UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

FACULTY OF ENGINEERING

GRADUATION PROJECT

SPATIAL DISTRIBUTION AND CHARACTERIZATION OF HONDURAN ELECTRICITY DEMAND

PRIOR TO OBTAINING THE DEGREE OF

ENERGY ENGINEER

PRESENTED BY:

21911416 GABRIELA MUNGUÍA DERAS

ADVISER: PHD. HÉCTOR VILLATORO

CAMPUS SAN PEDRO SULA; SEPTEMBER 2022

DEDICATION

I dedicate this thesis to my parents. To my father for giving me the opportunity to study through his financial support. Dad, this achievement is yours. And, to my mother for always being my number one supporter.

ACKNOWLEDGMENT

I thank EEH, especially Eng. Fernando Sarmiento, for providing information on municipal electricity consumption for this investigation, without it, this research would not have been possible.

I thank P. Eng Héctor Villatoro, for his excellence in research but also his objective advice during this period of thesis writing.

I thank P. Eng Alicia Reyes, who in addition to being my professor, has served as a guide and advisor for me.

I thank my colleagues of Energy Engineering, especially the faithful companions in the CRAI and "Los intensos" (Omar Enamorado, Celeste Parada and Laura Villanueva).

ABSTRACT

As energy demand rises and fossil fuel reserves become scarcer, energy planning is necessary in every country including Honduras. Chiefly, an energy planning that identifies the strength in decentralized energy systems in the national territory. The mapping and characterization of electricity consumption in Honduras, proposed in this investigation, is a significant step toward the goal of creating instruments that increase energy planning. The cartographic tool created provides relevant information of the behavior of the demand and the sectors that consume the most in each one of the 259 municipalities studied (only in 259 municipalities the electricity consumption data was accessible by the author from a total of 298 municipalities in Honduras). Through Tableau, a data visualization software, an interactive tool containing a choropleth map of Honduras based on electricity consumption filtered by department and municipality is achieved, allowing a greater level of disaggregation for this data. Through Excel, a univariable analysis to understand the relationship between 10 sociodemographic variables and electricity consumption is done. Through RStudio, which is an integrated development environment for the R programming language, dedicated to statistical computing and graphics, the sociodemographic variables that are significant for the electricity consumption in each municipality are described. In Honduras only 18 municipalities from the total of 259 municipalities studied consumed 80% of the electrical energy, i.e., in Honduras this resource is distributed unequally in the national territory. These results are valuable information for the governmental and municipal decision-makers that should implement procedures through public policies and regulations to evenly share the use of electricity throughout the nation and encourage the decentralization of electrical generation, installing on grid micro power plants near the municipalities with the highest electrical consumption or strategical located municipalities where this resource will improve the human and economic development for the population.

Keywords: *energy planning, decentralized energy systems, demand forecast model, geospatial-data integration, human development index.*

RESUMEN

A medida que aumenta la demanda de energía y escasean las reservas de combustibles fósiles, la planificación energética es necesaria en todos los países, incluido Honduras. Principalmente, una planificación energética que identifique la fortaleza de los sistemas energéticos descentralizados en el territorio nacional. El mapeo y caracterización del consumo eléctrico en Honduras, propuesto en esta investigación, es un paso significativo hacia la meta de crear instrumentos que incrementen la planificación energética. La herramienta cartográfica creada brinda información relevante del comportamiento de la demanda y los sectores que más consumen en cada uno de los 259 municipios estudiados (solo en 259 municipios los datos de consumo eléctrico fueron accesibles por el autor de un total de 298 municipios en Honduras). A través de Tableau, un software de visualización de datos, se logra una herramienta interactiva que contiene un mapa coroplético de Honduras basado en el consumo eléctrico filtrado por departamento y municipio, lo que permite un mayor nivel de desagregación de estos datos. A través de Excel, se realiza un análisis univariable para entender la relación entre 10 variables sociodemográficas y el consumo eléctrico. A través de RStudio, que es un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos, se describen las variables sociodemográficas que son significativas para el consumo de energía eléctrica en cada municipio. En Honduras solo 18 municipios del total de 259 municipios estudiados consumen el 80% de la energía eléctrica, es decir que en Honduras este recurso se distribuye de manera desigual en el territorio nacional. Estos resultados son información valiosa para los decisores gubernamentales y municipales que deben implementar procedimientos a través de políticas públicas y normativas para repartir equitativamente el uso de la energía eléctrica en todo el territorio nacional e incentivar la descentralización de la generación eléctrica, instalando en la red micro centrales cerca de los municipios con los de mayor consumo eléctrico o municipios ubicados estratégicamente donde este recurso mejorará el desarrollo humano y económico para la población.

Palabras claves: *generación descentralizada, índice de desarrollo humano, integración de datos geoespaciales, modelo de previsión de la demanda, planificación energética.*

TABLE OF CONTENTS

INDEX OF ILLUSTRATIONS

INDEX OF EQUATION

TABLE INDEX

I. Introduction

The integration of planning into the energy system of a country is a matter of vital importance to aspire for a sustainable future. OLADE (2017) distinguishes three energy plannning tools that help in this process: economic and energy information systems, energy prospective studies including conditions in an economic and social context at national and international levels, and models used for planning in the global and sub-sectorial energy levels. Based in these three tools, the mapping of electricity consumption in Honduras could be a significant step toward the goal of creating economic and energy information systems.

However, when studying the electricity consumption geographically in Honduras the disaggregation level reaches only a division of the country in three main regions using the nation's most recent information (2020). This disaggregation level does not allow to explore the electricity consumption by each one of the 18 departments in Honduras or to a more disaggregated state, by each one of the 298 municipalities in which Honduras is divided. The actual disaggregation level also precludes the study of the social and economic variables that have incidence in the electricity consumption behavior because the municipalities in the each one of the three regions do not share the same socioeconomical characteristics. A municipal disaggregation is important because access to electrical consumption information at this level empowers decision-makers such as municipal mayors to track the electrical demand as a step to start the process of energy planning. An energy planning that should focus on decentralized systems based on local energy sources which diversifies the resources and increases the self-sufficiency of a region. (Quintero P. , 2008).

Involving the mapping of electricity demand, many authors have studied the importance of mapping the demand, in United States of America a tool was develop by Saxum (2012) to understand the consumption in 13 regions of the country and in Latin America, Reza Torres (2015) mapped the consumption of electricity in the state of Jalisco in México and studied through a linear multiple regression a model of electricity consumption based on 31 independent sociodemographic variables through Minitab. In Honduras, where this study takes place, (W. C. Flores et al, 2021) created a tool that tracks the electricity consumption of the country year by year at national level.

Many authors have studied the relation between human development and electricity consumption; Reza Torres (2015) did it throught the regression model mentioned in the last paragraph, Mazur (2011) studied life expectancy with respect to energy and electricity consumption concluding that these two variables are essential for improving the wellbeing of people in less developed nations. Others have directly studied the Human Delopment Index (HDI) wich represents the life expectancy index, GDP index and education index of a nation. Arto et al (2016) studied the relationship between total primary energy demand and HDI whereas Leung & Meisen (2005) studied the relationship between electricity demand and HDI in medium and low HDI countries, concluding that a logarithmic relationship exist in both cases. Mohamed, Z., & Bodger, P (2005) related electricity consumption to sociodemographic variables like Gross Domestic Product, Population and Electricity Prices concluding that this three variables were significant for the model. In the United States of America, Haerer et al (2015) studied the relationship between employment and the energy sector, concluding that significant changes in employment occurred in industry sectors that support the operation and management activities of U.S. electric power industry.

Like the previously mentioned studies, this thesis also proposes the mapping of the electricity consumption in the national territory and evaluates the relationship between sociodemographic and economic variables with respect to electricity consumption, with the novelty that this study will be based in Honduras and will be considering the consumption of each one of the municipalities (259 municipalities where data was available) in the territory.

In this thesis, the data will be transformed using Tableau into a cartographic interactive tool that the public can access online. An univariable analysis will be carried out using Excel to examine the relations between each of the 10 sociodemographic variables per municipality into the model. Finally, using R, a multiple regression model is created to identify the variables that are significant to forecast the demand of electricity.

This document is structured as follows. In the second chapter "Problem Identification" the problem to be solved is explained. In the third chapter "Theorical Framework" the base theory to understand the actual work is indicated. In the fourth chapter "Methodology" the steps to obtain the results are described. In the fifth chapter "Results" the outcomes of this investigation are presented and discussed. In the next chapters, the conclusions, recommendations, and applicability of this research are described. Finally, the references and annexes are found in the last two chapters.

II. PROBLEM IDENTIFICATION

In the current chapter, the research problem identification is presented, providing a context about its precedents, its definition and how this study helps to solve it. Simultaneously, the research questions and the general and specific objectives that support these questions will be established.

2.1 PRECEDENTS OF THE PROBLEM

In the study of electrical systems, it is crucial to understand the behavior of the demand. In Honduras the per capita consumption of electrical energy per day is 1.9 kWh (Dirección General de Eléctricidad y Mercados, 2019), representing in a year a consumption of 693.96 kWh in a year per Honduran.

With the purpose of disaggregation of electrical demand in recent studies of the National Electric Power Company (ENEE by its acronym in Spanish: Empresa Nacional de Energía Eléctrica), the territory has been divided into three specific regions which are the Northwestern Region, the Atlantic Littoral Region, and the Central-South Region.

There are certain restrictions for this regions, for instance, Gracias a Dios, department of Honduras, is not considered in this division, due to the lack of electrical coverage by the National Interconnected System (SIN by its acronym in Spanish: Sistema Interconectado Nacional) it is reported that 69% of coverage in the department comes from off-grid systems (Empresa Nacional de Energía Eléctrica, 2019). Yoro, another department in the territory, takes part in two of the three regions mentioned, but it is unknown by which percentages.

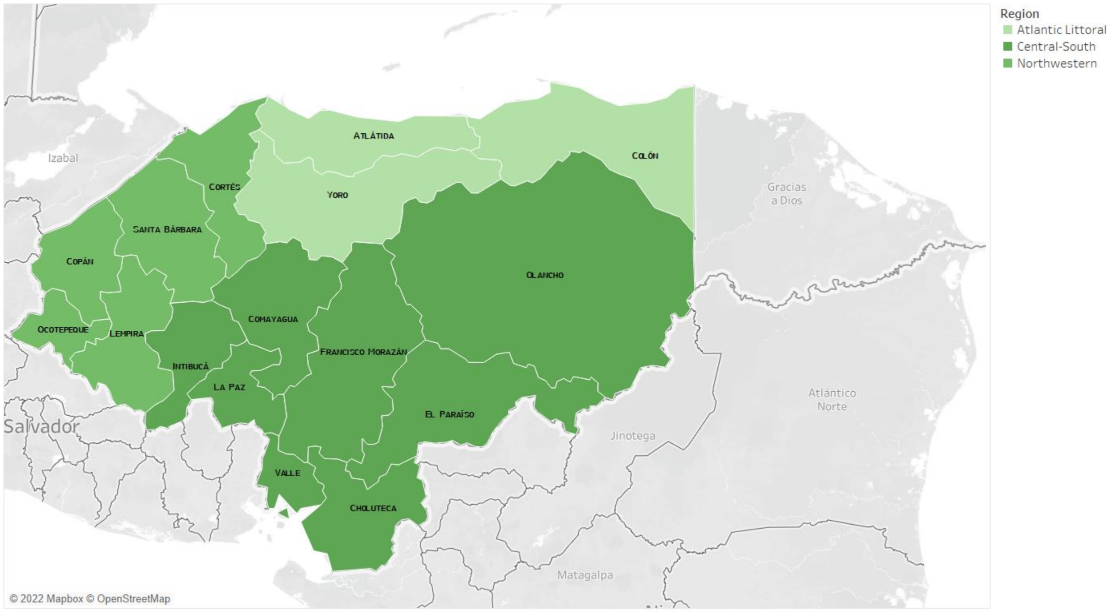


Figure 1: Geographic location of electrical demand regions in Honduras

Source: Own elaboration based on information of (Operador del Sistema, 2022)

The region with the most of electrical energy consumption in the last years (2007-2020) is the Northwestern Region, being accountable of 50% of the consumption in Honduras by 2020. The second region, the Central- South does not vary that much in comparison with Northwestern, being responsible for 40% of consumption by 2020. In the Atlantic Littoral Region, the consumption varies between 9% and 10% through the lapse of time from 2017 to 2020. (Operador del Sistema, 2022).

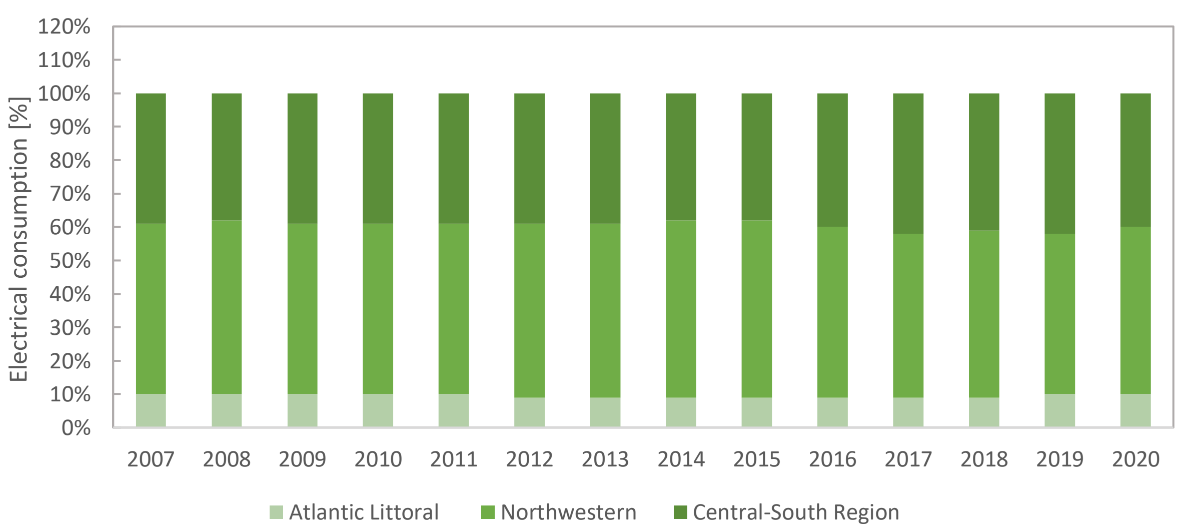


Figure 2: Percentage of electrical consumption by regions in Honduras

Source: Own elaboration based on information from (Operador del Sistema, 2022)

The electrical consumption or demand of a country is not a just a technical term, it can be related with social and economic variables such as growth domestic product (known by its acronym GDP).

> Regarding the relationship between real GDP and energy consumption, these findings show that for all countries, GDP has a positive effect on energy consumption. In the cases of Colombia and Venezuela, a 1% increase in real GDP increases energy consumption by 0.82%. (Jacobo Campo, 2013, pág. 248)

In Central America, comparing the GDP and the electrical consumption of some countries, it is notorious to conclude that they are directly proportional. It is recognizable that economic growth can be positively affected by the consumption of electricity.

