broad National Consumer Price Index (IPCA). POF data can be instrumental in developing data-driven marketing strategies (Silva, 2014; IBGE, 2018) in several ways, as per Appendix A.

This study uses a mixed-methods approach, focusing on the economic sphere of the consumption pattern of Brazilian households, particularly the demand for electrical materials, specifically in the city of São Paulo. The choice of the electrical materials sector was motivated by strategic reasons already mentioned in the introduction to this work. This industry sector is expanding in Brazil, as evidenced by its positive contribution to the Brazilian GDP in the third quarter of 2023 and by the POF data, which provides a detailed basis on the consumption of electrical products in Brazilian households, facilitating the application of SAE. In addition, using SAE in this sector allows us to capture the variation in demand in different microregions, which is interesting for understanding a market as diverse as that of electrical products in Brazil.

In turn, we choose São Paulo city due to its population representation within the state of São Paulo. It is one of the largest consumers of electrical materials and is Brazil's most significant economic power. This representation means that the analysis focused on São Paulo offers a significant and rich sample to examine demand variations within the sector. In addition, the code developed, which will be detailed later, indicates São Paulo as an area with a high concentration of demand for electrical materials, reinforcing the relevance of this choice for the research.

To achieve this, we used four primary sources of public data: the Household Budget Survey (POF), the IBGE Demographic Census, GeoSampa, and the SAEDE Foundation. Each dataset provides distinct and complementary data on socioeconomic indicators, geographic boundaries, and purchasing patterns, allowing for a comprehensive market demand analysis. There are three spheres of study in the POF: the food sphere, which focuses on understanding household food consumption patterns; the social sphere, which aims to understand the quality of life perceived by households; and the economic sphere, which will be the focus of this work (Brazilian demand for goods and services). The latter analyzes household finances, identifying the source of income and the destination of expenses (IBGE, 2018). It is important to note that the POF data are 12 months long due to the possible seasonality of consumption, income, and expenses throughout the year. Another reason for this duration is its difficulty operationalizing it, given its territorial scope (Silva, 2014; IBGE, 2018). For these reasons, the POF takes place, on average, every 8 years via in-person application of questionnaires. One shortcoming of the POF is the need for the exact location of the respondent. The data user can only reach a state level. Thus, it becomes necessary to add, merge, or complement it with complementary databases to make location inferences, one of the premises of SAE.

To make an empirical application feasible, we selected a specific segment of the Brazilian Market, electrical materials, and a specific municipality, São Paulo. The first provides access to POF data, population census data, and geographic coordinates from the Brazilian census. GeoSampa (2024), an entity in São Paulo, provided data on the municipality's local population mobility and geographic delimitation with its divisions. This database offers detailed information on population mobility in the city of São Paulo, which is essential to ensure the assertiveness and granularity of the results. The SAEDE Foundation (2023) provided socioeconomic data and GDP by social classes and neighborhoods for São Paulo. The ultimate goal is to indicate microregions with potential demand, defining the 20 main specific location points: 10 points with the highest purchase of electrical materials and 10 points with the lowest level.

To ensure clarity and consistency in the analysis, we developed a data dictionary summarizing the main variables selected from each data source. The data dictionary serves as a reference point, detailing the variables used, their descriptions, the measurement scales, and the year of data collection. Table 5 describes the main variables extracted from the IBGE, GeoSampa, and Fundação SAEDE databases. The data dictionary helps optimize the methodological process and ensures transparency in handling diverse datasets. Each entry in the dictionary contains the following columns:

1. Variable code: the specific name of the variable used in the datasets and code
2. Database: the database in which we found the indicated variable
3. Variable description: a brief explanation of what the variable represents
4. Scale: the type of scale applied
5. Collection date: the year in which we collected the data
6. Variable source: the institution that provided the data
7. Context of use: where we used the variable in the analysis or code
8. Data format: the file format in which we stored the data

The complete dictionary is available through the Open Science Framework (OSF)[1], repository, ensuring that all datasets are stored and shared in a standardized and accessible way.

[1]Repository link: https://osf.io/eskdw/?view_only=c9ca494db6dc48498966a4cf9a03dd90

Table 5 - Dictionary of Variables

Variable Code	Database	Variable Description	Scale	Data date	Variable source	Context of use	Data format
UF	collective expenses + individual expenses + resident	codes that represent each Brazilian state	nominal	2018	POF	throughout the code	.cvs
TIPO_SITUACAO_REG	collective expenses + individual expenses + resident	identifies whether the residence is rural or urban	nominal	2018	POF	throughout the code	.cvs
V9001	collective expenses + individual expenses	expense/acquisition type code	nominal	2018	POF	throughout the code	.cvs
V9002	collective expenses + individual expenses	way in which the consumer unit acquired the type of product or service	nominal	2018	POF	exploratory analysis	.cvs
V8000_DEFLA	collective expenses + individual expenses	deflated expense/acquisition value	ratio	2018	POF	exploratory analysis	.cvs
RENDA_TOTAL	collective expenses + individual expenses	total monthly gross income of the consumption unit or the individual in the consumption unit	ratio	2018	POF	exploratory analysis	.cvs
V0403	resident	age in years of the respondent	ratio	2018	POF	from the analysis of the selected UF	.cvs
V0404	resident	respondent's gender	nominal	2018	POF	from the analysis of the selected UF	.cvs
V0405	resident	color or race of the respondent	nominal	2018	POF	exploratory analysis	.cvs
nome_municipio	sexo_idade_sp	name of the cities of the selected UF	nominal	2022	IBGE	from the analysis of the	.xlsx

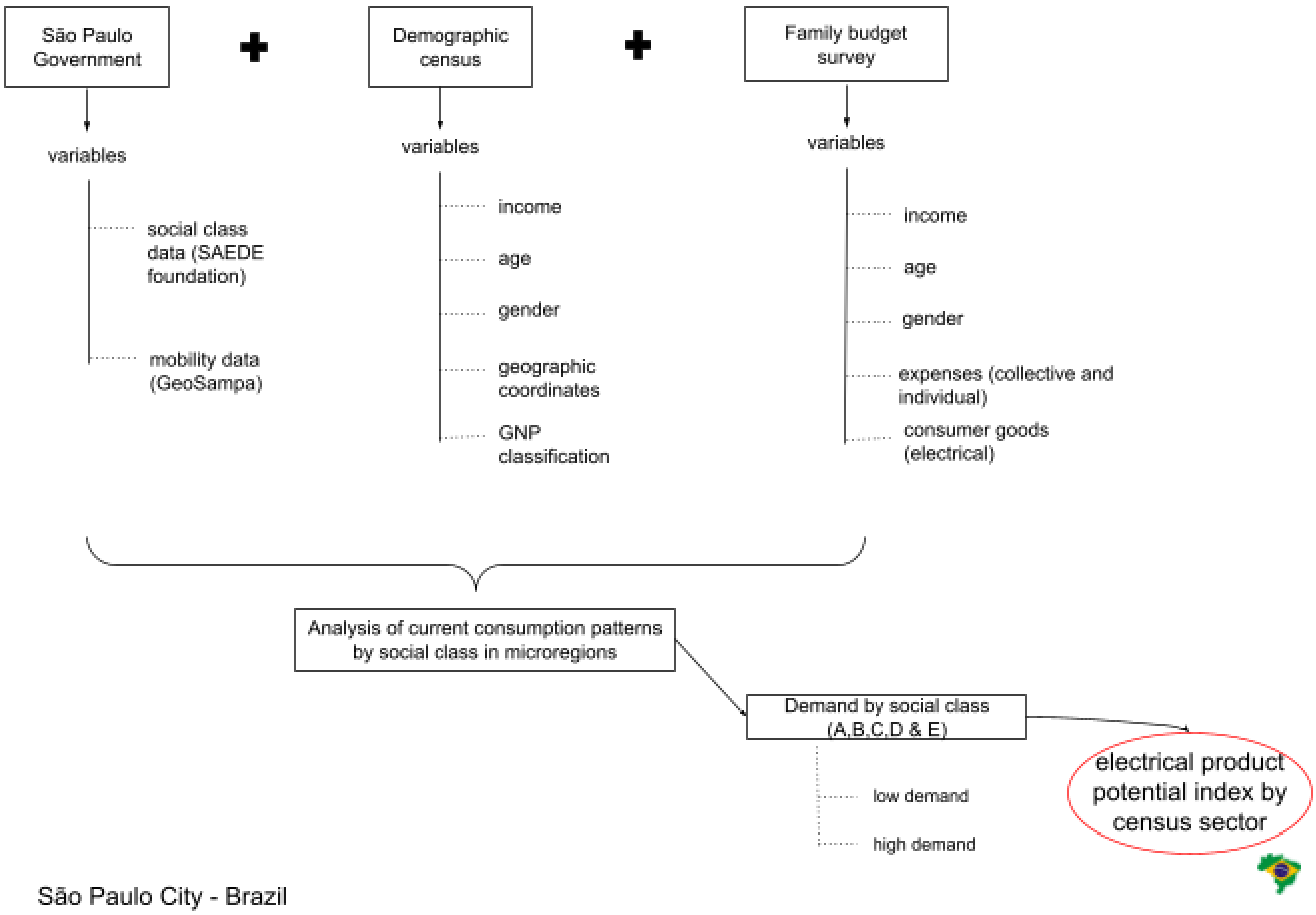
						selected UF	
codigo_municipio	sexo_idade_sp	code of the cities of the selected state	nominal	2022	IBGE	from the analysis of the selected UF	.xlsx
COD_MUN	coordenadas_enderecos_sp	code of the cities of the selected state	nominal	2022	IBGE	from the analysis of the selected UF	.cvs
COD_ESPECIE	coordenadas_enderecos_sp	identifies whether the domicile is a residence or an establishment	nominal	2022	IBGE	from the analysis of the selected UF	.cvs
LATITUDE	coordenadas_enderecos_sp	respondent latitude	nominal	2022	IBGE	from the analysis of the selected UF	.cvs
LONGITUDE	coordenadas_enderecos_sp	respondent's longitude	nominal	2022	IBGE	from the analysis of the selected UF	.cvs
Nome	DISTRITO_MUNICIPAL_SP_SMDUPolygon	name of the regions of the city of São Paulo	nominal	2024	GeoSampa	from the analysis of the selected UF	.shx
Nome	DISTRITO_MUNICIPAL_SP_SMDUPolygon	code of the regions of the city of São Paulo	nominal	2024	GeoSampa	from the analysis of the selected UF	.shx
Nome	pib_bairros_sp	name of the region of the city of São Paulo	nominal	2024	SAEDE	from the analysis of the selected UF	.xlsx
Classe_E_perc	pib_bairros_sp	percentage of class E in the regions of the city of São Paulo	ratio	2022	SAEDE	from the analysis of the selected UF	.xlsx