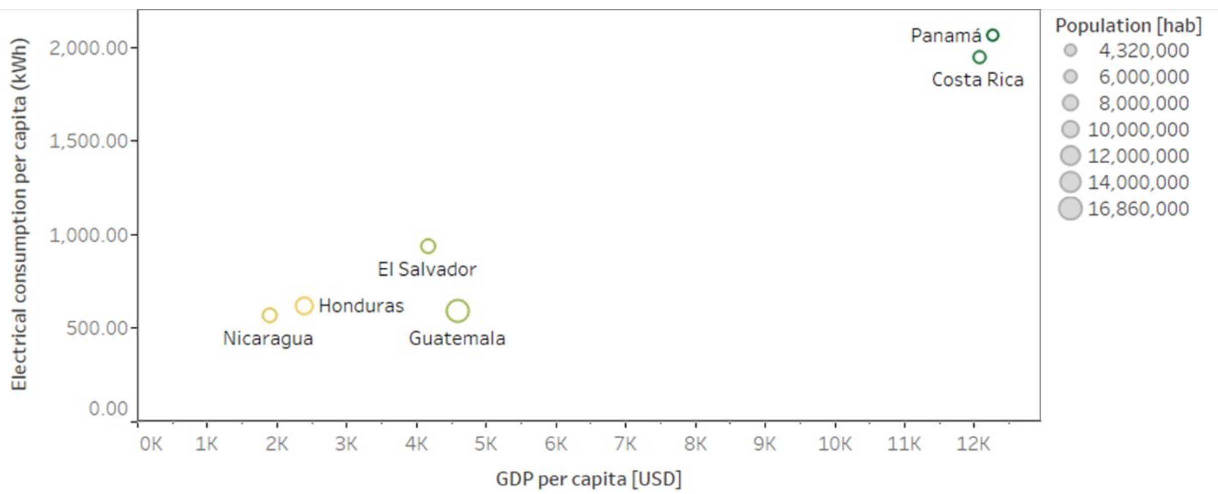


Figure 3: Relationship between energy consumption and GDP per capita in Central America

Source: Own elaboration based on information of (World Bank Open Data, 2020)

Another important variable that is highly affected by electricity consumption and is related to social variables such as life expectancy, education, and per capita income indicators is Human Development Index (known by its acronym HDI).

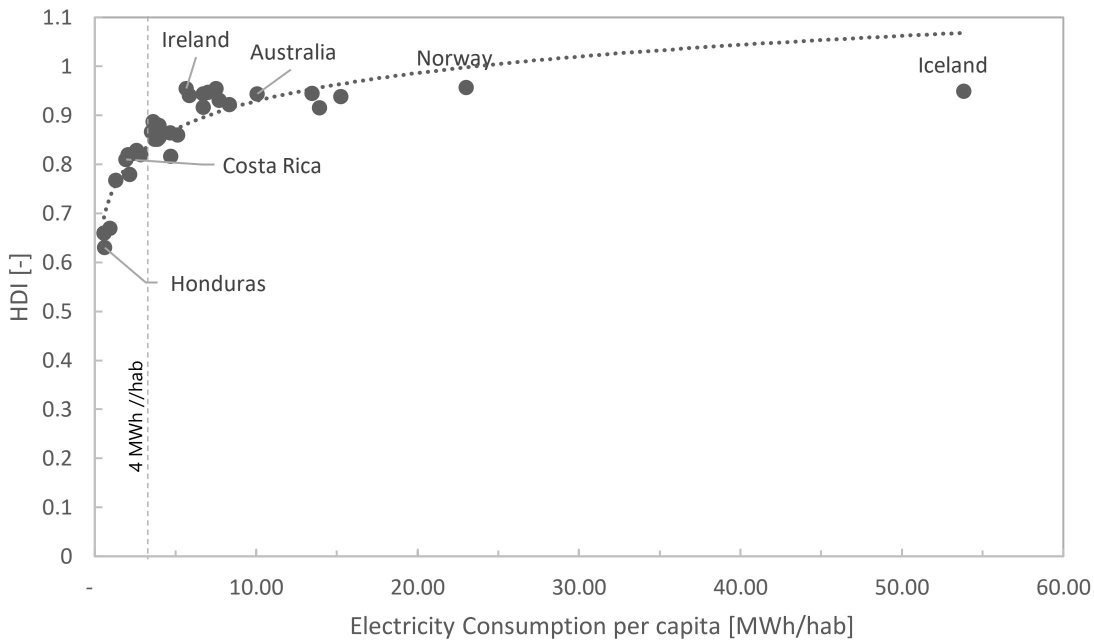


Figure 4: Relationship between Human Development Index and Electricity Use (2014) for thirty Countries including Honduras

Source: Own Elaboration based on information from (World Bank Open Data, 2020)

Pasternak (2000) describes the relationship between these two parameters: "In this correlation, HDI reaches a maximum value when electricity consumption is about 4,000 kWh per person per year, well below consumption levels for most developed countries but also well above the level for developing countries"

Which is true for the scatterplot above, given that Honduras en Costa Rica are below the 4,000 kWh per person per year asymptote as well as countries like Norway and Iceland are well above this value.

Zooming in to understand better the Central America situation, the following graph illustrates the relationship of these variables in the region, where the size of the circle is representative of the population.

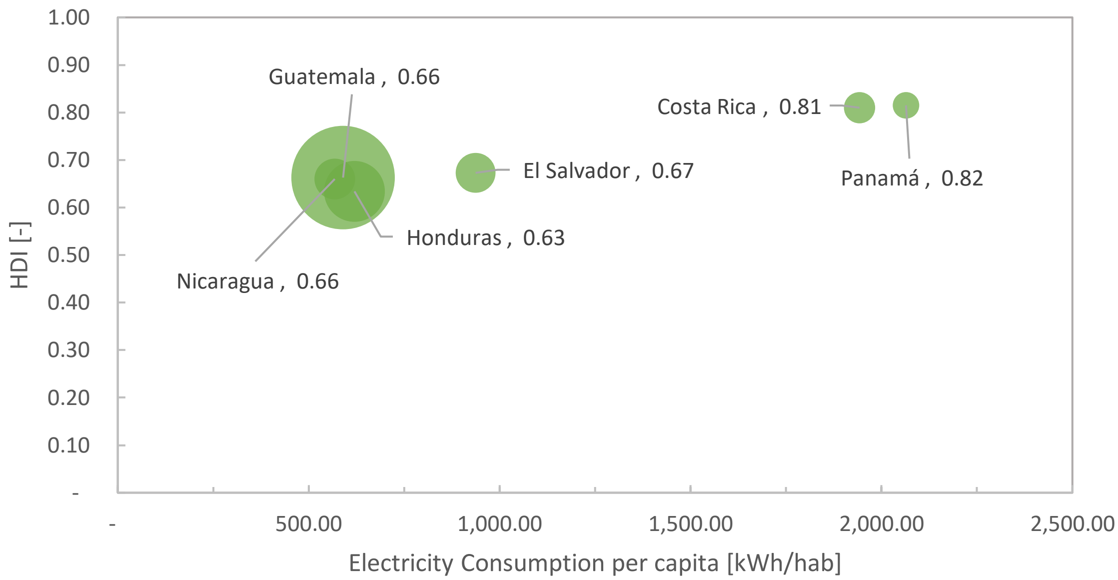


Figure 5: Relationship of HDI and Electricity Consumption per capita by population in Central America

Source: Own elaboration based on information of (World Bank Open Data, 2020)

2.2 Definition of the Problem

The demand of electricity and its growth rate is considered to plan the expansion of the transmission network and generation plants that will be a need in the country in the future ten years. However, the location of the demand is not a variable considered; besides, there is no geographical tool that locates the demand in each department and municipality (this information is just available by the regions mentioned in the previous section).

To create a geographical tool with the purpose to indicate the location and focus of demand in the Honduran territory is an opportunity to establish an information system for academia, industry, and the general population. Also, an invite for the distribution energy company (EEH by its acronym in Spanish: Empresa Energía Honduras) and operator (previously called ODS by its acronym in Spanish: Operador del Sistema, now called CND by its acronym in Spanish: Centro de Despacho Nacional) to start considering distributed generation close to demand, instead of planning for centralized generation with conventional fuels and renewable plants with a high environmental impact.

Distributed Generation (DG) systems can support the reduction of losses in the distribution network since DG systems are placed near to the load and improvement of the voltage in the network due to the reduction in reactive power loss. (K. Balamurugana, 2011).

2.3 JUSTIFICATION

The importance of a geographical tool that locates the demand relies on the fact that in Honduras the actual electrical coverage is 85.02% and the losses in the distribution network are approximately 31.6% of the energy generated (Pineda, 2022). The conditions of the actual electrical system are at risk with a debt of 66,088.10 million of lempiras (407, 256.30 USD; as for 2022: 1 USD = 24.55 HNL) from the statal company ENEE.

Despite this precarious scenario, the Honduran electricity subsector counts with installed capacity as distributed generation and the majority in the residential sector is photovoltaic technology. Unfortunately, the users that have elected to install these systems are still waiting for the payment of their electricity injection to the network (feed-in tariff). This is because in 2007 in Honduras, through decree 404-2013 in the "Ley General de la Industria Eléctrica" it was declared that the distribution companies are forced to buy the excess of energy from renewable sources generated by residential and commercial users which is injected back into the network, crediting them with the values in monthly billing. However, the decree has not been respected, and if the declaration becomes true, it will improve the opportunities of distributed generation and with the demand's location, the installation of the photovoltaic systems or others technology of conversion at small scale will increase in each municipality.

It is stated that "Una generación dispersa y basada en las fuentes energéticas locales, diversifica los recursos y aumenta la autosuficiencia de una región" [A dispersed generation based on local energy sources, diversifies the resources, and increases the self-sufficiency of a region] (Quintero J. P., 2008)

Improving the Hondurans' life quality by having greater ease of access to electricity and improvement of socioeconomical level are objectives that could be accomplished focusing the distributed generation near the demand with the help of the geographical tool developed through this investigation.

With this geospatial information tool, a greater municipal autonomy could be reach, the information of electricity consumption will be public and the decision-makers at municipal level could make judgments based on this information.

2.4 Research Questions

- Which departments and municipalities of Honduras have greater consumption of electricity in a month?
- Which sociodemographic and economic variables seem to have a high impact in the consumption of electricity in Honduras' municipalities?
- What is the per capita consumption of energy for each municipality in Honduras in a month?
- What is the electrical consumption per area for each municipality in Honduras in a month?

2.5 Objectives of the Project

In this section, the general objective of the research is presented, together with the specific objectives that are sought to be fulfilled.

2.5.1 General Objective

To develop a geographical tool to map the electricity demand of each department and municipality in Honduras with the purpose of analysis of electrical consumption and to supply a pioneering tool of planification in expansion and generation by location of demand.

2.5.2 Specific Objectives

1. Show by a Pareto diagram which of the municipalities of Honduras are the main consumers of electricity per capita by March 2022
2. Study electricity consumption at departmental level and the percentages that each municipality represents in the total departmental consumption by March 2022.

3. Provide an interactive cartographic tool in Tableau to track monthly electrical consumption in each of the 298 municipalities by March 2022.
4. Determine the sociodemographic and economic variables which electrical consumption depends on by a linear regression model with multiple input variables.
5. Determine which economical activities have higher impact in the total consumption of electricity in the commercial and industrial sector of electricity consumption by March 2022.

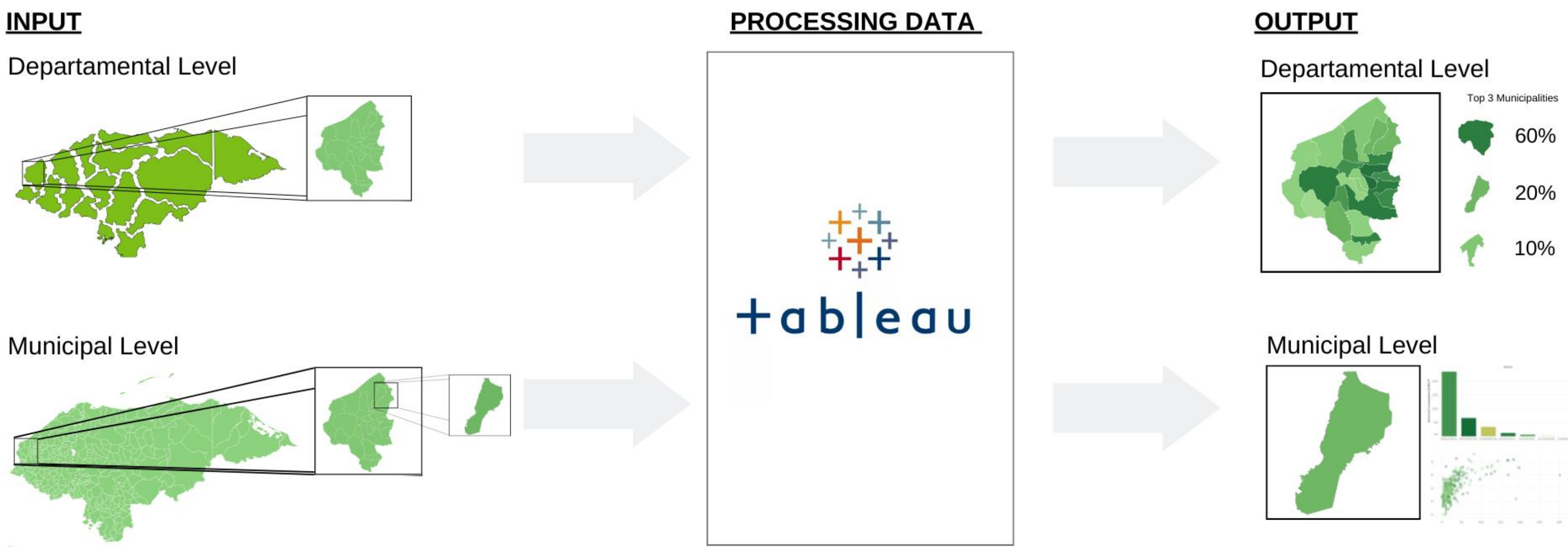


Figure 6: Investigation process diagram of this thesis

Source: Own elaboration

III. THEORETICAL FRAMEWORK

In this chapter, theories that sustain the actual work will be explained with the purpose of demonstrating the reader that this thesis is based on well-established ideas from academia.

3.1 ENERGY PLANNING

Energy planning can be defined as the governmental action of creating long term policies to manage and regulate the national energy system (Bhatia, 2014). While this process is guided by the government, it is necessary the addition of experts in the energy sector such as large and micro energy generators and academia. This last-mentioned actor has the task of analyzing data and providing results (based on the scientific method) that should be consider by decision-makers.

> "In energy planning, the principal emphasis is on the detailed and disaggregate analysis of the energy sector, with due regard for the main interactions within the sector itself, as well as the rest of the economy" (Munasinghe, 1990, pg. 19).

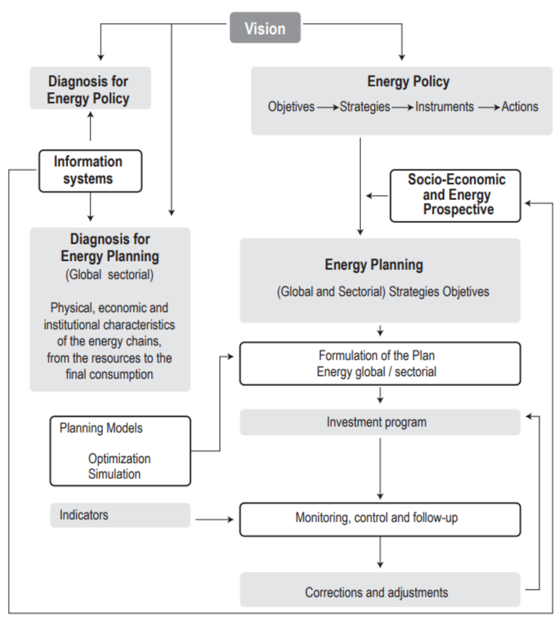


Figure 7: Stages in Energy Planning

Source: (OLADE, 2017) [OLADE by its Spanish acronym Organización Latinoamericana de Energía].

3.2 ENERGY PLANNING TOOLS

In (OLADE, 2017) various energy planning tools are mentioned:

- Economic and energy information systems.
- Energy prospective studies including conditions in an economic and social context at national and international levels.
- Models used for planning in the global and sub-sectorial energy levels.

3.2.1 ENERGY PLANNING TOOLS IN HONDURAS

Considering the tools mentioned in the section 3.2, it is essential to recognize which of these are propelling the advance of energy planning in the national territory.