Classe_D_perc	pib_bairros_sp	percentage of class D in the regions of the city of São Paulo	ratio	2022	SAEDE	from the analysis of the selected UF	.xlsx
Classe_C_perc	pib_bairros_sp	percentage of class C in the regions of the city of São Paulo	ratio	2022	SAEDE	from the analysis of the selected UF	.xlsx
Classe_B_perc	pib_bairros_sp	percentage of class B in the regions of the city of São Paulo	ratio	2022	SAEDE	from the analysis of the selected UF	.xlsx
Classe_A_perc	pib_bairros_sp	percentage of class A in the regions of the city of São Paulo	ratio		SAEDE	from the analysis of the selected UF	.xlsx

Source: Authors (2024)

Before presenting the detailed steps of our methodological approach, it is essential to understand how different data sources converge to build a comprehensive framework for analyzing the demand for electrical products in the selected city of São Paulo/SP. As shown in Figure 2, the study integrates data from the Demographic Census, POF, SAEDE Foundation and GeoSampa. Each of these sources contributes unique variables crucial to the analysis. The demographic census and POF provide sociodemographic and consumption data, offering information on income, age, gender, and geographic coordinates, while SAEDE Foundation and GeoSampa data reveal GNP (Gross National Product) of each neighborhood of São Paulo city patterns of mobility.

Figure 2 - Methodological Procedure



Source: Authors (2024) adapted from Silva (2014)

This data is synthesized in a multi-step process, combining current consumption trends with expected demand, segmented by social class (A, B, C, D, and E). This segmentation is essential to identify microregions in São Paulo/SP with varying demand levels, from very low to very high. Integrating sales and market share analysis across these microregions leads to constructing an electrical product potential index, which census tracts map us. This methodology ensures a geographic analysis of market opportunities, providing a basis for decision-making in business strategy.

A comprehensive approach was adopted to analyze the data, employing various combinations of methodologies to ensure robust and accurate estimates. By integrating multiple data sources and levels, the research leveraged several SAE techniques to address the complexities of the dataset, as outlined below:

1. Direct estimation: allowed direct calculation of purchases and demographic counts.
2. Indirect estimation of demographic data: aimed to improve estimates in regions with limited information.
3. Model-based estimates, mainly through linear mixed models, provided a differentiated analysis by accounting for fixed effects, such as age, gender, and income, and random effects, reflecting area-specific variations.
4. Spatial models: We used these methods later to enrich the analysis by incorporating geographic proximity and spatial dependencies, providing a more accurate and contextually relevant understanding of purchasing patterns across regions. We used the Python libraries Folium and GeoPandas to create interactive maps that display shopping demand across neighborhoods in the selected region. By applying spatial joins, we connected shopping data to specific geographic areas, allowing for a detailed examination of how economic class distribution correlates with purchasing behavior. Heat maps further enhanced this analysis by illustrating high- and low-demand areas.

Collectively, these methodologies have improved the accuracy of estimates and provided a more detailed and localized view of consumer spending. In other words, they have enabled a deeper understanding of market dynamics and supported more informed marketing decision-making.

3.2. Data Collection and Preprocessing

Initially, we conducted an exploratory analysis to investigate the consumption of electrical materials in Brazilian states using the POF - Family Budget Survey - data. We created a Python code in a Jupyter Notebook in Colab.

3.2.1. Loading the initial databases

The first step involved loading a database containing a detailed list of electrical materials into the POF dataset. This fundamental process allowed for selecting specific electrical material codes, which we used throughout the analysis. Next, we integrated two primary datasets: the collective expenditure file, representing household-level expenditure, and the individual expenditure file, which captured resident-level expenditure. Then, we sourced both datasets from POF. They underwent cleaning, validation, and filtering using the electrical material codes identified in the initial stage to provide a comprehensive overview of consumption patterns.

We filtered objective variables for the first two POF files and collective and individual expenses for subsequent analysis. The variables presented in Table 5 were those selected for the study analysis; we removed the others to ensure objectivity and clarity in the results: UF, TIPO_SITUACAO_REG, V9001, V9002, V8000_DEFLA, and RENDA_TOTAL. They were selected and filtered according to the classification of electrical materials, which we performed in the first step of preprocessing. We excluded variables irrelevant to the scope, such as non-electrical goods. We removed Records containing incomplete or inconsistent information to ensure a clean dataset. The goal was to capture a comprehensive picture of electrical material consumption, considering collective household purchases and individual expenses. This process allowed the generation of classification for Brazilian states based on their electrical material consumption, offering valuable data on the dynamics of regional demand.

3.2.2. Narrowing granular analysis

Once Brazilian states were classified, we restricted our analysis to a specific Brazilian state (Federal Unit, UF) for a more granular assessment of electrical materials consumption. This detailed exploration was essential to understanding the unique dynamics of a specific region, allowing us to examine how local economic conditions, demographic characteristics, and infrastructure influence consumption patterns. By focusing on a single state, the analysis can reveal region-specific factors that may need to be evident in broader national studies (Silva, 2014; Rao & Molina, 2015; IBGE, 2018; Ghosh, 2020). Based on the POF data, São Paulo (SP) and Minas Gerais (MG) emerged as the two Brazilian states with the highest electrical materials consumption, making them prime candidates for further investigation. This focus is significant for future research at the small-area level, where localized trends and demands can be better understood. We selected São Paulo city for further analysis, given its significant demand and influence in the national Market.

In the data collection and preprocessing stage, after narrowing the focus to the state of São Paulo, the POF resident database was included to provide data on the sociodemographic characteristics of the state. This dataset is crucial to understanding the population composition of the region, including age distribution, gender, and racial composition. We selected only three critical variables from the database for further analysis, along with the previously merged datasets, we also selected for further analysis. V0403: Represents the respondent's age of the respondent in years, providing insights into the age distribution of the population of São Paulo. V0404: Identifies the gender of the respondent, allowing for analysis of gender distribution. V0405: Captures the race or ethnicity of the respondent, allowing for exploration of racial demographics in the region. These variables are crucial in generating a comprehensive socio-demographic profile for São Paulo, supporting the broader analysis of electrical consumption patterns concerning population characteristics.

We used the most populous city as the determining factor to evaluate an even better geographic scope narrowing. Once again, we included a new IBGE database, which provides data on population density by age and gender for all cities in São Paulo state. When running the analysis, we identified the five cities with the highest population density, which are (in descending order) São Paulo, Guarulhos, Campinas, São Bernardo do Campo, and Santo André. Due to the significant number, we chose São Paulo city as the following scope for analysis, following the premise of finding business opportunities in a small area.

4. RESULTS

4.1 Exploratory Analysis

We conducted statistical analyses, which provided a detailed analysis of household spending on electrical materials, revealing data on consumption patterns in the state. As shown in Figure 3, total expenditure on electrical materials amounted to R$593,573.47, reflecting the aggregate expenditure of all households included in the POF. On average, each household spent R$130.86 on electrical materials, but the median expenditure is slightly lower at R$99.62, indicating that half of the households spent less than this amount. The standard deviation of R$301.82 suggests a wide variation in expenditure, with some households spending significantly more or less than the average. It is further highlighted by the minimum expenditure of just R$2.01 and the maximum expenditure reaching R$18,268.35, illustrating the disparity in consumption. When we check the numbers, the 25th percentile is R$61.09, and the 75th percentile is R$164.35, indicating that the middle 50% of households spent between these amounts. This distribution suggests that while most households have relatively modest expenses, a smaller population is responsible for much larger expenditures. Finally, the data shows a significant urban-rural divide, with 147,163 urban households compared to 30,186 rural households, suggesting that urban areas may drive a larger share of the state's electrical consumption.

Figure 3 - Analysis of Expenses with Electrical Materials in the State of São Paulo

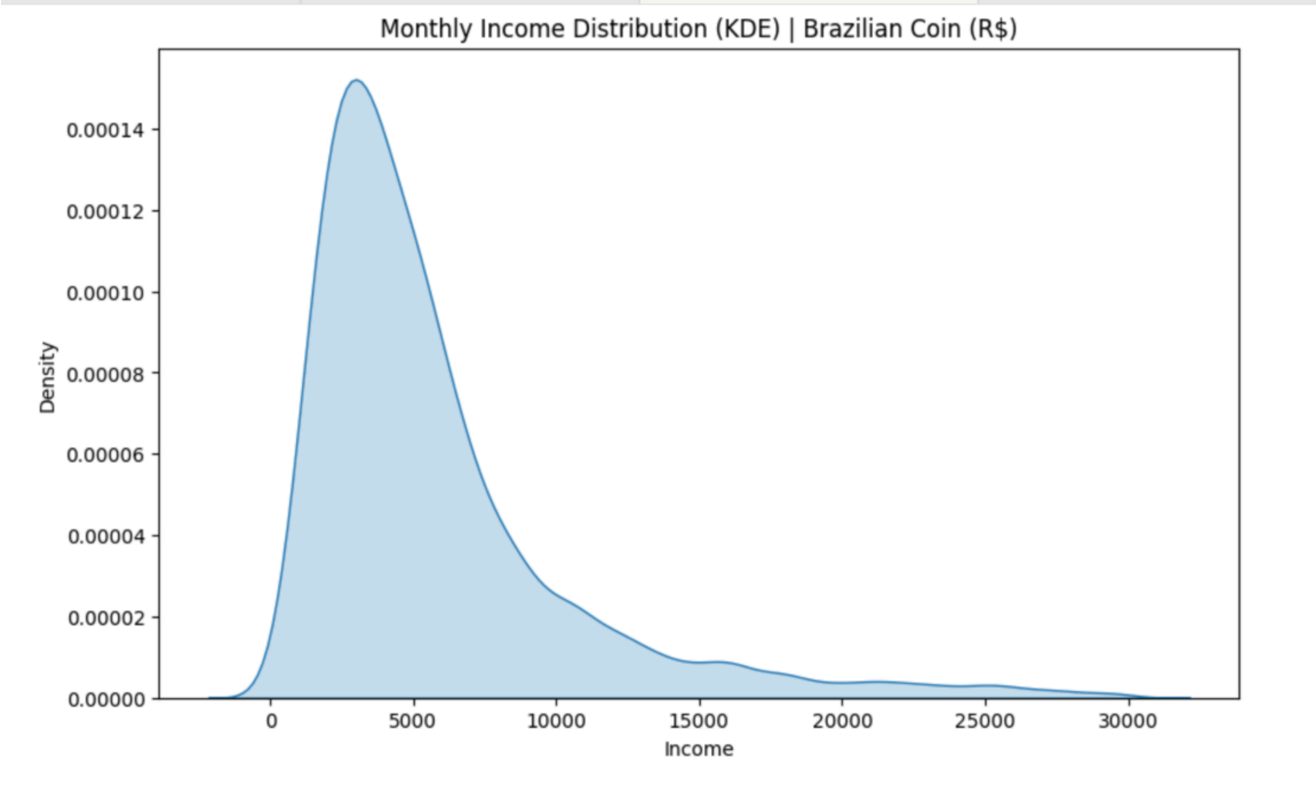
Statistic (São Paulo)	Value (R$)
Total Electrical Material Expenses	593,573.47
Mean Expense per Household	130.86
Median Expense per Household	99.62
Standard Deviation	301.82
Minimum Expense	2.01
Maximum Expense	18,268.35
25th Percentile	61.09
50th Percentile (Median)	99.62
75th Percentile	164.35
Number of Urban Households	147163
Number of Rural Households	30186

Source: Authors (2024).

Analyzing income distribution, electrical material purchases, and demographic characteristics in São Paulo reveals necessary consumer behavior and market segmentation insights. The Kernel Density Estimation (KDE) graph, as shown in Figure 4, for monthly income distribution, shows a significant concentration of the population earning between R$0 and R$5,000, with an average income of around R$3,500.00 and a sharp decline in the number of individuals earning above R$20,000. This skewed income distribution highlights the predominance of lower income brackets, directly influencing the region's purchasing power and consumer demand. These results are confirmed later when we analyze the social classes of the chosen city.

Historically, Brazil has had a scenario of high-income inequality, as demonstrated by data from the IBGE (2018). Although there have been advances in public policies to reduce these disparities, such as income transfer programs, inequality rates remain high. The Gini coefficient, which measures income concentration, places Brazil among the most unequal countries in the world. The COVID-19 pandemic has exacerbated these differences, disproportionately affecting the poorest populations (IBGE, 2023). This context of structural inequality is directly reflected in the consumption of goods and services, such as electrical materials, since a large part of the population does not have the purchasing power to carry out major renovations or purchase higher value-added products. Thus, consumption predominantly focuses on primary and affordable items that meet immediate home repair and maintenance needs.

Figure 4 - Monthly Income Distribution in the State of São Paulo



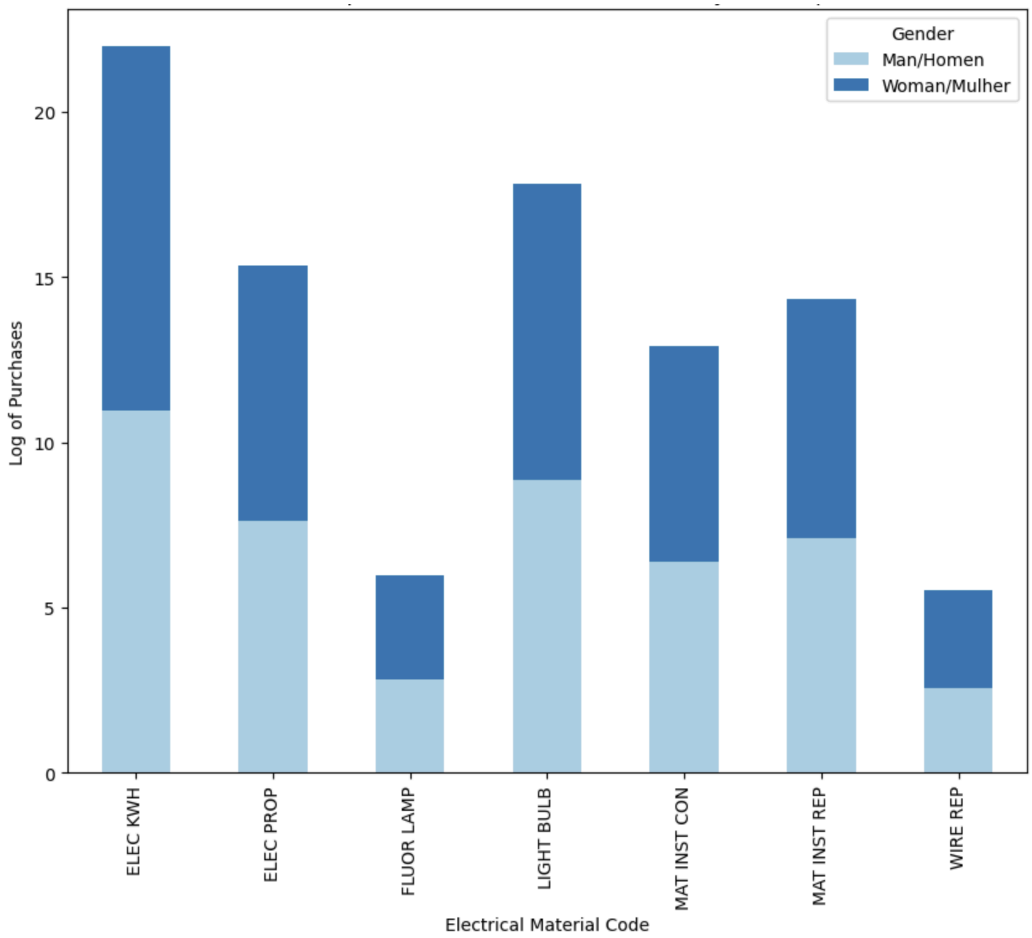
Source: Authors (2024)

A more in-depth analysis of electrical material purchases, using a stacked bar chart, as shown in Figure 4, highlights distinct purchasing patterns based on gender. Figure 4 analyzes the top 10 electrical material purchases in São Paulo state, segmented by gender, based on the electrical material code. The variable V0404, which distinguishes the respondents' gender gender of the respondents (1 for men and 2 for women), reveals differences in purchasing patterns between genders. The category "ELEC KWH" (electricity in kWh) stands out as the most consumed, predominantly by men, while other categories, such as "FLUOR LAMP" (fluorescent lamp) and "LIGHT BULB" (lamp), show a more balanced participation between genders. Purchases related to "MAT INST REP" (installation material - minor repairs) also indicate balanced participation, suggesting that both genders may purchase these items both genders may purchase these items in similar volumes. This graph provides a clear view of the preferences and purchasing behavior of electrical materials among men and women in the state.

We can delve deeper into some interpretations based on the observed patterns. The graph in Figure 4 also reveals a significant difference between genders in terms of consumption of electrical materials, with men showing greater participation in the categories "ELEC PROP" (electrical properties) and "MAT INST CON" (material for electrical installation - construction), which may suggest that they are more involved in electrical maintenance and construction tasks. On the other hand, categories involving smaller or everyday items, such as "FLUOR LAMP" and "LIGHT BULB," show more balanced participation, possibly reflecting purchases associated with home maintenance, where both genders are more involved.

The data in Figure 5 also highlights the concentration of purchases in "MAT INST REP" (minor repairs), which suggests a significant demand among both men and women for more superficial, more straightforward home repairs, which may indicate a potential market for DIY (Do It Yourself) products. The consumption behavior by gender, highlighted here, can provide essential data for targeting more effective marketing campaigns and adjusting communication and promotion strategies according to the consumer profile of each electrical material category.

Figure 5 - Distribution of electrical materials by gender in the State of São Paulo



Source: Authors (2024)

Heatmap analysis across age categories reveals that the highest purchasing intensity occurs among individuals aged 30-59, particularly for items such as “ELEC KWH” (electricity in kWh) and “MAT INST REP” (materials for electrical installation - minor repairs), as shown in Figure 6. In contrast, younger age groups, such as those aged 18-29, show lower engagement in these purchases, especially for products such as “WIRE REP” (wires for electrical repairs) and “MAT INST CON” (materials for electrical installation - construction and renovation). It suggests that middle-aged consumers, who generally have greater purchasing power and are more likely to own real estate, represent a key demographic for marketing strategies in the electrical materials sector in São Paulo state.