As economic and energy information systems, Honduras counts with multiple information sources:

a) Instituto Nacional de Estadística (INE)

On behalf of sociodemographic and economic matters the principal entity responsible is the National Institute of Statistics (INE by its Spanish acronym Instituto Nacional de Estadísitica). Nowadays the INE counts with sociodemographic and economic documents per municipality only for the years of 2013 and 2018, resulting in an outdated database.

b) Secretaría de Gobernación, Justicia y Descentralización (SGJD)

Fortuitously by 2017, surges the Ministry of the Interior, Justice, and Decentralization (SGJD by its Spanish acronym Secretaría de Gobernación, Justicia y Descentralización) with the initiative to create the municipality categorization, a database and data visualization tool of public access in Tableau (Secretaría de Gobernación, Justicia y Descentralización, 2020) for a range of years from 1992 to 2020 in which social and economic data for every municipality is recollected and shared geographically.

c) Balance Energético Nacional (BEN)

In aid of the necessity of energy information systems in Honduras, the document "National Energy Balance" (BEN by its Spanish acronym Balance Energético Nacional) is for the first time redacted by the Ministry of Energy in 2017 as a pioneering tool of energy planning. Since the first publication in 2017, the BEN has been annually published until 2020.

d) Sistema de Información Energética de Honduras

In the BEN 2020, an information system created by the OLADE is highlighted. OLADE created Sie Honduras (by its Spanish acronym Sistema de Información Energética) with the purpose of making simple the access to data and energy statistics for the period of 2010 to 2020.

e) Energy Observatory of Honduras

One of the most recently created tools of energy information is the Energy Observatory of Honduras, created by the initiative of the academia in the Engineering Faculty of the Universidad Tecnológica Centroamericana (UNITEC), with the purpose of planning and monitoring generation and demand at a national level (Flores, 2021). Energy generated per technology, national electricity demand, energy prices, electrical coverage and energy losses are main characteristics to be visualized in the observatory.

For models used in planning in the global and sub-sectorial energy levels, in Honduras certain models are used for energy generation and demand forecasting:

f) Forecast of energy generation in Honduras

The electrical subsector counts with an actual long-term plan spanning from 2022 to 2031, redacted by the operator of the electrical system.

The system Operator in Honduras utilizes a computational tool for generation expansion planning and regional interconnections called Opt Gen, in specific the model of Stochastic Hydrothermal Dispatch with network restrictions (SDDP) which allows to calculate the operating and production costs of the generation plants. The model also understands the actual technical restrictions and the actual scenario of the energy matrix. As an output, the model brings a response: the forecast of technic and economical dispatch in Honduras.

For the electrical subsector planning, Time Series Lab (TSL) is crucial to understand the behavior of renewable intermittent resources. The TSL was created by PSR Energy Consulting and Analytics (that is also the company responsible for SDDP and OptGen). TSL is applied to model the energy shedding of the non-conventional renewable technologies in the energy matrix, for instance, solar and wind energy.

In (Operador del Sistema, 2022) is mentioned the software (OptGen, SDDP and TSL) is fed by five main variables:

• Forecast of electrical demand (focus of this thesis)

• Details of the generator's operation and its future modifications

• Details of projects in development of generic type

• Technical characteristics of the system

• Required economic parameters

g) Forecast of electrical demand in Honduras

Actual forecasting of national electricity consumption in Honduras depends on independent variables such as Gross Domestic Product (GDP), energy prices, growth in the number of users, industrial added value, and the price of industrial machinery (Empresa Nacional de Energía Eléctrica, 2020).

The methodology selected to forecast the demand of each energy consumption sector is based on trend techniques, econometric analysis, double logarithmic regression, and expert judgment.

In Honduras, the projection of demand is carried out by sectors which are:

1. Residential
2. Commercial
3. Industrial: divided by ENEE into moderate and large industries.
4. Autonomous
5. Government

6. Municipal
7. Community service

The last four sectors (4,5,6 and 7) are grouped by ENEE into a larger sector called "Other Consumers" due to the small consumption they individually represent.

In the following table the methodology and independent variables necessary for each sector are indicated:

Table 1: Forecasting methodology of electricity demand by sector

Sector	**Methodology**	**Independent Variables**	**Dependent Variable**
Residential	Econometric Hybrid Model	Number of subscribers [-]	Annual demand growth rate [%]
		Price of electricity and kerosene as a substitute [$]	
		Gross Domestic Product (GDP) [$]	
		Elasticity of electrical demand [-]	
Commercial	Trend model and expert judgment	-	Annual demand growth rate [%]
Industrial (Moderate)	Econometric Hybrid Model	Industrial Aggregated Value (VAI)	Annual demand growth rate [%]
		Price of electricity [$]	
		Price of machinery [$]	
	Double log autoregressive model	Industrial Aggregated Value (VAI)	Annual demand growth rate [%]
		Price of electricity [$]	
		Price of substitution [$]	
		Electrical Consumption [MWh]	

Industrial (Large)	Interviews through the application of the Annual Technical Survey to Large Consumers	Expansion plans in term of five years	Annual demand growth rate [%]
Other Consumers	Double log autoregressive model	-	Annual demand growth rate [%]
	Trend model	-	

Source: Own elaboration based on information of (Empresa Nacional de Energía Eléctrica, 2020)

There are three main disadvantages of electrical demand planning in Honduras:

1. The methodologies are not accessible to academia or the general population.
2. The demand is projected at the national level and best-case scenario electrical consumption by regions is studied but not considered in planning. There is a need for data disaggregation.
3. Consequently, the information is centralized at a governmental level and does not provide space for autonomous municipal development and energy planning in Honduras.

3.3 FORECAST OF ELECTRICITY DEMAND MODEL

According to Esteves et al (2015) the models utilized electricity demand forecasting are:

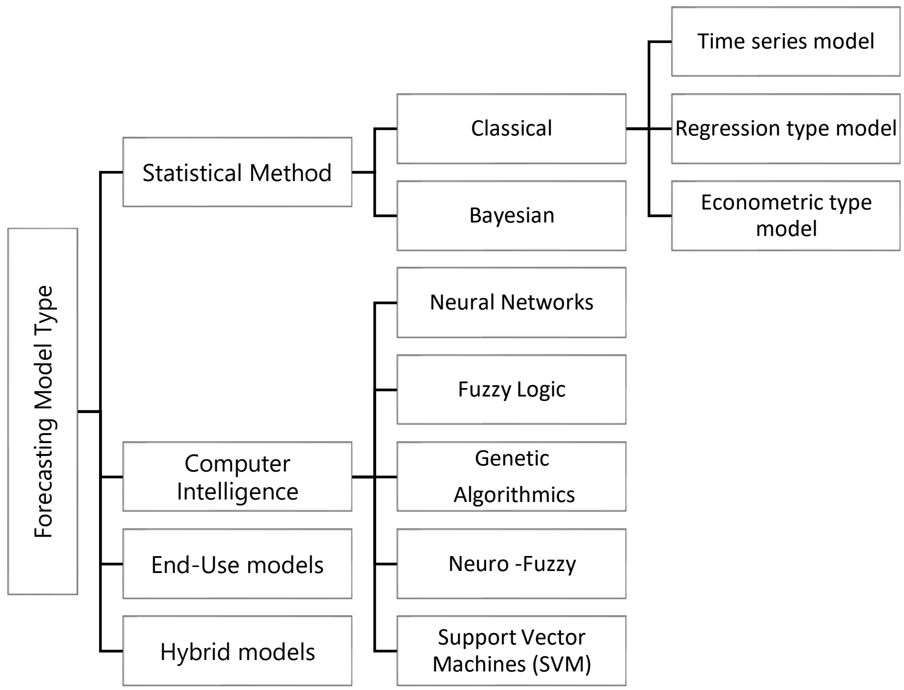


Figure 8: Forecasting Model Type in academic theses

Source: Own elaboration based on information from (Esteves et al, 2015)

It is known for a fact that statistical methods stand as the most studied model representing 59.3% of the methodologies utilized and 67.3% of the cited articles (Esteves et al, 2015). From statistical methods, the long-term forecasts with high level of utilization are Regression (30.3%) and Econometric (30.3%) methodologies.

Walpole et al (2012) defines regression analysis as : "Finding the best relationship between *y* and *x*, quantifying the strength of that relationship, and using methods that allow for prediction of the response values given values of the regressor *x*" (p. 389).

Where:

- y is known as a dependent variable or response

- x is an independet variable or a so called regressor in statistics

3.3.1 UNIVARIABLE LINEAR REGRESSION

The most ordinary type of regression model is linear regression, through the utilization of this model is possible to create a function *(y)* that depends on a unique regressor *(x)*. Due this last characteristic of the regressor the linear regression is also called an univariable regression model.

$$y_i = \beta_0 + \beta_1 \cdot x$$

Equation 1: Linear regression function

Source: (Walpole et al, 2012)

Where:

- x is the independent variable
- y is the dependent variable
- β_0 is the intercept of the regression model
- β_1 is the slope coefficient of the independent variable

In an univariable regression model is also natural to find a component β_0, which is described as an intercept in the dependent variable axis *(y)* (Walpole et al, 2012). When β_0 is zero the function intercepts the origin of the graph. Graphically, is evident that the component β_0 represents an upward or downward movement along the y-axis for the linear function as shown in the following graph.

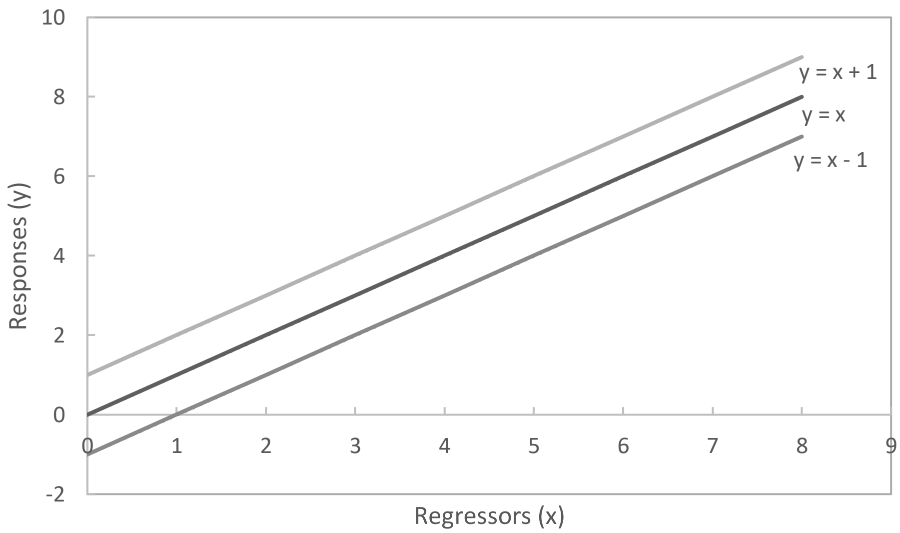


Figure 9: Evaluation of distinct values of interception in the y-axis (*β_0*)

Source: Own elaboration

Other important value is β_1 in equation 1, which is the slope of the linear function which defines the ratio between change in the dependent variable (*Δy*) and change in the independent variable (*Δx*).

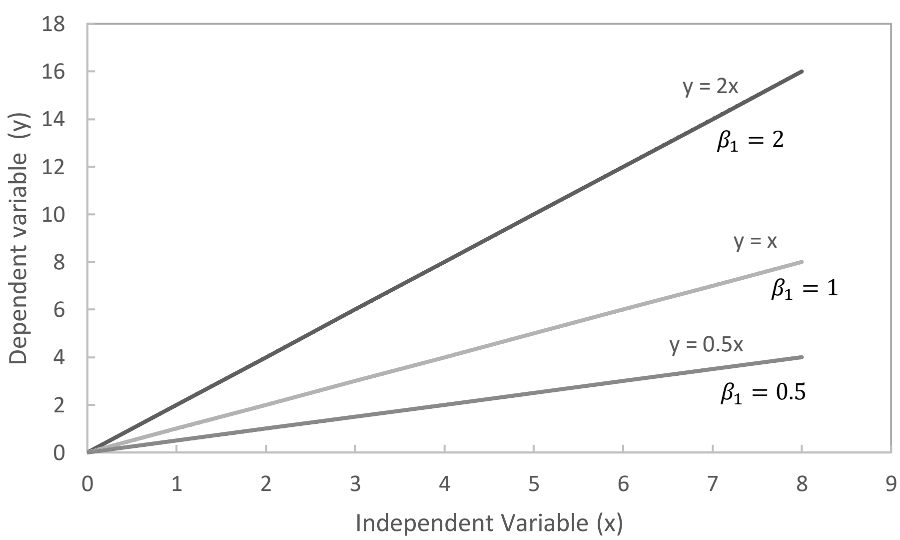


Figure 10: Evaluation of distinct values of slope (*β_1*)

Source: Own elaboration

The coefficients β_0 and β_1 are to be determined with the objective of reducing the error between the projected function ($\hat{y}_i$) and the real value of the response (y_i) for every regressor (x_i) where i has its limit in n, being n the number of pairs (x_i, y_i) available in the data given to train the regression model. (Ng, 2012)

For this to be achieved is necessary to find a linear function where the mean squared error (MSE) is at its minimum therefore minimizing the error ($y_i - \hat{y}_i$) for every i.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2$$

Equation 2: Definition of Mean Squared Error

Source : (Walpole et al, 2012)

Where:

- y_i is the real dependent variable
- $\hat{y}_i$ is the predicted dependent variable
- n is the number of samples

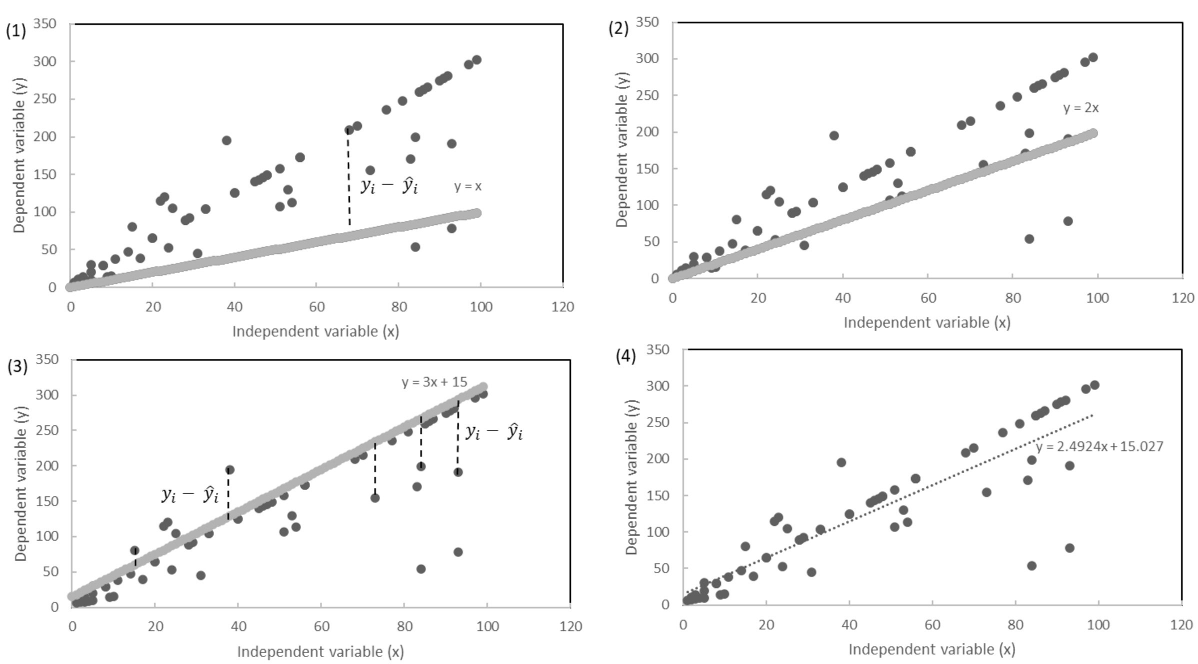


Figure 11: Graphical intuition of minimizing MSE

Source: Own elaboration

To build intuition in what it means to minimize the MSE, the graph above explains:

In (1) data recollected has been graphed, this data is usually called training set. Training set can be defined information that helps train the regression model. Also, a function is proposed ($\hat{y}$) where $\hat{y} = x$, but it doesn't seem to be the best fitted linear function for the data because the difference between where $\hat{y}$ and y is notable for every i.

By trial and error, in scenario (2) another regression model is proposed where $\hat{y} = 2 \cdot x$ (β_0 = 0 and β_1= 2), it is undoubtedly a better approach but not the best fitted linear function yet.

In scenario (3) where $\hat{y} = 3 \cdot x + 15$ (β_0=15 and β_1=3) most of the data seems to be fitted in the linear function but there are still notable differences between $\hat{y}$ and y for the rest of the data that doesn't allow MSE to be minimize.

Scenario (4) succeeds, after infinitely trials, the best fitted function is revealed, in this scenario the coefficients for β_0 and β_1 are 15.027 and 2.4924 respectively.

To obtain the function more immediately and without the need for infinite tests, various software (Excel, Minitab or R) performs this calculation immediately. Technologies such as machine learning with algorithms such as gradient descent are also able to predict a linear regression model.

Fumo et al (2015) introduces the utilization of univariable regression analysis to predict the energy consumption of a building based on climatological data in the site. Also utilizes multiple linear regression analysis to forecast energy consumption and concludes in the aggregation of the more independent variables in the study (in this case in particular drybulb temperature and solar radiation) improves the quality of the model.

3.3.2 MULTIPLE LINEAR REGRESSION

It is described that a multiple linear regression is necessary once the variable (y) doesn't depend just in one independent variable, but two or more.

Due to the aggregation of more independent variables, the coefficients now cannot be just β_0 and β_1 as they were in Univariable Lineal Regression. The coefficients, except β_0, depend on the number of independent variables and are equal in number. This quantity of independent variables is known as m.

$$\mathrm{y_i} = \beta_0 + \beta_1 \cdot \mathrm{x_{i1}} + \beta_2 \cdot \mathrm{x_{i2}} + \ldots \beta_m \cdot \mathrm{x_{nm}}$$

Equation 3: Multiple Linear Regression function

Source: (Walpole et al, 2012)

Where:

- y_i is the dependent variable.
- x_{ij} is the independent variable for the sample *i* and the feature *j*.
- β_j is the coefficient factor for each independent variable.
- n is the total number of samples.
- m is the total number of features or independent variables.

Usually, in multiple linear regression linear algebra is used, to reduce various values of a variable into a vector. Where the dot product between the vector of *x* and the vector of beta leads to obtain a scalar as the dependent variable.

$$\mathrm{Y} = \beta_0 + \vec{\beta} \cdot \vec{X}$$

Equation 4: Multiple Liner Regression function represented by vectors

Source: (Ng, 2012)

Where:

- Y is the dependent variable.
- $\vec{X}$ is the vector that contains all independent variables.
- $\vec{\beta}$ is the vector that contains all slope coefficients.
- β_0 is the intercept.

$$\vec{\beta} = [\beta_1, \beta_2, \beta_3, \ldots \beta_n]$$

$$\vec{X} = [x_1, x_2, x_3, \ldots x_n]$$

Equation 5: Definition of vectors

Source: (Ng, 2012)

Torres (2015) describes that multiple regression methodology applied to the consumption of electrical energy explores the relation of demographic independent variables with the dependent variable. Performing multiple linear regression model, attributes that influence consumption that are not obvious could be identified or attributes that seem to be highly correlated with electrical consumption could be proven to have zero influence.

3.3.3 NON-LINEAR REGRESSION

Data is modeled by a nonlinear function when the parameters are in a nonlinear combination and the independent value depends on one or more independent variables, for instance a nonlinear model is an exponential function.

Table 2: Comparison between linear and non-linear functions

Non-Linear Function	Linear Function
$\hat{y} = A \cdot {\beta_0}^x$	$\hat{y} = \beta_0 + \beta_1 \cdot x_1^1 + \beta_2 \cdot x_2^3 + \cdots \beta_2 \cdot x_p^r$

Source: Own elaboration based on information from (Walpole et al, 2012)

> "Confusion arises occasionally when we speak of a polynomial model as a linear model. However, statisticians normally refer to a linear model as one in which the parameters occur linearly, regardless of how the independent variables enter the model" (Walpole et al, 2012)

The applicability of this type of regression model in electricity demand forecasting has been studied, in Moral-Carcedo et al (2005) the relation between electricity demand and temperature is clearly nonlinear, both increase and decrease in temperature signifies a high consumption of energy.

3.3.4 QUANTILE REGRESSION

Quantile regression is often used to model specific conditional quantiles of the response *(y)*, is described as a semi parametric model in which there is no prior hypothesis of the relation between independent value and dependent values

Quantile regression estimates the conditional median contrary as linear regression that uses the method of least squares finding the conditional mean. (Dye, 2020)

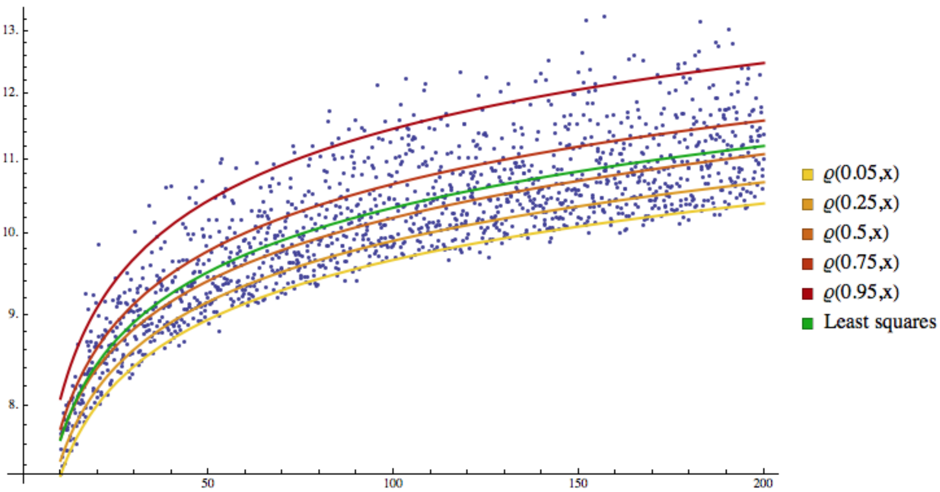


Figure 12: Data fitted with quantile regression

Source: (Dye, 2020)

The regression model for quantile level τ of the response is described as:

$$Q_{\tau}(y_i) = \beta_0(\tau) + \beta_1(\tau) \cdot x_{i1} + \beta_2(\tau) \cdot x_{i2} + \ldots . \beta_m(\tau) \cdot x_{nm}$$

Equation 6: Quantile regression function

Source: (Rodriguez , Yao, & SAS Institute Inc, 2017)

Where:

- $Q_{\tau}(y_i)$ is the response or dependent variable in the specific quantile.
- τ is the quantile level.
- x_{ij} is the independent variable for the sample i and the feature *j*.

- $\beta_j(\tau)$ is the coefficient factor for each independent variable in the specific quantile level.
- n is the total number of samples.
- m is the total number of features or independent variables

Where instead of minimizing the mean squared error as it was described for linear regression in section 3.3.1, the function to minimize is the median absolute deviation (MAD).

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_{i-}\beta_0(\tau) - \sum_{j=1}^{m}x_{ij}\cdot\beta_j(\tau)\right)$$

$$min_{\beta_0(\tau)\ldots\beta_p(\tau)} \text{in MAD}$$

Equation 7: Definition of median absolute deviation

Source: (Rodriguez , Yao, & SAS Institute Inc, 2017) and (Dye, 2020)

Where:

- y_i is the real dependent variable.
- $\beta_0(\tau)$ is the intercept at a specific quantile level.
- $\beta_j\ (\tau)$ is the coefficient factor for each independent variable at a specific quantile level.
- x_{ij} is the independent variable for the sample i and the feature *j*.
- n is the total number of samples.
- m is the total number of features or independent variables.
- ρ_τ is the check loss.

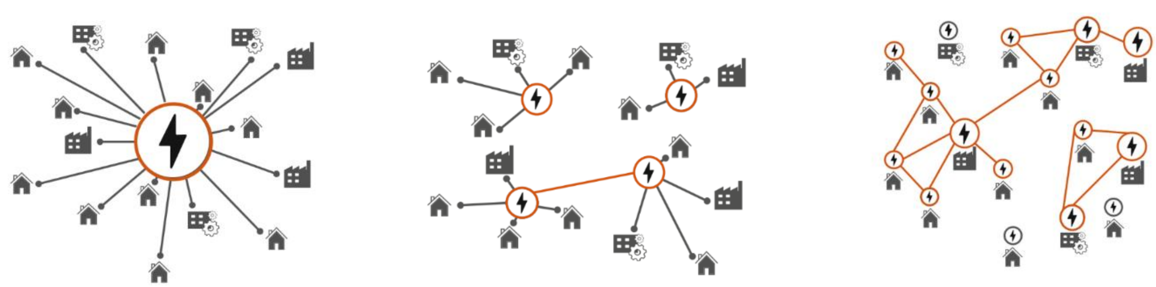
Medina (2011) carried out an investigation where the electrical demand in Spaniards households is foreseen, requiring the use of quantile regression comparing it to the method of least squares.

3.4 TYPES OF GENERATION SYSTEM STRUCTURES

According to Vezzoli (2018) there are three types of generation system structures:

1. **Centralised energy systems** have great capacity installed and are far from the point of use or consumers.
2. **Decentralized energy systems** represent small-scale energy generation units that provide electrical energy to local customer, due to this fact the distribution network is not large. These structures can be connected to each other to share resources such as energy surplus.
3. **Distributed energy systems** also represent a small-scale energy generation, but unlike decentralized energy systems, in distributed generation systems the users are the producers.

Generally, Distributed Generation (DG) is defined as the installation and operation of an electrical power generation unit connected to the costumer side of the meter (or the distribution network). The capacity in this generation units can vary between 1 W to 300 MW, suggesting that capacity of a generator does not have an impact in whether is considered DG or not DG. Distributed Generation can utilize renewable and not renewable resources (Ackermann, 2001).



a. Centralized energy systems **b. Decentralized energy systems** **c. Distributed energy systems**

Figure 13: Comparison of energy system structures

Source: Adapted from (Vezzoli, 2018)

The concept of energy flow being unidirectional has been modified with distributed generation which results in the modification of the electrical system itself including the improvement of distribution lines and installment of technical safety equipment to ensure a constant and secure power supply.

Although in many countries there is already the necessary regulation that promotes the use of cogeneration technologies and renewable distributed generation, there are still many social, economic, and technological barriers to be crushed to advance in the implementation. (Santamaria et al , 2014)

In Santamaría et al (2016) a method for assessing the status of distributed generation in a country called Distributed Generation Legislation Index (DGLI) is developed considering five aspects:

- Liberalized electricity market (A_1)
- Net metering—net billing (A_2)
- Renewable energy programs (A_3)
- Reduction of import duties for clean technologies and tax incentives for renewable energy resources (A_4)
- DG operation and lack of tolls for use of grids (A_5)

$$DGLI = p_1 \cdot A_1 + p_2 \cdot A_2 + p_3 \cdot A_3 + p_4 \cdot A_4 + p_5 \cdot A_5$$

Equation 8: Definition of Distributed Generation Legislation Index

Source: (Santamaría et al, 2016)

The aspects $(A_1, A_2, A_3, A_4, A_5)$ are allowed to be 1 or 0 values, 1 for the affirmative case of the aspect and o for the negative case of the aspect. This equation links the aspects mentioned with the DGLI with the use of a parameter p, where:

$$p = p_1 = p_2 = p_3 = p_4 = p_5 = 20$$

Equation 9: Definition of parameters p

Source: (Santamaría et al, 2016)

Country	IR_{DG}					
	(A_1) Liberalized electricity market	(A_2) Net metering –net billing	(A_3) Renewable energy programs	(A_4) Reduction of import duties for clean technologies and tax incentives for RES and DG operation	(A_5) Not tolls for use of grids	
Mexico	X	X	X			60
Belize	X					20
Guatemala	X	X	X	X	X	100
El Salvador	X	X	X	X		80
Honduras	X	X	X	X		80
Nicaragua	X		X	X		60
Costa Rica	X		X	X		60
Panama	X	X	X	X	X	100
Jamaica		X	X			40
Dominican Republic	X	X	X	X		80
St. Kitts and Nevis			X			20
Antigua and Barbuda			X			20
Dominica		X	X			40
Saint Lucia			X	X		40
St. Vincent and the Grenadines			X	X		40
Barbados			X	X		40
Grenada		X	X			40
Trinidad and Tobago			X	X		40
Colombia	X	X	X	X		80
Venezuela			X			20
Brazil	X	X	X	X		80
Ecuador	X		X	X		60
Peru	X	X	X	X	X	100
Bolivia	X		X			40
Chile	X	X	X		X	80
Paraguay			X			20
Uruguay		X	X	X	X	80
Argentina	X		X	X		60

Figure 14: Evaluation of regulatory aspects for the promotion of DG in Latin America and the Caribbean

Source: (Santamaría et al, 2016)

3.4.1 DISTRIBUTED GENERATION IN HONDURAS

Honduras' Distributed Generation Legislation Index (DGLI) is 80 out of 100, this indicates a high support of governmental authorities to impulse distributed generation in Honduras, the only aspect missing is the opportunity of grid usage without economical repercussion as noted in Figure 14. Showing a high performance of DGLI, Honduras seems to have everything in its favor for the greater penetration of distributed generation; however, although the regulation has been published, many of the laws that promote selfconsumption are not complied, the energy injected into the distribution lines is not paid at residential and commercial levels.

Distributed Generation is defined differently in Honduras from the general definition of the term. Nationally DG is defined as those energy generation plants connected to the distribution network in specific at 34.5 kilovolts voltage level distribution.

Based on the prior definition, in 2010, 16% of the energy generated in Honduras was recognized to proceed from distributed generation, but only 9% of this distributed generation comes from renewable sources, 4% represented by minor hydroelectric and 5% by biomass. (Figueroa Rivera, 2012)

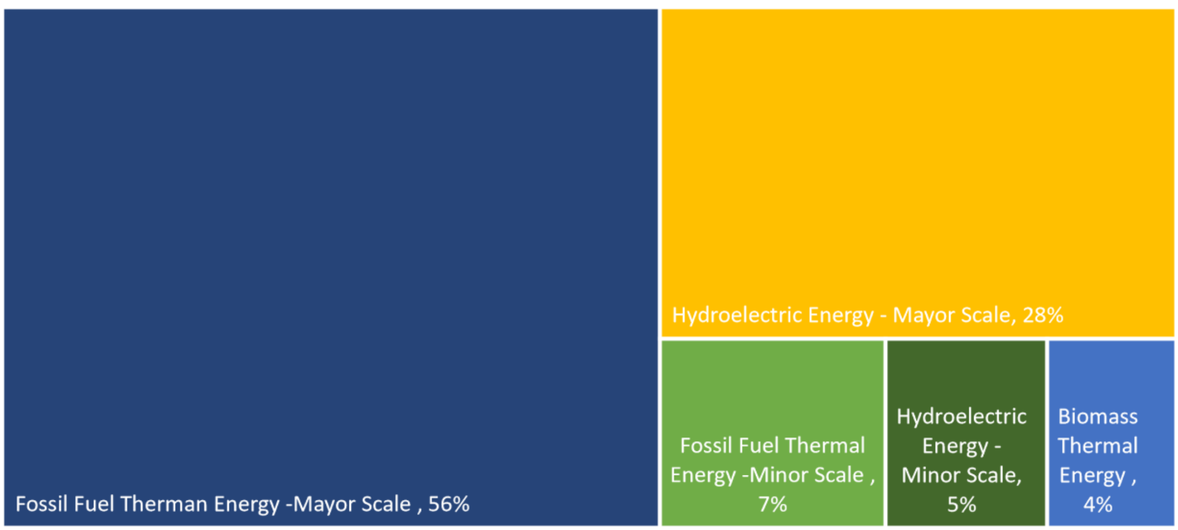


Figure 15: Participation of DG sources in the electricity subsector in Honduras 2010

Source: Own elaboration based on information from (Figueroa Rivera, 2012)

IV. METHODOLOGY

The methodology is the collection of logical steps taken to accomplish the general objective and specific objectives that steers a research study.