The graph used a logarithmic scale to adjust for discrepancies observed in the consumption data, especially in the case of “ELEC KWH.” Since electricity consumption showed highly disparate values compared to other electrical materials, the log scale allows for more precise and comparative visualization of the different purchase categories. This way, the distribution of purchases among the different age groups is better understood, and patterns that significant variations in the data could obscure become more evident. In addition, items such as “LIGHT BULB” and “ELEC PROP” also stand out with a concentration of purchases among individuals aged 40 to 59 and over 60. These results indicate that marketing strategies should focus primarily on middle-aged and older adults, the state's leading buyers of electrical products. Therefore, understanding these sociodemographic demographic dynamics is essential to developing effective market segmentation and consumer targeting approaches in the electrical materials sector.

Figure 6 - Distribution of Purchases by Age in the State of São Paulo

Source: Authors (2024)

These findings suggest that middle-aged consumers, who have greater purchasing power, represent a crucial demographic for targeted marketing strategies in the electrical materials sector in the state of São Paulo. Understanding these dynamics is essential for developing effective market segmentation and consumer targeting approaches, considering the context of inequality and income distribution in Brazil.

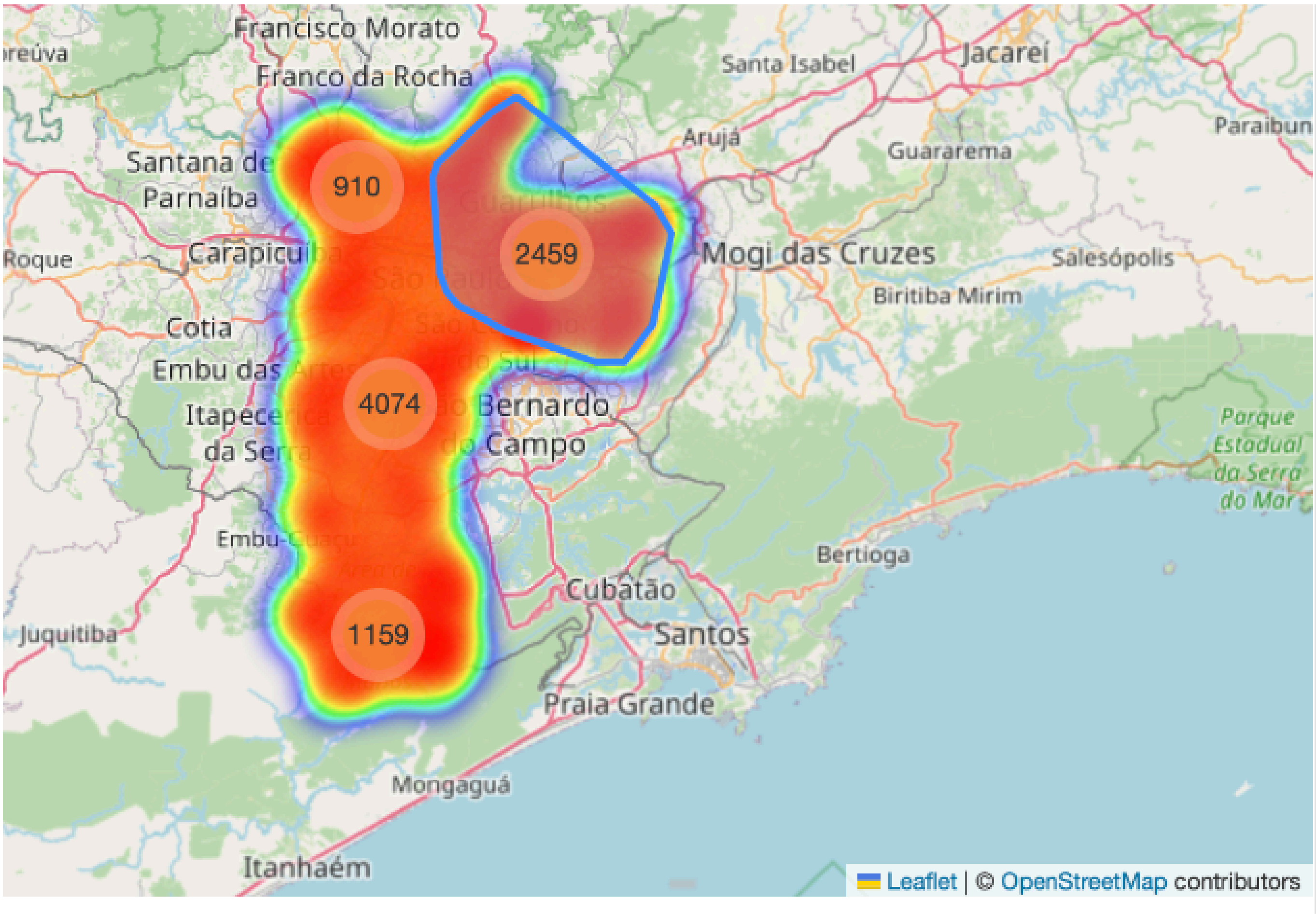
4.2. Municipal Analysis and Small Area Cutout

Still, in exploratory analysis, the data revealed that the city of São Paulo stood out as the main center of consumption of electrical materials in the state, being the largest in terms of population and economic relevance in the state of São Paulo. The consumption behavior in this city offers valuable data on how demographic distribution and purchasing power are reflected in the demand for these products. In addition to refining our focus on microregions, in this section, we present the specific results for the city of São Paulo, highlighting how population concentration and socioeconomic inequalities affect the consumption pattern. This analysis, focused on the largest city in the state, provides a detailed view of the market opportunities and challenges faced by companies in the sector. Thus, it offers valuable support for developing more targeted and effective marketing strategies.

The gender distribution in the city of São Paulo is close to equality. Men comprise 46.56% of the population, while women comprise 53.44%. Until now, we have filtered the database by adding new data to complement and contribute to more precise analyses. When we reached this cutoff point (city), we identified a barrier: the lack of geographic coordinates for POF respondents. Before adding the available data and a new database with the coordinates, the available data were all compiled in the "sao_paulo_df" dataset. To deal with the difficulty of accurately identifying the POF respondents' location, we inferred the location of POF respondents; we chose to infer it from IBGE census data. This new database, inserted from topic 4.2.2 of the code, has latitudinal and longitudinal data for census respondents according to the address type of address (home, establishment, or building). Since the POF is primarily intended for households, we selected coordinates for the city of São Paulo only at the residential level. To complement the 'sao_paulo_df' dataset, we randomly distributed the geographic coordinates among the respondents. This approach was chosen as the most appropriate to ensure that the geographic coordinates reliably represented the population analyzed without compromising the integrity of the data. This method allowed us to associate the respondents' data with their respective regions, preserving the geographic representativeness of the study.

To make the dataset even more robust and provide auxiliary information that increases the accuracy of the SAE model, we included additional data from GeoSampa. Since one of the premises of SAE is using reliable auxiliary variables to improve estimates in areas with lower data density, including mobility data helps better capture the patterns of movement and population distribution, enriching the analysis with a spatial and demographic dimension. With this, the objective was to create an interactive heat map, showing shoppers' locations, the locations of shoppers (demand), and points of greater and lesser mobility in the city, as shown in Figure 7. We implemented a dynamic mapping approach toTo visualize shopping behavior and mobility patterns in São Paulo; we implemented a dynamic mapping approach.

Figure 7 - Interactive Heat Map (Demand and Mobility) of São Paulo City



Source: Authors (2024)

The process began by creating a base map centered on São Paulo (latitude: -23.5505, longitude: -46.6333), providing a detailed city view. This map served as the foundation for subsequent data layers. First, we added a heatmap layer to represent purchase data. Each point on the map corresponded to a specific location where each buyer made purchases. The intensity of the heatmap was determined by the number of purchases (count_elec), with areas of higher concentration displayed in warmer colors (e.g., red) and areas of lower concentration in more excellent colors (e.g., blue or green). This visualization highlighted regions with significant purchase activity. Individual purchase locations were then delineated using circular markers. These markers varied in size, with the radius proportional to the number of purchases. The color scheme differentiated between male and female shoppers, with red borders and orange fills representing male purchases and blue borders with purple fills representing female purchases. Additionally, each marker features a pop-up that displays the exact number of purchases when clicked, providing an interactive element to the map.

To enhance the analysis, we overlaid a second heat map layer with GeoSampa mobility data, which could indicate pedestrian traffic, vehicle movement, or other forms of mobility in the city. This heat map used a gradient from blue (indicating less mobility) to red (indicating more mobility), allowing a visual comparison of movement patterns with purchase data. Due to the large number of mapped points, which exceeds 8,000 records, we focused on analyzing the ten10 regions with the highest and lowest amounts of electrical equipment purchases. This approach allows us to concentrate the visualization and interpretation on the extremes of the distribution, facilitating the identification of relevant patterns without overloading the analysis with an excessive volume of data. Selecting the top 10 and bottom 10 provides a clearer view of the areas with the highest and lowest demand for electrical equipment, allowing us to observe significant contrasts between regions. This data reveals the discrepancies in consumption between areas with a high concentration of purchases and those with substantially lower demand where demand is substantially lower.

Thus, we have data on purchasing behavior patterns in the city of São Paulo, distinguishing between shoppers' genders and the genders of shoppers. The combination of purchasing concentration and mobility data allows for identifying areas with high purchasing activity and significant movement, providing information for strategic business decisions, such as targeted marketing or optimal store placements. The map serves as a tool to understand the interaction between mobility and consumer behavior in urban environments.

As a subsequent step, we highlighted, for better visualization, on different maps, the classification of neighborhoods with the highest and lowest density of electrical material purchases in the city of São Paulo. These points have latitudinal and longitudinal coordinates, allowing an exact analysis of the consumer's location. To do this, we cropped the dataset so that, in this analysis, it would be objective and present only the following relevant data: neighborhood, latitude and longitude, purchase quantity, and gender. It is worth remembering that the heat map in Figure 5 also presents this data interactively. The geographic coordinates of the 20 points, 10 with the highest and 10 with the lowest demand, are available in Table 6.