4.1 APPROACH

The approach of this investigation is mixed, it is considered quantitative since the data will be analyzed statistically and the correlation between dependent and independent variables will be achieved through regression analysis. It is considered qualitative because to analyze the results, interpretive richness is achieved thanks to this approach.

It is intended to design a cartographic tool in Tableau where monthly electricity consumption is identified per municipality as well as other electricity related data such as electrical consumption per capita, electricity consumption per subscriber and electricity consumption per square meter. It is also expected to analyze the relationship (or lack of association) between sociodemographic data and electrical consumption data for each municipality through a linear regression model.

4.2 RESEARCH VARIABLES

4.2.1 INDEPENDENT VARIABLES

Independent variables for this research were selected to explore the influence that socioeconomic and sociodemographic data has on electricity consumption in Honduras. Ten variables were selected, if one is discarded due to no influence in the final electricity consumption, this variable could be substituted with one of the remaining nine variables. This process of discarding and substituting will be repeated the number of iterations necessary to obtain only independent variables that show correlation with the dependent variable.

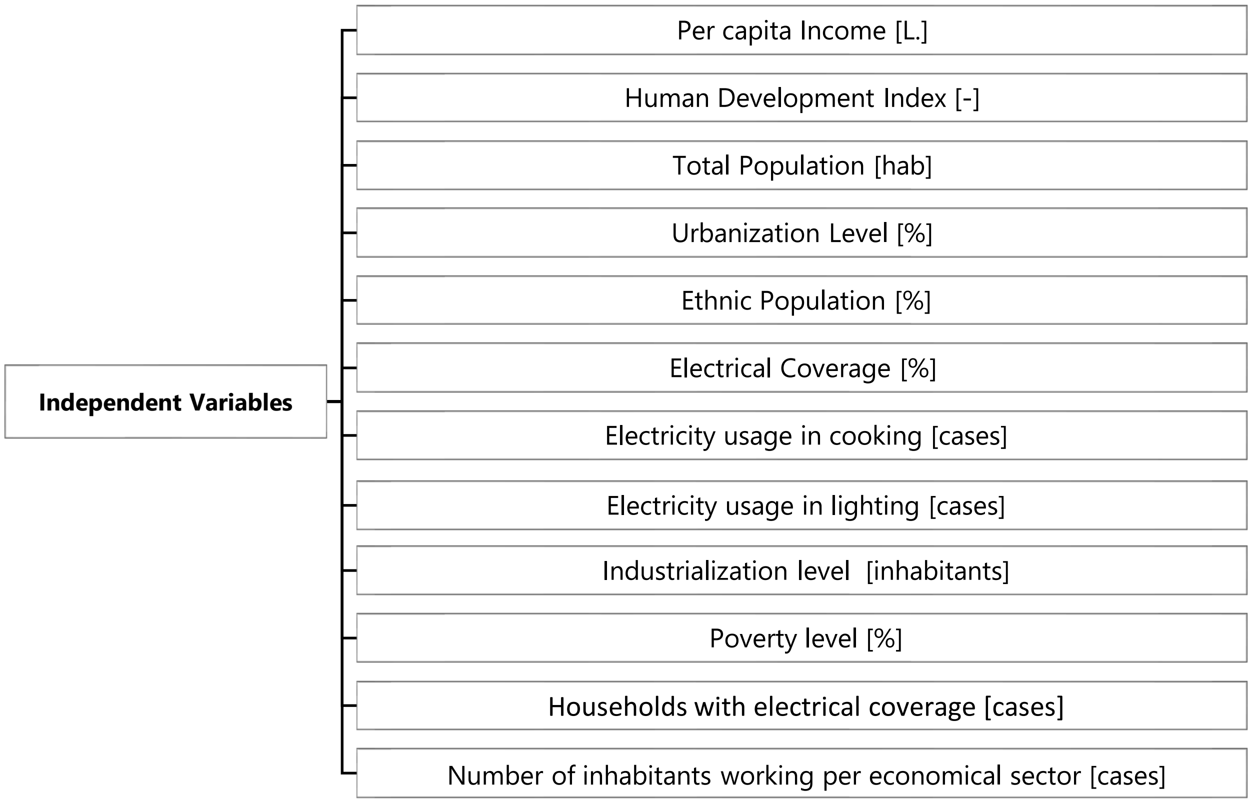


Figure 16: Studied independent variables

Source: Own Elaboration

a) Per capita income [L.]

Its construction is done through the quotient of the municipality´s income divided by the total population of the municipality. It can be assumed that the higher the value of per capita income, the better development scenario the municipality presents.

$$Per\ capita\ income = \frac{Total\ income\ of\ municipality\ [L.]}{Total\ population\ [hab]}$$

Equation 10: Definition of per capita income

Source: (Secretaría de Gobernación, Justicia y Descentralización, 2020) *b) Human development index [-]*

The Human Development Index (HDI) is known to contribute as a measure of general humanitarian development for a region. A high HDI means a great life expectancy, education index and gross domestic product. The data obtained is a projection for 2020 done by the Honduran government.

$$HDI = \frac{1}{3}(Life\ expectancy\ index) + \frac{1}{3}(Education\ index) + \frac{1}{3}(\ GDP\ index)$$

Equation 11: Definition of Human Development Index

Source: (United Nations Development Programme, 2007)

c) Total population [hab]

When referring to total population, a projection of inhabitants per municipality for 2021 done by the government of Honduras was considered for this thesis (Secretaría de Gobernación, Justicia y Descentralización, 2020).

d) Urbanization Level [%]

The urbanization level of a municipality is obtained dividing the number of households that are in urban areas by the total number of households in the municipality. The greater the urbanization level there are less habitant in rural conditions. Data was obtained for the year of 2020. It is expressed in percentage; it varies from 0% to 100%.

$$Urbanization\ Level = \frac{Households\ in\ urban\ areas}{Total\ households\ in\ municipality}$$

Equation 12: Definition of Urbanization Level

Source: (Secretaría de Gobernación, Justicia y Descentralización, 2020) *e)* *Ethnic Population [%]*

This variable has been considered since it is desired to explore electricity consumption in regions where ethnic society focus on Honduras. Percentage of ethnic population is calculated dividing the ethnic population of a municipality by the total population in the same area. Data was obtained for the year of 2013. It is expressed in percentage; it varies from 0% to 100%.

$$Ethnical\ Population = \frac{Ethnical\ population\ [hab]}{Total\ population\ [hab]}$$

Equation 13: Calculation of percentage of ethnic population

Source: Own deduction

f) *Electrical Coverage [%]*

Electrical coverage represents which regions have the necessary infrastructure to transmit and distribute electrical energy to the final consumers, in specific residential subscribers, in the municipality. The data obtained is from the year 2019. It is expressed in percentage; it varies from 0% to 100.

$$Electrical\ Coverage = \frac{Households\ with\ access\ to\ the\ distribution\ network}{Total\ households\ in\ municipality}$$

Equation 14: Calculation of percentage of electrical coverage

Source: (Empresa Nacional de Energía Eléctrica, 2019)

g) *Electricity usage in cooking [cases]*

In Honduras, electricity utilization is not common for cooking in most of the rural areas, firewood is preferred for its economic value and accessibility. Due to this peculiarity, this variable is important to understand if this behavior at least represents a negative correlation with respect to total electricity consumption. The data is obtained for the year 2013.

h) Electricity usage in lighting [cases]

In Honduras, various ways of luminous energy are considered to meet the need for lighting, these alternatives are kerosene lamp, wax candles, solar photovoltaic modules, private electrical system or the most used in the cities the electricity of the public lighting system. The data is obtained for the year 2013.

i) Industrialization level [inhabitants]

The occupation level in industry will be determined by the sum of all industrial and commercial level jobs that the population have in each municipality.

j) Poverty level [%]

The poverty level of a municipality is obtained through the population census from 2013 in Honduras. Based on these unsatisfied basic needs, the poverty level for each municipality was obtained by the INE. It is expressed in percentage; it varies from 0% to 100%.

k) Number of electrified households [cases]

The number of electrified households per municipality was obtained for the sectorial data results in specific for the residential sector in Honduras. The data source in this case is ENEE for the year of 2019.

l) Number of inhabitants working per economical sector [cases]

Is obtained through the population census from 2013. The number of inhabitants working per economical sector was obtained for the sectorial data results in specific for the commercial and industrial sector in Honduras. With the objective of understanding the relationship between electrical consumption and the inhabitants working on commercial or industrial sector.

4.2.2 DEPENDENT VARIABLES

It was already intuited in the section 4.2.1, that electricity consumption was considered a dependent variable for this study. There are other concepts that derived from electricity consumption, that are also considered to be dependent variables such as electricity consumption per area, per capita and per subscriber.

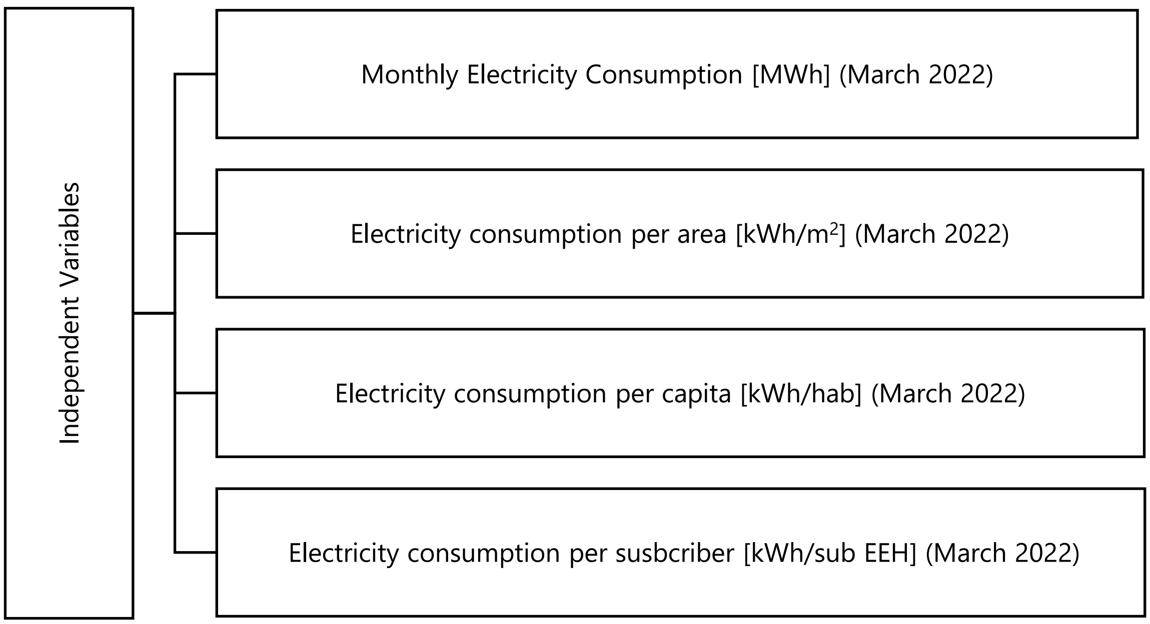


Figure 17: Studied independent variables

Source: Own Elaboration

a) Electricity Consumption [MWh]

The total electricity consumption per municipality was obtained via email in direct contact with EEH which is the national electrical distribution company in Honduras for a sample of 259 municipalities.

It is necessary to mention that a comparison between municipalities based on total electricity consumption is not totally valid because is a value that has not been normalized. Better approaches are developed comparing factors such as electricity consumption per area and electricity consumption per capita.

b) Electricity consumption per area [kWh/m²] or Electricity Density

The electricity consumption per area was obtained dividing the total electricity consumption of the municipality by the total superficial area of the municipality. This factor normalizes the consumption of electricity, this makes the comparison between municipalities valid.

$$Electicity\ Density = \frac{Total\ electricity\ consumption\ [kWh]}{Total\ superficial\ area\ [m^2]}$$

Equation 15: Definition of Electricity Density

Source: Own Deduction

c) Electricity consumption per capita [kWh/hab]

The electricity consumption per capita was obtained dividing the total electricity consumption of the municipality by the total number of inhabitants. This factor normalizes the consumption of electricity, this makes the comparison between municipalities valid.

$$Electicity\ consumption\ per\ capita = \frac{Total\ electricity\ consumption\ [kWh]}{Total\ number\ of\ inhabitants\ [hab]}$$

Equation 16: Definition of electricity consumption per capita

Source: Own Deduction

d) Electricity consumption per subscriber [kWh/sub EEH]

The electricity consumption per subscriber was obtained dividing the total electricity consumption of the municipality by the total number of subscribers to EEH. This factor normalizes the consumption of electricity, this makes the comparison between municipalities valid.

$$Electicity\ consumption\ per\ subscriber = \frac{Total\ electricity\ consumption\ [kWh]}{Total\ number\ of\ EEH\ suscribers\ [hab]}$$

Equation 17: Definition of electricity consumption per subscriber

Source: Own deduction

4.3 TECHNIQUES AND INSTRUMENTS APPLIED

This section will describe the essential tools for obtaining results and monitoring the methodology programmed for this research.

4.3.1 MICROSOFT EXCEL

This spreadsheet specialized software will be used to tabulate the input data of the lineal regression model and the data for the generation of maps and visualizations in Tableau. The 11 variables will be tabulated including dependent and independent variables for 259 municipalities in Honduras, a total of 2,590 inputs will be manually introduced in Excel. The graphs to explore the correlation between the 10 sociodemographic variables and electricity consumption will be done in Excel and the pareto diagram, a graph that arranges data in descending order, from left to right and where the data is presented by bars, will also be elaborated in this software. In this instance, the pareto diagram will be useful to know which municipality is primarily in charge of the consumption of electricity.

4.3.2 TABLEAU DESKTOP

Tableau is an interactive data visualization tool which makes simple to explore and manage data. Tableau is also available in the web and the visualizations created in Tableau can be shared with no charge or fee. Tableau gives a trial of one year to students, which allows the opportunity of understanding the software and enhanced data visualization abilities. The actual work was achieved thanks to the student trial.

In this thesis Tableau will be utilized to create scatterplot graphics and bar graphs that interact with each other with the proposed of a user-friendly visualization of electricity demand in Honduras.

In tableau, choropleth maps will be also created. A choropleth map is a specific kind of thematic map in which regions are shaded in varying shades, frequently from the same color range, to indicate various values of a statistical variable that is particular to that geographic area.

4.3.3 RSTUDIO

R is a programming language with a statistical approach, RStudio is a program which allows the user to handle R and use it more comfortably. The learning curve for R is relatively straight forward, that why is ideal for deliverables and rapid analysis. RStudio will provide the analysis for the multiple regression model; will assist with the discard of variables and improvement of the model.

4.4 POPULATION AND SAMPLE

This research will focus on the country of Honduras with a total of 298 municipalities. The sample to be utilized will be 259 municipalities for which the information was achievable, the rest of the municipal data will be obtained through the multiple linear regression model that will be explain in the future section.