Table 6 - Location of the 20 Points in the City of São Paulo (High and Low Demand)

Demand	Region	Purchase Quantity (units sold)
High - Top10	Marsilac	5,424,003
High - Top10	São Lucas	5,005,429
High - Top10	Jabaquara	4,859,514
High - Top10	Anhanguera	4,484,502
High - Top10	Cangaíba	2,454,300
High - Top10	Jaguaré	2,454,300
High - Top10	Jardim São Luís	2,405,214
High - Top10	Vila Prudente	2,264,900
High - Top10	Iguatemi	2,264,900
High - Top10	Cangaíba	2,219,602
Low - Bottom10	Marsilac	122,715
Low - Bottom10	Jardim Angela	113,245
Low - Bottom10	Campo Grande	98,172
Low - Bottom10	Marsilac	90,596
Low - Bottom10	Jaraguá	73,629
Low - Bottom10	Anhanguera	67,947
Low - Bottom10	Marsilac	49,086
Low - Bottom10	Ermelino Matarazzo	45,298
Low - Bottom10	Parelheiros	24,543
Low - Bottom10	Jardim Helena	22,649

Source: Authors (2024)

4.3. Market Potential Index for Electrical Products

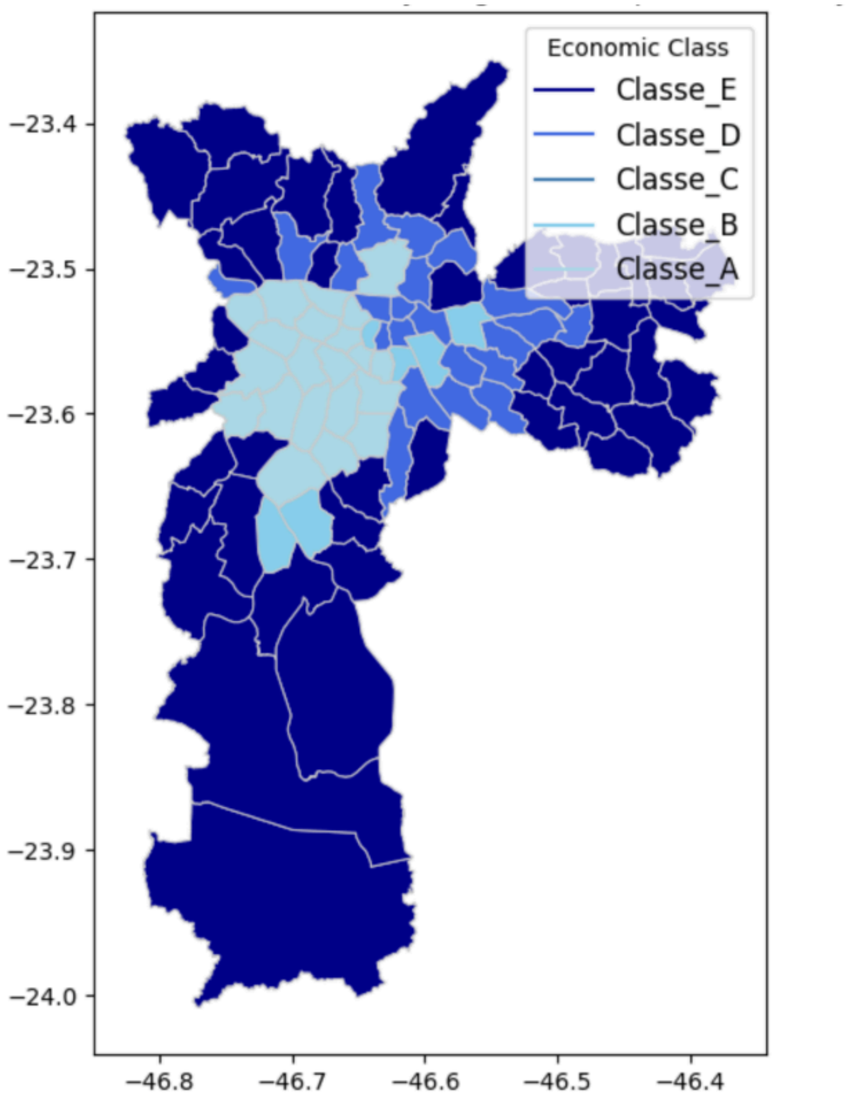
With the ultimate goal of presenting an index of purchasing potential for electrical materials by social class (as shown in Figure 2), we included another database in the dataset with statistical power to classify demand by social class. The IBGE defines population classes according to monthly income. They are, in ascending order: class E (income below R$2,000.00); class D (between R$2,000.00 and R$4,000.00); class C (between R$4,000.00 and R$10,000.00); class B (between R$10,000.00 and R$20,000.00); and class A (above R$20,000.00).

Once again, as one of the premises of SAE, we added a new database to our primary dataset, "sao_paulo_df," with data on the GDP by neighborhoods of the city of São Paulo. The objective was to classify the locations defined in Figure 7 according to the social class of the neighborhood where the respondent was. We adopted this approach as a broad sampling method to define the geographic area in which the consumer is located. However, it is worth noting that the code used provides the exact geographic coordinates of each buyer, still according to Figure 7, allowing for a granular and precise location, which is essential for the SAE methodology. With this, it is possible to identify the specific point of each purchase, ensuring a detailed analysis of the variations in demand between the different microregions of the city. The methodology used to define the class was to list the predominant one in each of the neighborhoods of São Paulo city. It is essential to highlight that, for the most part, all classes are present in all city neighborhoods. However, there is one that prevails. It was the variable we chose to determine the index by social class.

By combining geographic data with economic class distribution data, we determined the predominant economic class in each neighborhood, ranging from class E to class A. Neighborhoods were categorized according to the dominant economic class, using a color scale to facilitate visualization of socioeconomic differences. Each class is represented by a variation in hue, where darker colors indicate lower-income classes and lighter colors indicate higher-income classes: from dark blue for class E to light blue for class A. This coding facilitates visual identification of neighborhood income distribution patterns across neighborhoods, allowing for more intuitive and immediate analysis. Figure 8 effectively illustrates the socioeconomic landscape of São Paulo city, highlighting economic disparity and concentration areas. The visualization provides insights into the geographic distribution of wealth in the city, which can inform urban planning, social policy, and targeted economic interventions.

The map in Figure 8 shows that the darker shades of blue, representing classes E and D (the lowest economic classes), are predominantly concentrated in the outskirts of the city of São Paulo. It suggests that the most economically disadvantaged populations are on the city's outskirts. On the other hand, the lighter shades, representing classes A and B (economically, higher), are found closer to the city center and in specific neighborhoods, indicating areas of wealth concentration. This spatial distribution indicates socioeconomic segregation in São Paulo, where the wealthiest individuals are clustered in more central and well-served areas. At the same time, the poorest populations reside in more peripheral and less accessible regions. The gradient from light blue (class A) to dark blue (class E) reflects the intensity of economic disparity, providing a clear visual representation of the divide between the rich and the less privileged. This visualization not only describes the distribution of wealth in São Paulo but also serves as a powerful tool for identifying areas needing economic and social attention. DemarcatingClearly demarcating neighborhoods by economic class can guide more informed decision-making, aiming to reduce economic disparities and promote inclusive development in the city.

Figure 8 - Map with Dominant Social Classes by Neighborhood



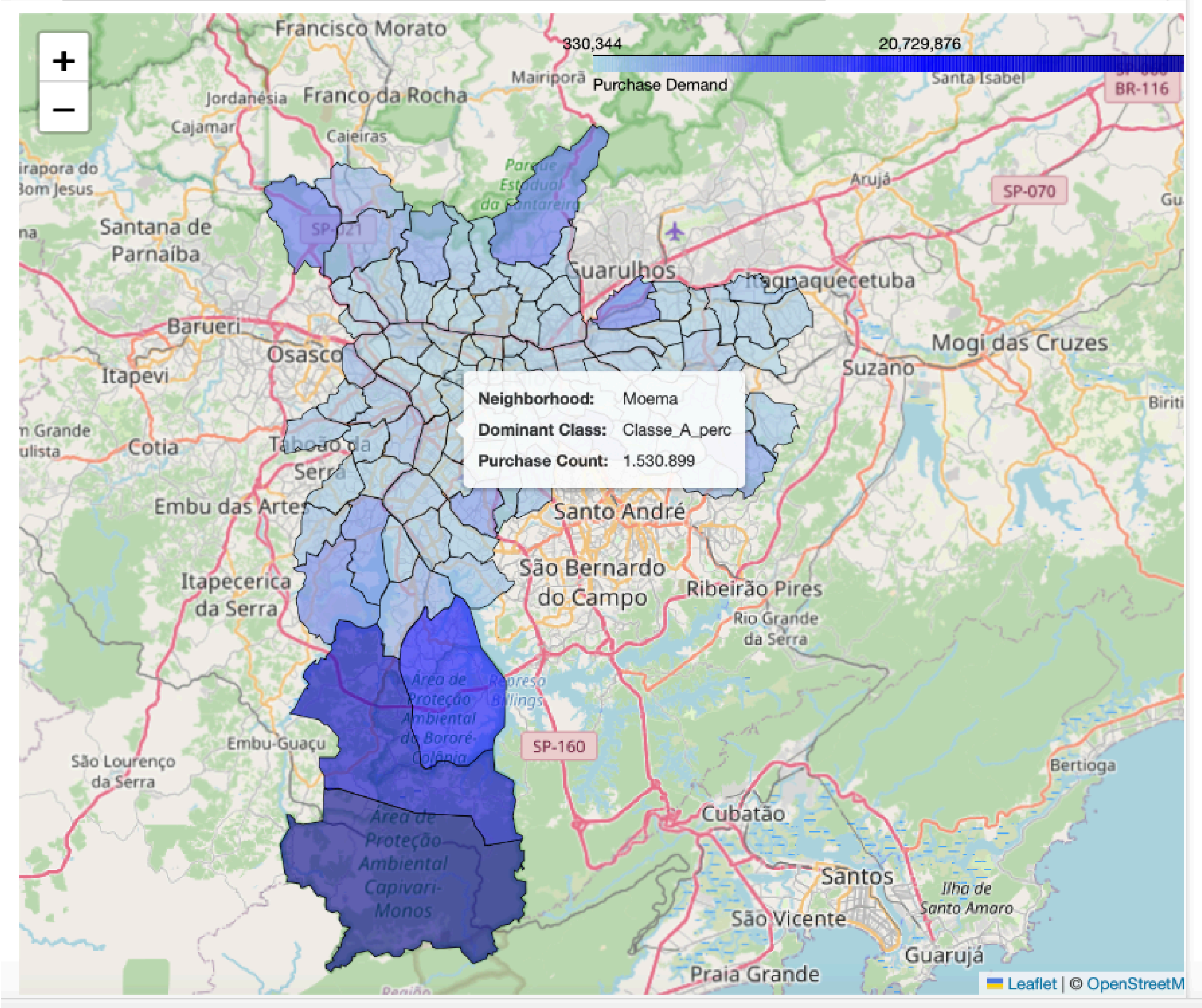
Source: Authors (2024)