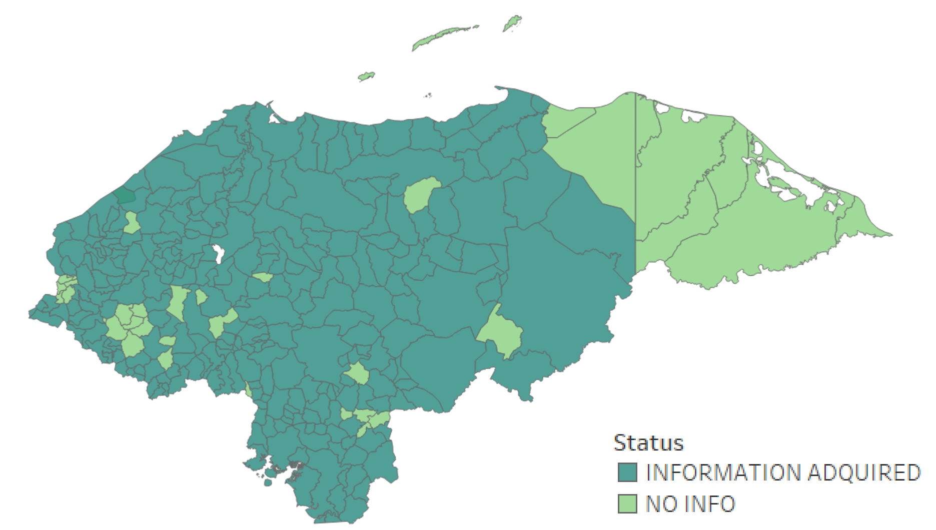


Figure 18: Status of information per municipality

Source: Own Elaboration

4.5 Research Methodology

The methodology considered most convenient to respond to the research questions raised in chapter two is explained below.

4.5.1 Compilation of sociodemographic and electricity consumption data at municipal level in Honduras

The first phase of data compilation started by contacting the local electrical distribution company (EEH) via email. The information requirements were detailed. For this thesis the goal was to achieve a municipality level of disaggregation in electricity consumption data.

The raw data was obtained by April 2022. In this data, certain cities and villages were taken as entire municipalities even though they are not. For example, Chamelecón is part of the municipality of San Pedro Sula; however, in the raw data it appears as being one entire municipality. So, a pre-treatment to correct these "errors" was made. In the end, it was possible to obtain the information of 259 municipalities (with their villages and cities grouped within) as the sample for electrical consumption data.

The second phase of data compilation was to identify governmental and legitimate data sources to obtain sociodemographic attributes for each of the 259 municipalities traced at first. For this purpose, three principal data sources were consulted: the National Institute of Statistics (INE), the Ministry of the Interior, Justice, and Decentralization (SGJD) and the National Electric Power Company (ENEE).

Table 3: Sociodemographic attributes compilated per municipality

Sociodemographic Attribute	Measure	Source
Electrical Coverage	Percentage [%]	ENEE 2019
Electrical Usage for cooking	Percentage [%]	INE 2013
Electrical Usage for lightning	Percentage [%]	INE 2013
Human Development Index	A dimensional [-]	SGJD 2020
Income per capita	Lempiras [L.]	SGJD 2020
Poverty Level	Percentage [%]	INE 2013
Urbanization Level	Percentage [%]	SGJD 2020
Total Population	Cases [inhabitants]	SGJD 2020
Ethnic Population	Percentage [%]	INE 2013
Industrialization Level	Cases [inhabitants]	INE 2013
Households with electrical coverage [cases]	Cases [houses]	INE 2013
Occupation per economical sector [cases]	Cases [inhabitants]	INE 2013

Source: Own elaboration

The variables above were selected as the main sociodemographic data for this investigation for each municipality.

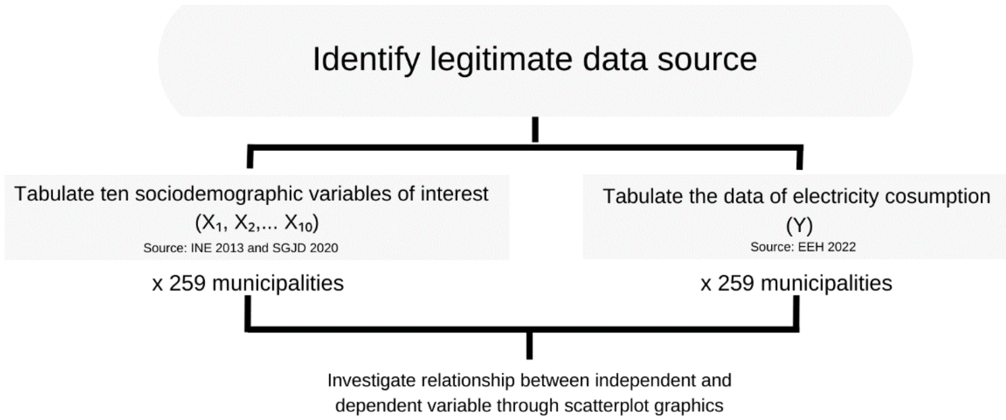


Figure 19: Followed steps for data compilation

Source: Own elaboration

The objective of the first eleven variables, including the electricity consumption and first ten sociodemographic parameters in Table 3, is to understand the behavior of electrical demand depending on sociodemographic data, also to identify if there is high or low determination coefficient between them (also known as R^2; this concept will be explained in section 4.5.2).

4.5.2 LINEAR REGRESSION MODEL

As shown in Chapter three, section 3.2.1 of this work, the methodologies of electricity demand forecast in Honduras are not accessible to academia, the forecast of demand is not considering location and the planification is centralized and does not provide space for autonomous municipal development and energy planning in Honduras.

Due to these obstacles, from the academia the actual study has as an objective to build a linear regression model based on sociodemographic variables as the independent variables and electrical consumption as a dependent variable. This model is created with the objective of understanding if these variables do affect electrical consumption.

R program is necessary to improve, approve or discard variables in the regression model. Following the requirements of statistical literature, the analysis of data in RStudio results in an ANOVA; an analysis of variance that constitutes a basic tool for studying the effect of one or more factors on the mean of a continuous variable.

For the Analysis of Variance (ANOVA) three parameters will be considered to discard independent variables or to change the regression model:

1. Adjusted coefficient of determination (R^2): This value demonstrates how much variability of the response the regression model based on the independent variables can predict; the adjusted value is utilized for linear multiple regression.

$$\mathbf{\bar{R}^2 = 1 - \frac{\frac{SSE}{m-(n+1)}}{\frac{SST}{m-1}}}$$

Equation 18: Definition of adjusted coefficient of determination

Source: (Richardson, 2015)

Where:

- m is the number of samples
- n is the number of independent variables
- SSE is the sum of squared estimate of errors
- SST is the total sum of squares

These two last parameters mentioned are calculated for the regression model, as follows:

$$SSE = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

Equation 19: Definition of Squared Estimate of Error

Source: (Richardson, 2015)

Where:

- $\hat{y}_i$ is the predicted value of the dependent variable

- y_i is the real value of the dependent variable
- m is the number of samples.

And;

$$SST = \sum_{i=1}^{m} (y_i - \bar{y})^2$$

Equation 20: Definition of Total Sum of Squares

Source: (Richardson, 2015)

Where:

- $\bar{y}$ is the mean of the dependent variable
- y_i is the real value of the dependent variable
- m is the number of samples

The accepted values of R^2 for certain model varies according to the literature but a range of **0.7 to 1 is acceptable** for the regression model to be determined in this thesis.

```
Call:
lm(formula = Y ~ X1 + X2, data = R)

Residuals:
      Min        1Q    Median        3Q       Max
-27522627   -579924    381753    792056  55325733

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.439e+06  4.084e+05  -3.523 0.000505 ***
X1           1.034e+02  3.047e+00  33.948  < 2e-16 ***
X2           1.810e+05  1.166e+06   0.155 0.876777
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4541000 on 256 degrees of freedom
Multiple R-squared:  0.8411,    Adjusted R-squared:  0.8398
F-statistic: 677.3 on 2 and 256 DF,  p-value: < 2.2e-16
```

Figure 20: Acceptable adjusted R^2 value in R

Source: Own Elaboration

2. Significant F / p-value for the F-test / Critical Value of F

   The accepted values of Critical Value of F for certain model varies according to the literature. For this methodology, when Critical Value of F is less than **0.05**, the overall model is significant, and when it is greater than **0.05**, the overall model is not significant.

```
Call:
lm(formula = Y ~ X3 + X4, data = R)

Residuals:
      Min        1Q    Median        3Q       Max
-13742314  -3156766   -823783   1630783 124000898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -59362069   15201491  -3.905 0.000121 ***
X3           90867763   24984709   3.637 0.000334 ***
X4               4878       3713   1.314 0.190064
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10470000 on 256 degrees of freedom
Multiple R-squared:  0.1558,    Adjusted R-squared:  0.1492
F-statistic: 23.62 on 2 and 256 DF,  p-value: 3.854e-10
```

Figure 21: Demonstration of significant models in R

Source: Own Elaboration

3. Slope coefficient and p-value for each independent variable

   If the slope coefficient called β in this investigation, that accompanies the independent variable is too small, variable can be discarded. It is also necessary to apply for the p-value of each coefficient the 0.05 rule. If p-value is less than **0.05**, the impact of the variable is significant, and when it is greater than **0.05**, the impact of the variable is not significant.

```
Call:
lm(formula = Y ~ X3 + X4, data = R)

Residuals:
      Min        1Q    Median        3Q       Max
-13742314  -3156766   -823783   1630783 124000898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -59362069   15201491  -3.905 0.000121 ***
X3           90867763   24984709   3.637 0.000334 ***
X4               4878       3713   1.314 0.190064
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10470000 on 256 degrees of freedom
Multiple R-squared:  0.1558,    Adjusted R-squared:  0.1492
F-statistic: 23.62 on 2 and 256 DF,  p-value: 3.854e-10
```

Figure 22: Evaluation of p-value per each independent variable

Source: Own Elaboration

In RStudio that is the compiler for R, the function "readxl ()" is necessary to upload the training set and the function "lm ()" is required to call for a multiple linear regression. Regressors (independent variables) that are not significant in the model can be identified and discarded.

In linear regression models in R, lm (Y~.) Considers the dependent variable and all the independent variables and lm (Y~X1+X2+X3) Considers the dependent variable and only the first three independent tabulated in the training set.

Even though, in this investigation, the model is required to be a multiple linear regression, R present the opportunity of exploring a specific case of quantile regression which will be also utilized.

This process of linear regression will be done for general consumption of electricity and will be done for the electricity consumption of the three main sectors: residential, commercial, and industrial. The code utilized for this linear regression model was written in R and is annexed to the document.

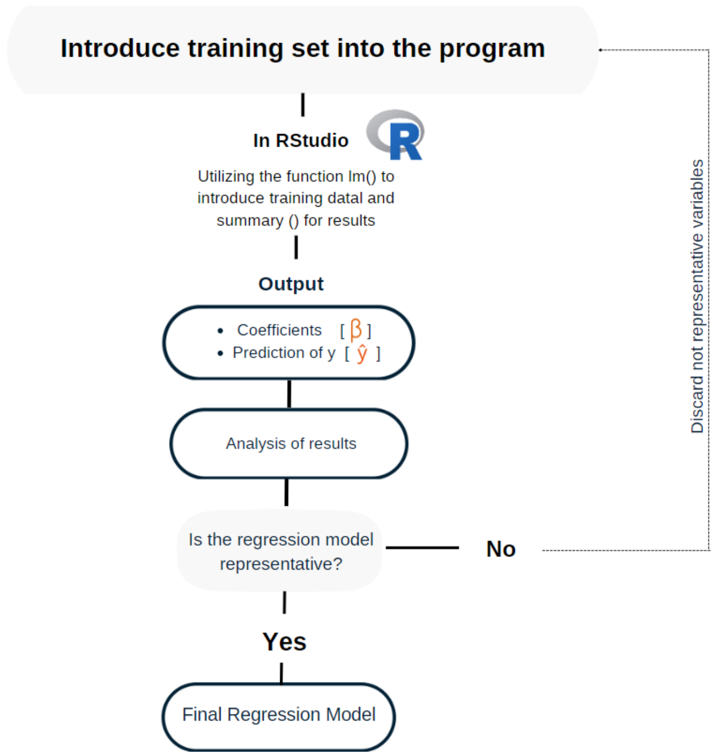


Figure 23: Linear Regression Model process

Source: Own Elaboration

4.5.3 DATA VISUALIZATION: MUNICIPAL MAP AND GRAPHS CREATION

It was mentioned in section 4.3 the utilization of Tableau Desktop; this tool is utilized with the purpose of providing a spatial distribution of electricity demand to be considered in national energy planning and making accessible the data of electricity consumption to general population, academia, and other interested sectors.

The information gathered in Excel will be transformed into a digestible format to reach a better understanding of the behavior of the electrical demand in Honduras. Thanks to the regression model already approved, it will be possible to obtain the electrical consumption of the 298 municipalities of Honduras whereas in raw data the only 259 municipalities were included.

Obtaining the national data, it is possible to create the subsequent visualizations in Tableau per department (18) and municipality (298) in Honduras:

a) Choropleth map of electricity consumption [MWh]

A choropleth map is a geographical representation where areas are differentiated by colors of the same chromatic range depending on certain variable, in this case the variables is electricity consumption.

Input:

- Shapefile of the map with the municipal division of Honduras obtained via web in Gobierno de la República de Honduras (2019).
- Electrical consumption per municipality data [MWh].

Process:

- Shapefile of municipal division of Honduras' map Geometry is charged into a Tableau´s sheet.
- Electrical consumption per municipality data is uploaded and linked with the shapefile document through municipal code (an array of four characters unique per each municipality). This action is made by the "Blend Relationship" option in Tableau.
- Municipality name and Department name is assigned to Tableau´s sheet as a "Detail" Mark.
- Electrical consumption per municipality data [MWh] is assigned to Tableau´s sheet as "Color" Mark.
- Electrical consumption per municipality data [MWh] is assigned to Tableau´s sheet as "Tooltip" Mark.
- Choose Department name as Filter in Tableau´s sheet.

Output:

- Honduras choropleth municipal map colored by electricity consumption [MWh] with interactive design filtered by department and municipality. In this case, the department shown is Choluteca because is the department from which all the municipal information was held.

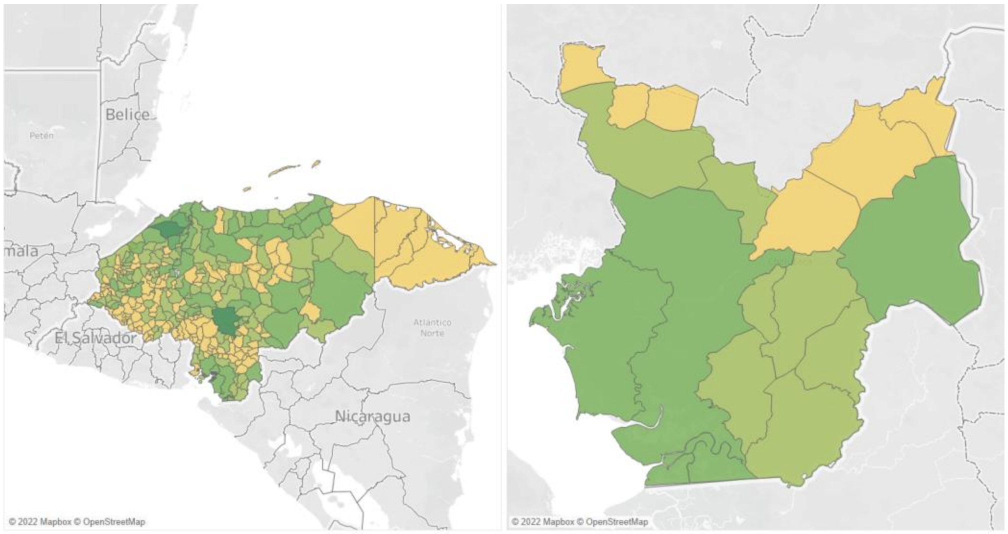


Figure 24: Preview of Honduras map [left] and the department of Choluteca [right] by electricity consumption of March 2022 [MWh]

Source: Own Elaboration

b) *Interactive design of Choropleth map + Bar graph of electricity consumption specified by sector [MWh/sector]*

Having already discussed the elaboration of the Honduras choropleth municipal map colored by electricity consumption [MWh] with interactive design filtered by department. The bar graph elaboration will be discussed.

Input:

- Electrical consumption by sector per municipality [MWh /sector]

Process:

- Electrical consumption by sector per municipality data is uploaded and linked with the shapefile and global electricity consumption [MWh] document through the municipal code (an array of four characters unique per each municipality).
- In "Show Me" section the graph of interest is chosen. In this case Bar Graph.
- Name of Sector is assigned to Columns and Electrical consumption by sector per municipality is assigned in Rows. Bar Graph will be completed.

- Create a new dashboard; assigned both sheets (Choropleth map and Bar graph) to the new dashboard and use a filter in the Choropleth map by municipality.

Output:

- Choropleth map by municipality and electricity consumption classified by sector within.

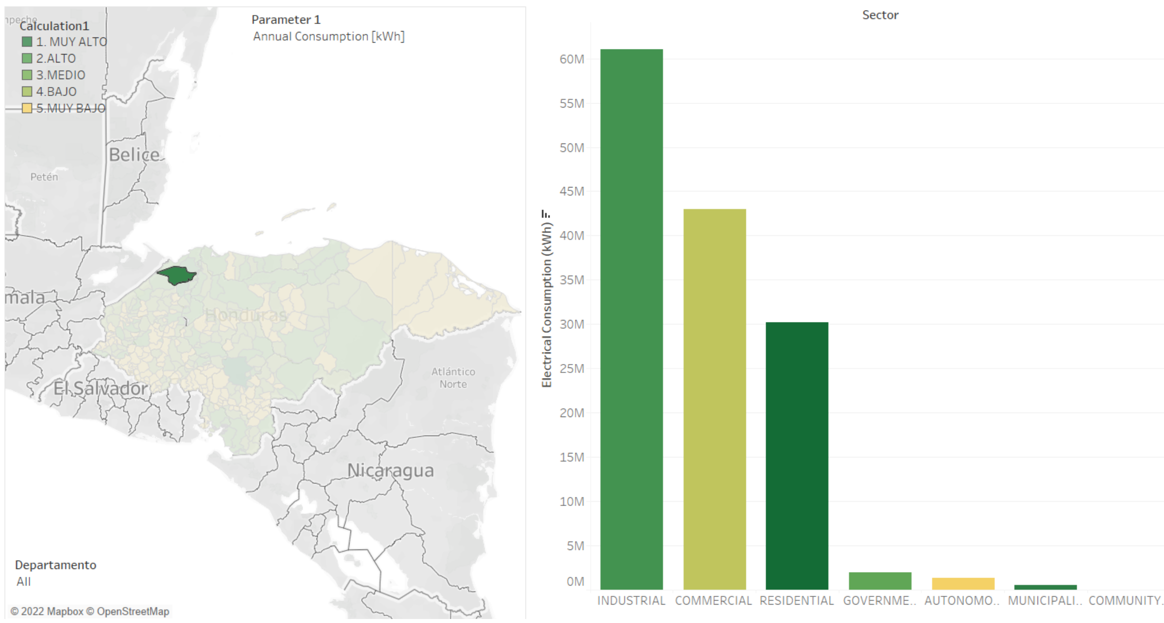


Figure 25: Preview of electricity consumption by sector for certain municipality selected [kWh]

Source: Own Elaboration

c) Choropleth map of global electricity consumption per capita [kWh/hab]:

Input:

- Shapefile of the map with the municipal division of Honduras obtained via web in Gobierno de la República de Honduras (2019).
- Electrical consumption per municipality data [kWh]
- Population of each municipality by 2021 [inhabitants]

Process:

- Obtain global electricity consumption per capita (ECPC) by municipality following the next equation:

$$\text{ECPC} = \frac{\text{Electrical consumption per municipality [kWh]}}{\text{Population 2021 per municipality [hab]}}$$

Equation 21: Definition of electricity consumption per capita

Source: Own Deduction

- Shapefile of municipal division of Honduras' map Geometry is charged into a Tableau´s sheet.
- Electricity consumption per capita data is uploaded and linked with all data utilized in the model. This action is made by the "Blend Relationship" option in Tableau.
- Municipality name and Department name is assigned to Tableau´s sheet as a "Detail" Mark.
- Electricity consumption per capita per municipality data [kWh/hab] is assigned to Tableau´s sheet as "Color" Mark.
- Electricity consumption per capita per municipality data [kWh/hab] is assigned to Tableau´s sheet as "Tooltip" Mark.
- Choose Department Name as Filter in Tableau´s sheet.

  Output:
- Honduras choropleth municipal map colored by electricity consumption per capita per municipality data [kWh/hab] with interactive design filtered by department.

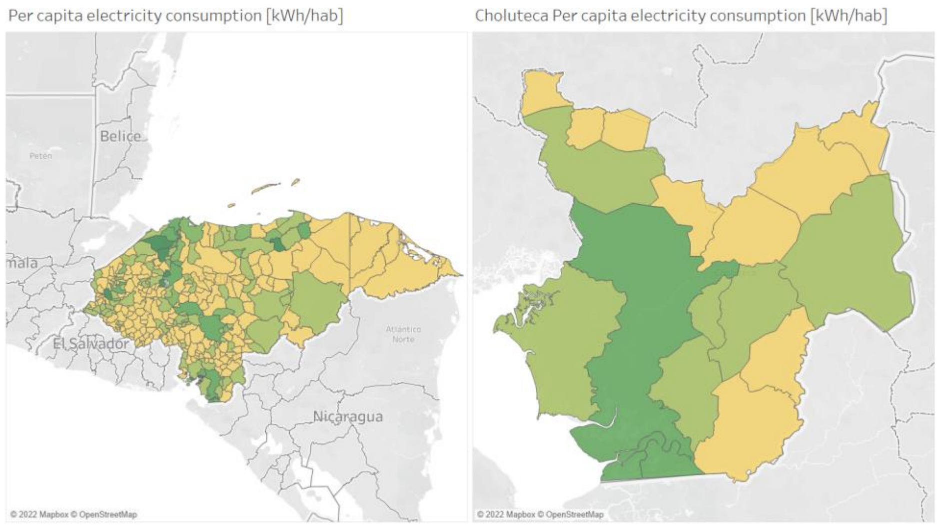


Figure 26: Preview of Honduras map [left] and the department of Choluteca [right] by electricity consumption per capita of March 2022 [kWh/inhabitant]

Source: Own Elaboration

d) Choropleth map of adjusted electricity consumption per capita [kWh/EEH subscriber]

Input:

- Shapefile of the map with the municipal division of Honduras obtained via web in Gobierno de la República de Honduras (2019).
- Electrical consumption per municipality data [kWh]
- Population of each municipality by 2021 [inhabitants]
- Electrical Coverage [%]

Process:

- Obtain electricity consumption per EEH subscriber (ECPS) by municipality following the next equation:

$$\text{ECPS} = \frac{\text{Electrical consumption per municipality [kWh]}}{\text{Population 2021 per municipality [hab]} * \text{Electrical coverage [\%]}}$$

Equation 22: Definition of electricity consumption per EEH subscriber

Source: Own Elaboration

- Shapefile of municipal division of Honduras' map Geometry is charged into a Tableau´s sheet.
- Electricity consumption per EEH subscriber data is uploaded and linked with all data utilized in the model. This action is made by the "Blend Relationship" option in Tableau.
- Municipality name and Department name is assigned to Tableau´s sheet as a "Detail" Mark.
- Electricity consumption per EEH subscriber per municipality data [kWh/sub] is assigned to Tableau´s sheet as "Color" Mark.
- Electricity consumption per EEH subscriber per municipality data [kWh/sub] is assigned to Tableau´s sheet as "Tooltip" Mark.
- Choose Department Name as Filter in Tableau´s sheet.

Output:

- Honduras choropleth municipal map colored by electricity consumption per EEH subscriber per municipality data [kWh/sub] with interactive design filtered by department.

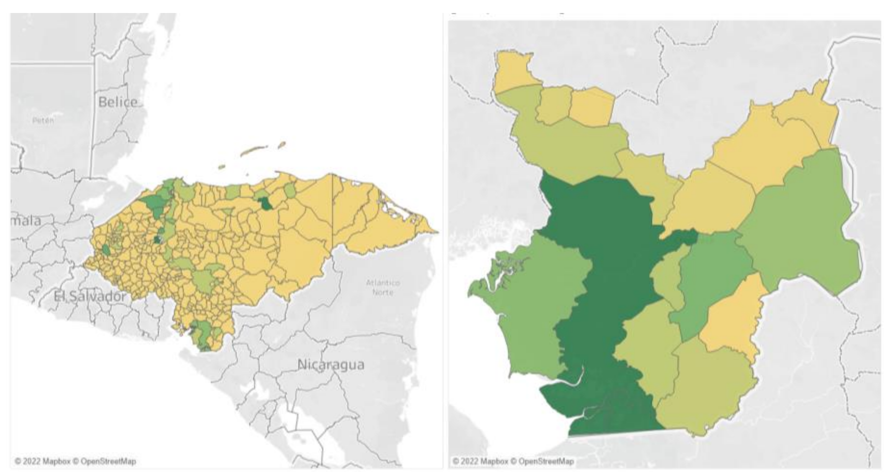


Figure 27:Preview of Honduras map [left] and the department of Choluteca [right] by electricity consumption per EEH subscriber of March 2022 [kWh/subscriber]

Source: Own Elaboration

e) Choropleth map of density of electricity consumption[kWh/m2]:

Input:

- Shapefile of the map with the municipal division of Honduras obtained via web in Gobierno de la República de Honduras (2019).
- Electrical consumption per municipality data [kWh]
- Superficial area of each municipality [m^2]

Process:

- Obtain electricity density (ED) by municipality following the next equation:

$$\mathrm{ED} = \frac{\text{Electrical consumption per municipality [kWh]}}{\text{Superficial area of each municipality } [\mathrm{m}^2]}$$

Equation 23: Definition of electricity density

Source: Own Elaboration

- Shapefile of municipal division of Honduras' map Geometry is charged into a Tableau´s sheet.
- Electricity density data is uploaded and linked with all data utilized in the model. This action is made by the blend relationship option in Tableau.
- Municipality name and Department name is assigned to Tableau´s sheet as a "Detail" Mark.
- Electricity density per municipality data [kWh/m^2] is assigned to Tableau´s sheet as "Color" Mark.
- Electricity density per municipality data [kWh/m^2] is assigned to Tableau´s sheet as "Tooltip" Mark.
- Choose Department Name as Filter in Tableau´s sheet.

Output:

- Honduras choropleth municipal map colored by electricity density per municipality data [kWh/m^2] with interactive design filtered by department

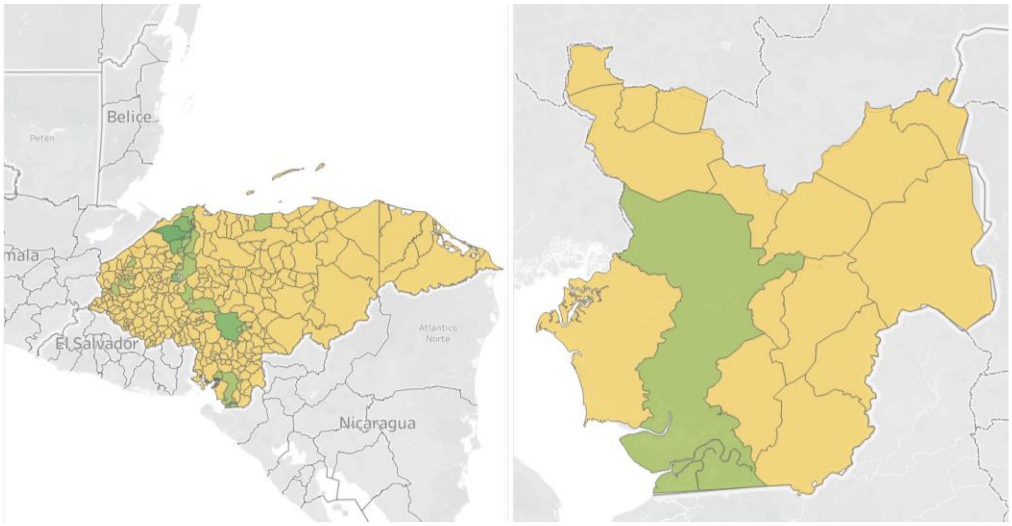


Figure 28: Preview of Honduras map [left] and the department of Choluteca [right] by electricity density of March 2022 [kWh/m²]

Source: Own Elaboration

f) Choropleth map of electricity consumption per capita specified by sector [kWh/hab]:

Input:

- Shapefile of the map with the municipal division of Honduras obtained via web in Gobierno de la República de Honduras (2019).
- Electrical consumption per municipality data [kWh]
- Electrical consumption per sector data [kWh]
- Population of each municipality by 2021 [inhabitants]

Process:

- Obtain sectorial electricity consumption per capita (SEPC) for each municipality following the next equation:

$$SECPC = \frac{Electricity\ consumption_{mn}}{Population_m}$$

Equation 24: Calculation of sectorial electricity consumption per capita

Source: Own deduction

Where:

- m represents the municipality
- n represents the sector, which can be:
    1. Residential
    2. Commercial
    3. Industrial: divided by ENEE into moderate and large industries.
    4. Autonomous
    5. Government
    6. Municipal
    7. Community service

This equation will be utilized seven times by each municipality, to calculate the seven parameters of SECPC.

- Shapefile of municipal division of Honduras' map Geometry is charged into a Tableau´s sheet.
- Sectorial electricity consumption per capita data is uploaded and linked with all data utilized in the model. This action is made by the blend relationship option in Tableau.
- Municipality name and Department name is assigned to Tableau´s sheet as a "Detail" Mark.
- A "Parameter "named "List of Sectors" must be created in Tableau where the seven sectors are shown as a drop box list for which the user can decide which choropleth map visualize.
- A "Calculation Field" named "Sector" must be created in Tableau where a conditional code must be written to connect the drop box list created as a parameter with the excel worksheet.
- The Calculation Field named "Sector" must be placed to Tableau´s sheet as "Color" Mark.
- The Calculation Field and the Parameter named "Sector" and "List of Sectors" respectively are assigned to Tableau´s sheet as "Tooltip" Mark.

- Choose Department Name as Filter in Tableau´s sheet.

Output:

Honduras choropleth municipal map colored by sectorial electricity consumption per capita per municipality data [kWh/hab] with interactive design filtered by department.

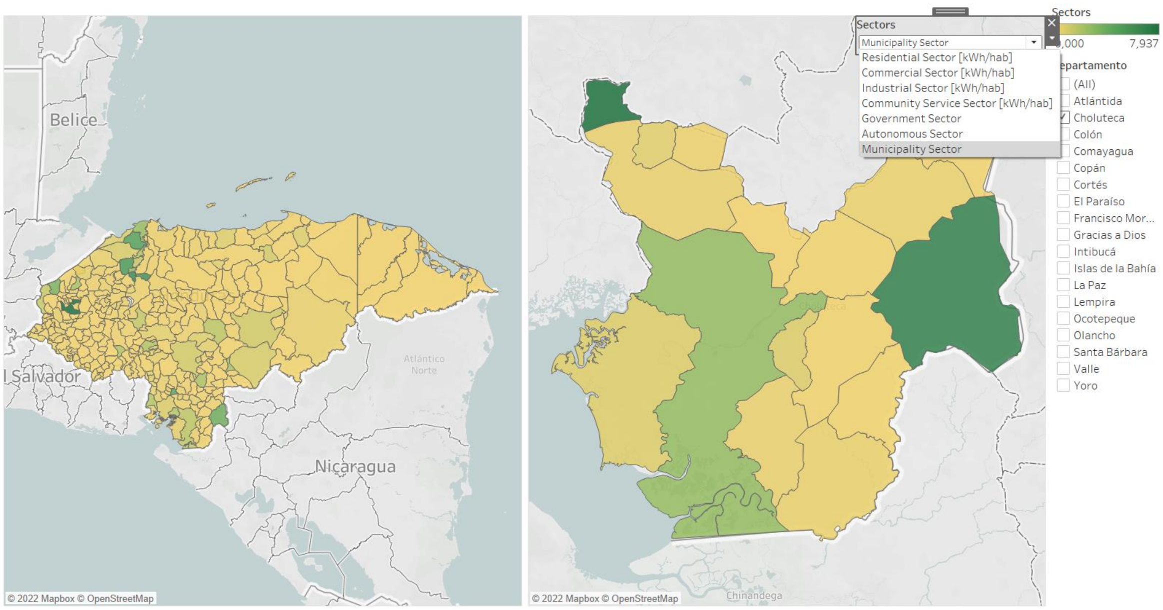


Figure 29: Preview of Honduras map [left] and the department of Choluteca [right] colored by sectorial electricity consumption of March 2022 [kWh/inhabitant/sector]

Source: Own Elaboration

g) Scatterplot of HDI [-] vs Electrical Consumption [GWh]:

Logarithmic scale will be utilized for the x- axis, through this a better approach to analyze the data will be achieve, dividing the municipalities by clusters where similar characteristics for both variables are achieved.