The analysis revealed a counterintuitive result: low-income areas such as Marsilac, where the E class is predominant, presented a significant volume of purchases of electrical materials, as demonstrated in the interactive map in Figure 9. Although we expected regions with higher purchasing power would dominate the demand for electrical products, this result suggests that additional factors influence consumption behavior in these areas. First, one can consider the population density of these regions, which, despite the low per capita income, can generate high aggregate demand. Another relevant factor may be the urgent need to maintain or replace for maintenance or replacement of precarious electrical infrastructure. In low-income neighborhoods, it is common for old or poorly maintained electrical systems to need to be updated, generating a demand for essential electrical products. In addition, the search for improvements in quality of life, even in low-income areas, can drive the consumption of items essential for the safety and functionality of homes. These factors indicate that, in certain situations, the demand for durable and essential goods is linked to purchasing power, structural conditions, and basic needs that encourage consumption. This finding highlights the importance of considering variables beyond income when analyzing consumption patterns, especially in economically disadvantaged areas.

On the other hand, wealthier neighborhoods such as Moema and Perdizes, dominated by the upper class, have lower purchasing demand, possibly due to existing ownership of such assets. In addition, wealthier residents may have greater access to maintenance services, extending the lifespan of their electrical products and further reducing the need for replacement. This differentiated relationship between economic class and purchasing behavior challenges the expectation that higher income directly correlates with higher purchase volumes. The findings highlight the importance of considering multiple factors when analyzing consumer behavior.

The interactive map in Figure 9 also reveals a gradient in purchasing behavior, with middle-class neighborhoods exhibiting moderate demand, reflecting their intermediate economic status. These findings illustrate a complex interplay between income levels, existing asset ownership, and purchasing patterns, challenging the simplistic notion that higher income inevitably leads to higher consumption. The data derived from this analysis are essential for businesses and policymakers, as they highlight the need for targeted marketing strategies and policies that consider the specific needs and behaviors of different economic classes in São Paulo city. Understanding these dynamics can help design more effective interventions to improve residents' quality of life aimed at improving the quality of life of residents across socioeconomic strata.

Figure 9 - Interactive Map with Purchase Density and Social Class of the City of São Paulo



Source: Authors (2024)

According to the results presented, the index of purchasing potential for electrical materials, segmented by social class, offers a detailed view of demand in the city of São Paulo. By integrating mobility, income, and geographic location data, it was possible to identify consumption patterns beyond purchasing power, revealing the influence of structural and demographic factors. The analysis revealed that although higher-income neighborhoods, such as Moema and Perdizes, have lower demand for electrical materials due to existing ownership and maintenance of the products, low-income areas, such as Marsilac, show significant demand. This behavior reflects basic needs for infrastructure upgrades, showing that consumption in economically disadvantaged regions can be driven by factors that transcend simple purchasing power.

5. DISCUSSION AND CONCLUSION

Through the application of SAE in this study, it was possible to reveal insights into the demand for electrical products in the city of São Paulo. The analysis, together with statistical methods such as mixed models and spatial effects (Pfeffermann, 2013; Rao & Molina, 2015; Ghosh, 2020), demonstrated the ability of SAE to make reliable inferences for specific geographic areas or subpopulations, even in the absence of a single and complete database. The combination of multiple data sources and direct and indirect estimation techniques allowed the precise identification of critical variables (Ghosh, 2020) influencing purchasing behavior across different economic classes.

The results of this study have strategic implications for companies that wish to optimize their marketing and distribution operations. In high-demand regions, where lower economic classes predominate, the segmentation and positioning strategy should focus on value solutions that highlight the cost efficiency and benefits of upgrades in electrical materials. In regions of lower demand, dominated by higher economic classes, marketing communications should highlight the reliability and quality of products, reinforcing the company's reputation and the satisfaction of existing customers.

Applying SAE methodology to the electrical materials market in São Paulo city revealed significantly more complex consumption dynamics than initially expected. The analysis demonstrated that purchasing behavior across social classes is influenced by purchasing power and strongly conditioned by structural needs and specific local contexts, such as the precariousness of electrical infrastructure in low-income regions. This complexity highlights the importance of considering a broader range of variables when analyzing the consumption of essential goods in addition to traditional income metrics (Silva, 2014; Wagner et al., 2022). As a result, this research offers a more detailed and multidimensional view (Rao & Molina, 2015) of the socioeconomic demands in São Paulo neighborhoods, providing fundamental support for developing more assertive market strategies and public policies. By aligning these initiatives with the real needs of different social classes, it is possible to promote more effective interventions capable of meeting the specificities of each region and maximizing the economic and social impact (Rao & Molina, 2015; Ghosh, 2020).

The practical application of SAE techniques has proven effective in identifying market opportunities that could go unnoticed in traditional analyses (Ghosh, 2020). By segmenting and targeting marketing campaigns based on granular insights, companies can increase their market penetration, optimize resource allocation, and maximize return on investment (Bazotti et al., 2016). Therefore, the research provided an affirmative answer to the work's central question:

Does SAE microdata allow the identification of market opportunities? The analysis of the results indicated that SAE is a tool for identifying market opportunities at the local level, answering the research question. Based on the analyzed microdata, it was possible not only to identify geographic areas with high demand potential but also to understand the underlying factors that drive this demand, which is crucial for the formulation of more effective marketing and distribution strategies (Carvalho & Pereira, 2012; Rao & Molina, 2015; Ghosh, 2020).

The importance of Python as an analytical tool in this work cannot be overstated. Python's ability to manipulate data, run complex statistical models, and visualize results was necessary for applying SAE techniques. Using libraries such as Pandas, NumPy, Statsmodels, and SciPy, it was possible to perform large-scale data analyses with a high degree of accuracy. Python allowed for the efficient combination of multiple data sets, the execution of mixed and spatial models, and the prompt generation of granular insights. Without this tool, the execution of this study would have been more challenging, limiting the scope and accuracy of the results.

This empirical study validates the integration of SAE into business strategies, especially in a context where identifying localized opportunities is essential for business growth and success. The findings indicate that by leveraging localized insights into the demand for electrical products, companies can identify, evaluate, and capitalize on market opportunities more accurately and in an informed manner (Rao & Molina, 2015; Silva, 2017; Ghosh, 2020). Furthermore, the approach offers a model that can be replicated in other sectors and regions, promoting more strategic and data-driven decision-making. The use of Python was essential in achieving these results, highlighting the importance of advanced analytical tools in researching and applying research and application of complex statistical techniques such as SAE.

6. LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE RESEARCH

While the benefits of leveraging POF data for data-driven marketing strategies are substantial, we highlight four challenges that must be addressed to capitalize on this resource fully. The first challenge is data integration (Carvalho & Pereira, 2012; Silva, 2014; Bazotti et al., 2016) with other sources, such as points of sale and digital surveys, for a more comprehensive market view, which requires advanced analytical systems and tools. The second challenge is privacy and ethics in the use of consumer data (Bazotti et al., 2016), requiring compliance with legislation, such as the General Data Protection Law (LGPD), and responsible use of data. The third involves market dynamics, marked by economic fluctuations (Pintos-Payeras, 2009; Silva, 2014) that require continuous strategic adjustments, while the fourth is the time lag of the data since the POF is only updated every six years (IBGE, 2018).

These challenges highlight the importance of a careful and ethical approach to POF data without disregarding its potential to understand regional consumption patterns and demographic segmentations. Companies that can overcome these barriers are better positioned to develop effective and responsive marketing campaigns for the Brazilian Market. It allows for a deeper connection with target audiences and agile adaptation to local dynamics.

When robust data sources like POF inform data-driven marketing strategies, they offer a powerful approach to understanding and engaging with the Brazilian Market. By leveraging POF data, companies can gain a nuanced understanding of consumer behavior across different regions and demographic segments (Silva, 2014; IBGE, 2018), enabling them to develop targeted and effective marketing campaigns that resonate with their audiences. However, to fully realize the potential of data-driven marketing, companies must address the challenges of data integration, privacy, and market dynamics, ensuring that their strategies are practical and ethically sound.

While this study provided valuable data on using SAE to identify market opportunities for electrical products in São Paulo, it is essential to acknowledge some limitations that may have influenced the results and conclusions. First, the availability and quality of the data used represent a significant limitation. The analysis relied on multiple data sources, each with its limitations regarding accuracy, timeliness, and comprehensiveness. Combining data from different levels and sources (national, state, and municipal) may introduce bias, especially when there are discrepancies or inconsistencies between the data sets. Although advanced statistical techniques, such as mixed models and spatial effects, were used to mitigate these issues, the inherent limitations of the available data may still have impacted the accuracy of the estimates.

Another limitation is related to the geographic specificity of the analysis. This study focused exclusively on São Paulo city, meaning that the results and conclusions may not directly apply to other regions or cities, even in Brazil. Market dynamics, consumer behaviors, and demographic characteristics can vary significantly across different areas, limiting the generalizability of the findings of this study to other geographic contexts. In addition, the use of specific statistical models imposes some limitations. We based our choice of mixed and spatial models on best practices in the literature, but we could have explored other methods or approaches to verify the robustness of the results. Alternative models, such as nonparametric models or machine learning techniques, could provide complementary insights or identify patterns that the models used did not capture.

Another important consideration is the interpretation of the results obtained. Although SAE has proven effective in identifying areas with high or low demand for electrical products, the underlying causes of these patterns have yet to be explored in depth. Contextual factors, such as public policies, local infrastructure, and economic changes, may have significantly influenced and played a significant role in demand variations but were not directly incorporated into the analysis model. There is an apparent temporal limitation in this study. The data reflect a specific period of the POF (2017-2018), and although techniques to deal with lag were applied, the conclusions may only remain valid in the future with frequent updates. Market and consumption dynamics evolve rapidly, and what applies now may not be relevant in a future context.

Finally, there is a temporal limitation to this study. The data used reflect a specific point in time. Although the analysis included techniques to deal with changes over the years, the conclusions may only apply to future periods without constant updates and reassessments. Market dynamics constantly evolve, and what is valid today may not be valid tomorrow. In summary, although this study offers significant essential contributions to the application of SAE in business strategies, the limitations identified should be considered when interpreting the results and when attempting to apply these insights to other contexts or sectors. Future research could explore additional approaches to mitigate these limitations and expand the applicability of the findings. Based on the limitations discussed, future research could consider the following directions:

1. **Data update and longitudinal analysis:** To maintain the relevance of the results, future research could focus on incorporating more recent POF data when available. It would allow for longitudinal analysis that captures changes in consumer behavior over time, providing insights into emerging trends in the electrical materials sector.
2. **Geographic expansion:** Future research could apply SAE to other regions of Brazil, comparing consumption patterns between different states or cities. It would help assess the validity of the conclusions for other areas and allow the identification of market opportunities in different geographic contexts.
3. **Integration of alternative models:** Besides mixed and spatial models, methods such as machine learning or neural networks could complement the analysis, offering a more robust predictive approach and helping to detect complex patterns not captured by traditional methods.

4. **Exploring contextual factors:** Public factors such as public policies, urban infrastructure, and regulatory changes can significantly influence the demand for electrical materials. Future research could include these variables in the analysis model to provide a more comprehensive view of the determinants of demand.

This study has laid the foundation for the application of SAE in the electrical materials sector, highlighting significant market opportunities in São Paulo city. However, incorporating more recent data and exploring other methodologies could further refine the results, increasing the accuracy of the analyses. By advancing in these areas, it will be possible to identify more dynamic consumption patterns and develop more effective marketing strategies tailored to local realities.

IV. CONCLUSION

This study presented a comprehensive analysis of the potential of Small Area Estimates (SAE) in identifying business opportunities and supporting strategic decisions. Based on a systematic review of the literature presented in the first article and a methodological application presented in the second article, it was possible to highlight the contributions of SAE to the business context, especially in the electrical products sector. The results of both articles show that, by offering a more granular view of the market, SAE enables a more precise understanding of microregion trends, creating more aligned and informed strategies.

Through the first article, we highlighted that, despite the significant potential of SAE, there still needs to be a gap in its application in identifying marketing and business strategies. The literature highlights the ability of SAE to provide granular data on consumer behavior and market dynamics. However, its adoption is still limited due to methodological challenges and the familiarity of potential applicants. Overcoming such barriers requires greater awareness of the benefits of SAE and training and improvement of modeling techniques. This work helps to advance these issues.

In the second article, we demonstrated how this approach can generate valuable data on demand through the practical application of SAE in the city of São Paulo/SP. The analysis showed that different contextual and structural factors influence purchasing behavior in different socioeconomic classes, highlighting the importance of adapting marketing strategies to the peculiarities of microregions. Using analytical tools, such as Python, was essential to obtain these results, highlighting the importance of advanced techniques amid microdata analysis.

It was possible to verify that the combination of SAE-based methodologies provides a significant competitive advantage, especially in heterogeneous markets. By identifying demand patterns in different microregions, companies can create more efficient marketing strategies, optimize resources, and increase the return on the amount invested. The practical application showed that SAE overcomes the limitations of more traditional analyses and reveals nuances of consumer behavior that might otherwise go unnoticed.

In addition, this study also reinforces the importance of using contemporary analytical tools when implementing more complex techniques, such as SAE. The ability to manipulate large databases and precision in statistical analyses are fundamental for obtaining reliable results. By adopting approaches based on SAE and robust analytical technologies, public and private institutions can transform granular data into strategic information, promoting a continuous cycle of adaptation to market demands.

Finally, the combination of theoretical and empirical analyses validates the role of SAE as a strategic and assertive tool for those seeking to identify and invest in market opportunities at a local level. By incorporating SAE into their practices, institutions can improve economic performance and contribute to the more sustainable development of the communities in which they operate. The ability to adapt strategies to the peculiarities of each microregion, with the use of robust tools, positions SAE as a valuable resource to face competition and market complexity.

V. REFERENCES

Abell, D. F. 1980. Defining the Business: Starting Point of Strategic Planning. Prentice Hall, Englewood Cliffs, NJ

Abotaleb, I. S., & El-adaway, I. H. (2017). Construction bidding markup estimation using a multistage decision theory approach. Journal of construction engineering and management, 143(1), 04016079.

Associação Brasileira da Indústria Elétrica e Eletroeletrônica. Panorama Econômico, 2024. Available at <https://www.abinee.org.br/organizacao/decon/panorama/>.

Baran, R. (2012). A introdução de veículos elétricos no Brasil: avaliação do impacto no consumo de gasolina e eletricidade.

Barua, S., Abedin, Z., Nath, A., & Biswas, C. (2019, May). A Statistical Estimation of Solar Power for Energy Mix in Bangladesh. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-5). IEEE.

Bazotti, A., Finokiet, M., Conti, I. L., França, M. T. A., & Waquil, P. D. (2016). Tabagismo e pobreza no Brasil: uma análise do perfil da população tabagista a partir da POF 2008-2009. Ciência & Saúde Coletiva, 21, 45-52.

Bickel, D. R. (2012). Controlling the degree of caution in statistical inference with the Bayesian and frequentist approaches as opposite extremes.

Bickel, D. R. (2014). Small‐scale Inference: Empirical Bayes and Confidence Methods for as Few as a Single Comparison. International Statistical Review, 82(3), 457-476.

Booth, A., Sutton, A., & Papaioannou, D. (2016). Systematic Approaches to a Successful Literature Review. Sage Publications.
<https://books.google.com.br/books?id=A1lvCAAAQBAJ&lpg=PP1&hl=pt-BR&pg=PP1#v=onepage&q&f=false>

Buckwell A, Heissenhuber A, Blum W (2014) Sustainable Intensification of European Agriculture. RISE Foundation, Brussels: https://risefoundation.eu/wp-content/uploads/2020/07/2014_-SI_RISE_FULL_EN.pdf

Buganova, K., Luskova, M., Kubas, J., Brutovsky, M., & Slepecky, J. (2021). Sustainability of business through project risk identification with use of expert estimates. Sustainability, 13(11), 6311.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296.

Carvalho, C. H. R. D., & Pereira, R. H. M. (2012). Gastos das famílias brasileiras com transporte urbano público e privado no Brasil: uma análise da POF 2003 e 2009.

Chandra, H., Chambers, R., & Salvati, N. (2012). Small area estimation of proportions in business surveys. Journal of Statistical Computation and Simulation, 82(6), 783-795

Cliquet, G. (Ed.). (2013). Geomarketing: Methods and strategies in spatial marketing. John Wiley & Sons.

Cliquet, G. (2021). From geomarketing to spatial marketing. Spatial Economics Volume II: Applications, 277-305.

Costa, R., Conceição, M. M., Da Silva, A. R., Conceição, J. T. P. (2021). Marketing verde – A importância do consumo sustentável para as empresas. Research, Society and Development, 10 (7).

Chopra, S., Dougan, D., & Taylor, G. (2001). B2b. Supply Chain Management Review, 51.

de Almeida, A. N. (2011). Elasticidades renda e preços: análise do consumo familiar a partir dos dados da POF 2008/2009. Núcleo de Economia Regional e Urbana da Universidade de São Paulo (NEREUS).

Eid, R., Trueman, M., & Moneim Ahmed, A. (2002). A cross‐industry review of B2B critical success factors. Internet Research, 12(2), 110–123.doi:10.1108/10662240210422495

Eisenhardt, K. M., & Zbaracki, M. J. (1992). Strategic decision making. Strategic management journal, 13(S2), 17-37.

Erciulescu, A. L., Franco, C., & Lahiri, P. (2021). Use of administrative records in small area estimation. Administrative Records for Survey Methodology, 231-267.

Fabrizi, E., Ferrante, M. R., & Trivisano, C. (2018). Bayesian small area estimation for skewed business survey variables. Journal of the Royal Statistical Society Series C: Applied Statistics, 67(4), 861-879.

Fast, N. J., & Schroeder, J. (2020). Power and decision making: new directions for research in the age of artificial intelligence. Current opinion in psychology, 33, 172-176.

Feizizadeh, B., Omarzadeh, D., Sharifi, A., Rahmani, A., Lakes, T., & Blaschke, T. (2022). A GIS-based spatiotemporal modelling of urban traffic accidents in Tabriz City during the COVID-19 pandemic. Sustainability, 14(12), 7468.

Ferreira, L. B. L. (2011). A importância da capacitação profissional para empresas do agronegócio.

Fink, A. (2019). Conducting research literature reviews: From the internet to paper. Sage publications. <https://books.google.com.br/books?hl=pt-BR&lr=&id=0z1_DwAAQBAJ&oi=fnd&pg=PP1&dq=conducting+research+literature+reviews&ots=16GuaZSTbC&sig=YTyuFtBLuC8hTz6bJKMYeZ4lYMk#v=onepage&q=conducting%20research%20literature%20reviews&f=false>

Fundação Sistema Estadual de Análise de Dados (2023). Fundação SAEDE. Governo Estadual de São Paulo. Available at <https://www.seade.gov.br/>.

Galvão, T. F., Pansani, T. D. S. A., & Harrad, D. (2015). Principais itens para relatar Revisões sistemáticas e Meta-análises: A recomendação PRISMA. Epidemiologia e serviços de saúde, 24, 335-342.

Geng, Z., Wang, Z., Peng, C., & Han, Y. (2016). A new fuzzy process capability estimation method based on kernel function and FAHP. IEEE Transactions on Engineering Management, 63(2), 177-188.

Ghosh, M., & Rao, J. N. (1994). Small area estimation: an appraisal. Statistical science, 9(1), 55-76.

Ghosh, M. (2020) : Small area estimation: its evolution in five decades. Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, Vol. 21, Iss. 4, pp. 1-22,

Gonzalez, M.E. and Waksberg, J. (1973), Estimation of the Error of Synthetic Estimates, in Paper presented at the First Meeting of the International Association of Survey Statisticians,Vienna, Austria.

Goodman, S. K. (1993). Information needs for management decision-making. Information Management, 27(4), 12.
Harrison, E. F., & Pelletier, M. A. (2000). The essence of management decisions. Management decision, 38(7), 462-470.

Gruber, M., MacMillan, I. C., & Thompson, J. D. (2008). Look before you leap: Market opportunity identification in emerging technology firms. Management science, 54(9), 1652-1665.

Hallikainen, H., Savimäki, E., & Laukkanen, T. (2020). Fostering B2B sales with customer big data analytics. Industrial Marketing Management, 86, 90-98.

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983), An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys, Journal of the American Statistical Association, 78, 776–793.

Instituto Brasileiro de Geografia e Estatística - IBGE. Pesquisa de Orçamentos Familiares 2017-2018. Rio de Janeiro: IBGE, 2018.

Instituto Brasileiro de Geografia e Estatística - IBGE. Pesquisa Nacional por Amostra de Domicílios Contínua de 2023. Rio de Janeiro: IBGE, 2024.

Instituto Brasileiro de Geografia e Estatística - IBGE. Censo Demográfico de 2022. Rio de Janeiro: IBGE, 2024.

Jiang, Y., Kong, P., Xu, H., Zhang, Q., Quan, X., & He, B. (2021, July). Temporal Mapping of Grassland Aboveground Biomass in Qinghai Province from Landsat 8 and Sentinel-2. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 6901-6904). IEEE.
Kotler, P., & Pfoertsch, W. (2008). Gestão de marcas em mercados B2B. Bookman Editora.

Leek, S., & Christodoulides, G. (2011). A literature review and future agenda for B2B branding: Challenges of branding in a B2B context. Industrial marketing management, 40(6), 830-837.

Li, X., Luo, G., Wang, W., Wang, K., Gao, Y., & Li, S. (2021). Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation. IEEE Journal of Biomedical and Health Informatics, 26(3), 1140-1151.

Liu, Q., Chen, B., Wang, F., Luan, G., & Hu, D. (2021, September). Vehicle Distance Estimation Based on Monocular Vision and CNN. In 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI) (pp. 638-641). IEEE.

Matthews A (2020) The new CAP must be linked more closely to the UN Sustainable Development Goals. Agric Econ 8:19.https://doi.org/10.1186/s40100-020-00163-3

Marchau, V. A., Walker, W. E., Bloemen, P. J., & Popper, S. W. (2019). Decision making under deep uncertainty: from theory to practice (p. 405). Springer Nature.

MARTÍNEZ-LÓPEZ, F. J. et al. Fifty years of the European Journal of Marketing: a bibliometric analysis. European Journal of Marketing, [s. l.], v. 52, n. 1/2, p. 439–468, 2018.

Melnyk, L., & Nyzhnyk, L. (2018). Geomarketing is an innovative technology business. Industry 4.0, 3(3), 141-143.

Miller, D. J., M. J. Fern, L. B. Cardinal. 2007. The use of knowledge for technological innovation within the diversified firm. Acad. Management J. 50 308–326.

Ministério de Minas e Energia do Brasil - MME. Portal de Dados Aberto, 2023. Available at <https://dadosabertos.mme.gov.br/>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. International journal of surgery, 88, 105906.

Paul, J., & Criado, A.R. (2020). The art of writing literature review: What do we know and what do we need to know? International Business Review, 29, 101717.

Petrucci, A., & Salvati, N. (2004). Small area estimation using spatial information. The rathbun lake watershed case study. Dipartimento di Statistica "G. Parenti" viale morgagani, 59-50134.
Pfeffermann, D. (2002). Small area estimation‐new developments and directions. International Statistical Review, 70(1), 125-143.

Pfeffermann, D. (2013). New important developments in small area estimation.

Pignone, M., Saha, S., Hoerger, T., Lohr, K. N., Teutsch, S., & Mandelblatt, J. (2005). Challenges in systematic reviews of economic analyses. Annals of internal medicine, 142(12_Part_2), 1073-1079.

Pintos-Payeras, J. A. (2009). Estimação do sistema quase ideal de demanda para uma cesta ampliada de produtos empregando dados da POF de 2002-2003. Economia Aplicada, 13, 231-255.

Prefeitura de São Paulo (GeoSampa). Mapa Digital da Cidade de São Paulo, 2024. Available at <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx>.

Rao, J. N., & Molina, I. (2015). Small area estimation. John Wiley & Sons.

Rodrigues, Â. C. (2007). Impactos socioambientais dos resíduos de equipamentos elétricos e eletrônicos: estudo da cadeia pós-consumo no Brasil (Doctoral dissertation, Faculdade de Engenharia, Arquitetura e Urbanismo da Universidade Metodista de Piracicaba. Programa de Pós-Graduação em Engenharia de Produção.).

Rotejanaprasert, C., Ekapirat, N., Areechokchai, D., & Maude, R. J. (2020). Bayesian spatiotemporal modeling with sliding windows to correct reporting delays for real-time dengue surveillance in Thailand. International journal of health geographics, 19, 1-13.

SAIPE, 2023. Available at <https://www.census.gov/programs-surveys/saipe.html>. Accessed on: May 12, 2023.

Silva, C. D. A. (2014). Mercado de comida japonesa no Distrito Federal: análise das oportunidades de negócio por meio de Geomarketing e Máquinas de Suporte Vetorial.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, 333-339.

Vaiman, V., Scullion, H., & Collings, D. (2012). Talent management decision making. Management Decision, 50(5), 925-941.

Vaz, D. V., & Hoffmann, R. (2020). Elasticidade-renda e concentração das despesas com alimentos no Brasil: uma análise dos dados das POF de 2002-2003, 2008-2009 e 2017-2018. Revista de Economia, 41(75).

Wagner, Y. G., Coelho, A. B., & Travassos, G. F. (2022). Análise do consumo domiciliar de pescados no Brasil utilizando dados da POF 2017-2018. Revista de Economia e Sociologia Rural, 61(3), e250494.

You, Y., & Rao, J. N. K. (2002). Small area estimation using unmatched sampling and linking models. Canadian Journal of Statistics, 30(1), 3-15.

Zanutto, E. and Zaslavsky, A. (2002). Using administrative data to improve small area estimation: an example from the U.S. Decennial Census. Journal of Official Statistics 18: 559–576.

APPENDIX A

Table 4 - Possible applications of POF for developing marketing strategies

Marketing Strategies	Description	Benefits	Related Studies	Featured in this study?
Consumer segmentation	Analyze POF data to identify distinct consumer segments based on income levels, consumption patterns, and demographic characteristics	It enables targeted marketing efforts tailored to specific groups, resulting in more effective and relevant campaigns.	**Bazotti et al. (2016)** - Explores the profile of tobacco consumers in Brazil, highlighting factors such as age, gender, and education in different regions. **Carvalho & Pereira (2012)** - Discusses public and private urban transportation spending in Brazil, focusing on transportation consumption patterns by income bracket.	Yes
Product development and pricing	Using POF data to understand how different income groups allocate their budgets to inform product design and pricing strategies	Helps design products and pricing strategies that align with the purchasing power and priorities of target audiences	**Vaz & Hoffmann (2020)** - Analyzes income elasticity and concentration of food expenditure in Brazil, suggesting differences in consumption between income brackets. **Pintos-Payeras (2009)** - Estimates demand elasticities for different product categories, highlighting consumption patterns by price range.	No
Regional Market Insights	Leveraging regional breakdowns of POF data to gain insights into economic conditions and consumer behavior in different parts of Brazil	Supports expansion into new regions or optimization of existing operations by focusing on high potential areas	**Silva (2014)** -Identifies business opportunities in the Japanese food market in the Federal District using Geomarketing to map areas with high demand and low	Yes

			supply. **Wagner et al. (2022)** - Analyzes the factors that influence fish consumption in Brazil, highlighting the importance of location and income.	
Trend Analysis	Using longitudinal POF data to track changes in consumer behavior over time and identify emerging trends	Enables businesses to adapt marketing strategies to evolving consumer demands while maintaining a competitive advantage	**Carvalho & Pereira (2012)** - It tracks the increase in spending on private transport over time, reflecting changes in urban mobility patterns **de Almeida (2011)** - Explores changes in food consumption patterns over the past few decades, including the impact of income on food choices.	Yes

Source: Authors (2024)