4.6 SCHEDULE

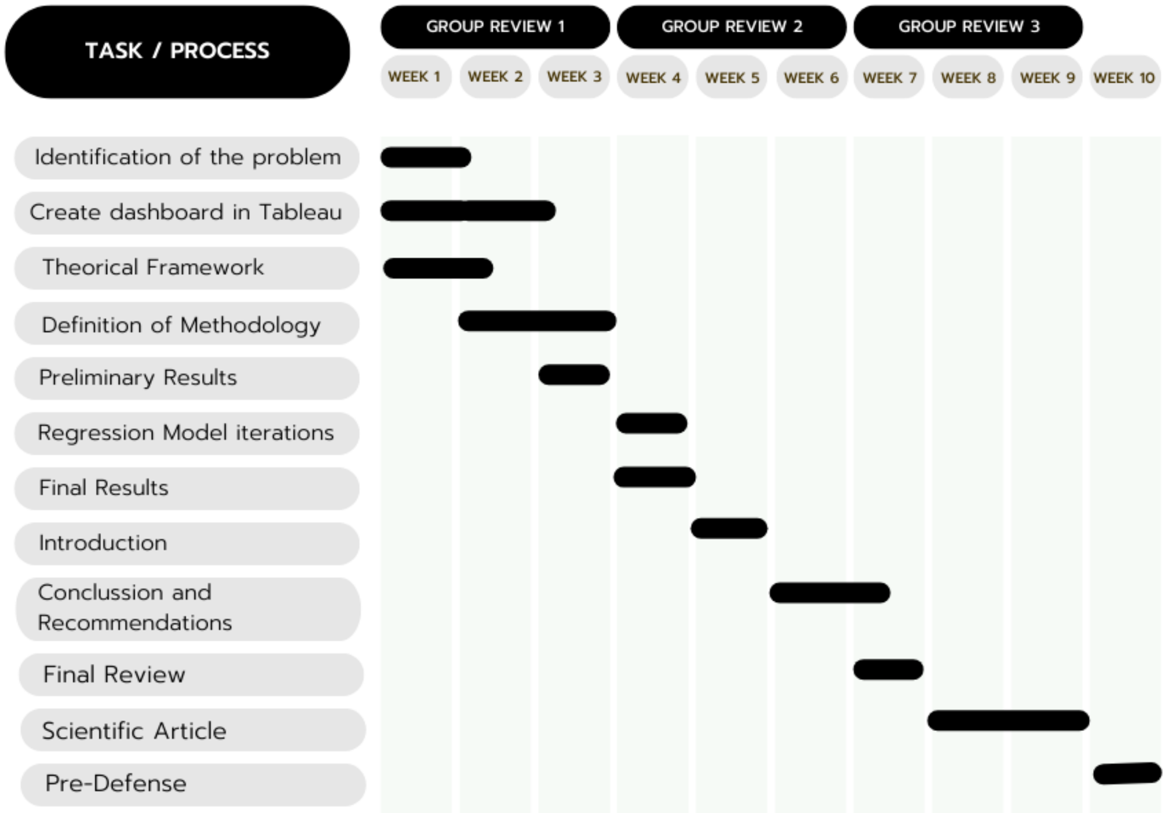


Figure 30: Schedule of thesis drafting

Source: Own Elaboration

V. RESULTS

In this chapter, the results of the investigation methodology will be exposed and analyzed through instruments that illustrate better the behavior of data. This section provides the necessary information to arrive to rigorous conclusions about the electricity consumption in Honduras.

The results will be divided into five sections,

1) The first section is a global analysis of consumption by municipality. This section is where through Pareto diagrams the municipalities that are the largest consumers of total electricity consumption, per capita electricity consumption and electricity density in the national territory will be identified.
2) The second section contains an elaborated analysis to indicate the correlation of electricity consumption with 10 sociodemographic and socioeconomic variables through the usage of scatterplot graphs. The purpose of this section is finding the highest coefficient of determination (R^2) trying different functions by which the relation of the two variables (one of the 10 social variable and electricity consumption) can be best explained.
3) The third section contains the explanation of the regression model analysis and its iterations in the search for the best fitted multiple linear regression.
4) The fourth section contains the results of the data visualization tool in Tableau and a departmental analysis of the electrical demand in Honduras showing the percentages that the municipalities represent of the total departmental consumption. Also, the total electricity consumption, per capita electricity consumption and electricity density will be calculated for each department as it was done in the first section per municipality.
5) The fifth section involves a sectorial analysis. This sectorial analysis is made to understand which sector consumes more electricity (therefore, who should be the first to implement energy efficiency policies) at a national, departmental, and municipal level. Finally, in the residential sector, an economic analysis per household

in each municipality will be made to identify how much of the income in households is spent on electricity services.

5.1 GLOBAL ANALISYS OF ELECTRICITY CONSUMPTION

Having already recollected all the information necessary of the 259 municipalities for the sample, the analysis of the monthly electricity consumption for March 2022 (as the model month) can be done.

There is one main conclusion reached by analyzing the data: the consumption of electricity in Honduras is unequally distributed. Only 18 municipalities represent 80% of the electricity consumption for the month of March 2022. (In other words, only 6.94% of the municipalities represents 80% of electricity consumption).

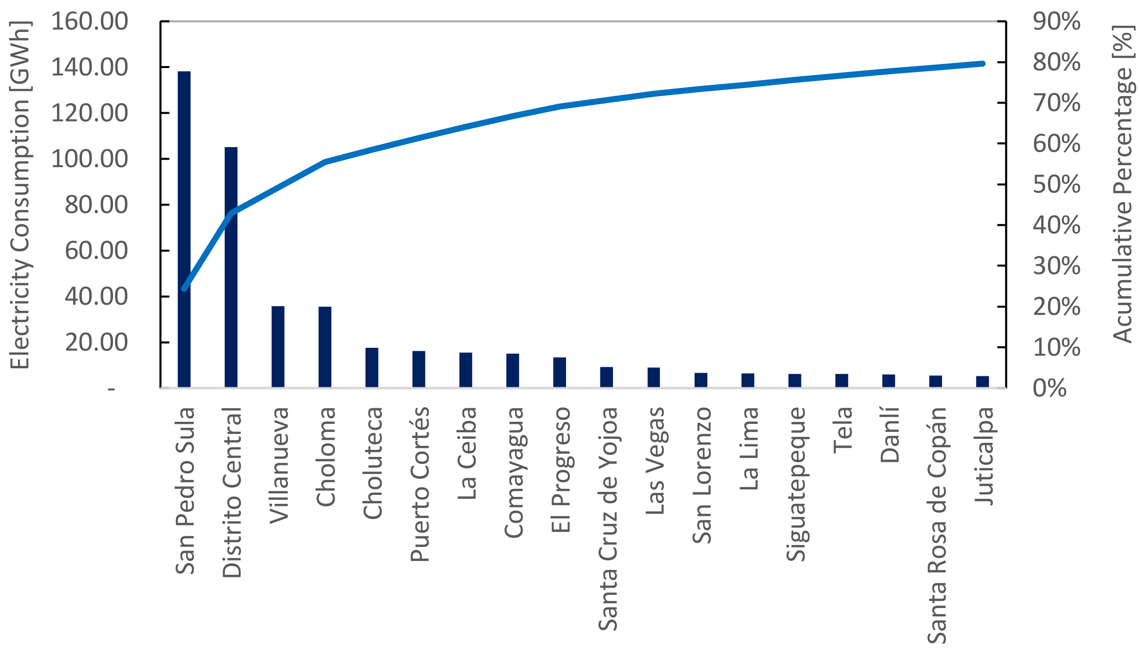


Figure 31: Pareto Diagram of the 18 municipalities that are the main consumers of electricity in Honduras

Source: Own Elaboration with information obtained via email from (EEH , 2022)

The analysis done by the pareto diagram can be utilized also for other percentages besides 80%. For instance, 90% represents only the consumption of 40 municipalities out of 259 that are being studied. This means that 219 municipalities are responsible for only 10% which is insignificant and quantitatively reveals the levels of inequality in electricity consumption.

This analysis by municipalities might not be as revealing. To analyze the population involved is more representative of the actual situation in Honduras. In the 19 municipalities where accumulative electricity consumption represents 80%, the population is four million of Hondurans. The rest, approximately, 5.5 million consume only 20% of the electricity resource.

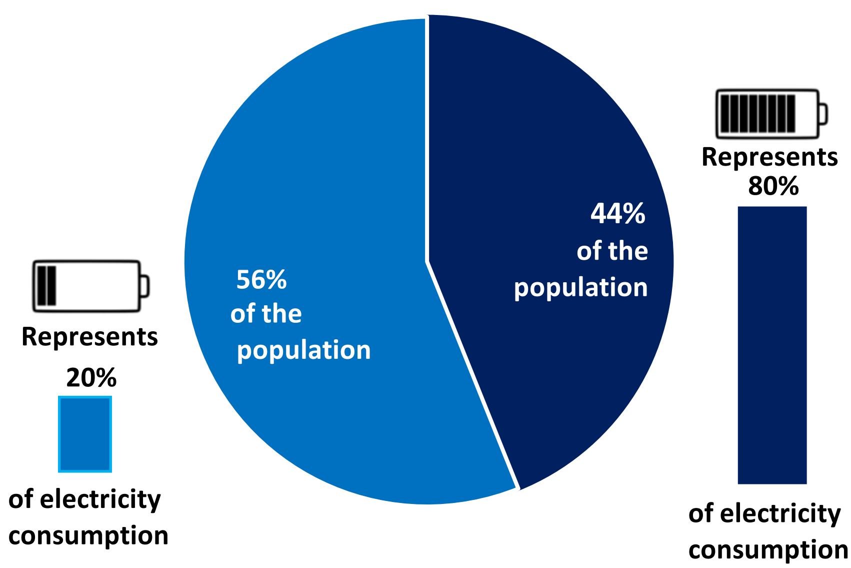


Figure 32: Inequality in electricity consumption presented by population

Source: Own Elaboration with information obtained via email from (EEH , 2022)

Given that electricity is a variable that has and effect in human development, an unequal distribution of this good could translate into a not advanced human development.

Another way of understanding electricity consumption is analyzing the per capita electricity consumption, this will provide a raw idea of the municipalities where the population has more electrical security. The per capita electricity consumption will also provide a variable of comparison given the normalization of electricity consumption being divided by the population of the municipality as shown in equation 18.

In an ideal scenario, where no disparity in electricity consumption distribution exists, all municipalities should have the same value of per capita electricity consumption signifying that all Hondurans consume the same quantity of electricity.

The value of electricity consumption per capita for each municipality can be found in Annex 1.

The graph below shows the Top 15 municipalities that stand out with levels of electricity consumption per capita between 350 and 100 kWh consumed per inhabitant in the municipal territory.

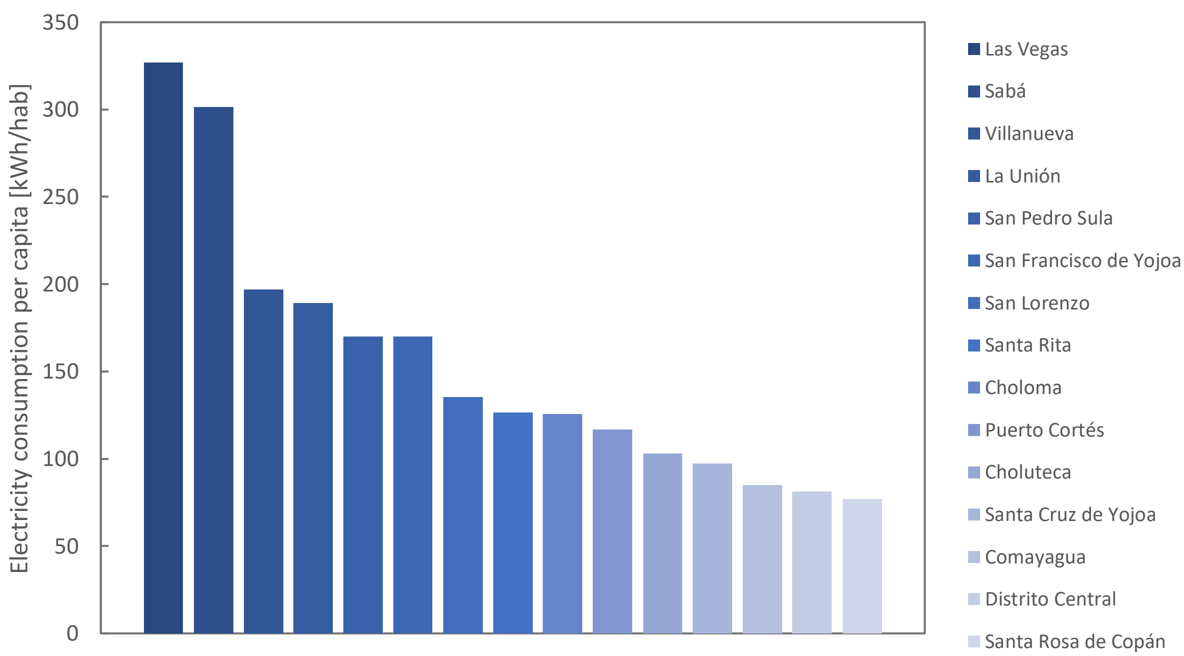


Figure 33: Top 15 municipalities with highest electricity consumption per capita

Source: Own Elaboration with information obtained via email from (EEH , 2022)

Las Vegas, Sabá and La Unión represent greater values of electricity consumption per capita than San Pedro Sula, being the last one known as the most industrialized municipality in the country and one of the most urbanized territories. This occurrence has an explanation: population in the first three municipalities mentioned is less than in San Pedro Sula. This results in a smaller value as a denominator in the equation 18 therefore a bigger value as a response.

Individually these municipalities were identified to present the subsequent scenarios that could also give response to the occurrence:

- Las Vegas is known as the mining capital of Honduras; the consumption of electricity is mainly just in the industrial sector. The quantity of inhabitants is not representative of the electricity consumption.
- Sabá is where the Standard Fruit company is located, it is commonly known by the locals that this company is connected to the electricity network as a residential client. The quantity of inhabitants is not representative of the electricity consumption.
- La Unión is home of the mines of San Andrés, representing that this mining activity is the responsible of most of the electricity consumption in the municipality. The quantity of inhabitants is not representative of the electricity consumption.

It is also important to mention that Distrito Central, where the capital city of Honduras is located, it is found in 14th place, revealing that in the Top 15 only Santa Rosa de Copán has a smaller value of electricity consumption per capita.

Whereas in this Top 15 municipalities the ranges go from 350 to 100 kWh per capita, in the Bottom 15, ranges go from 5 to 0 kWh per capita.

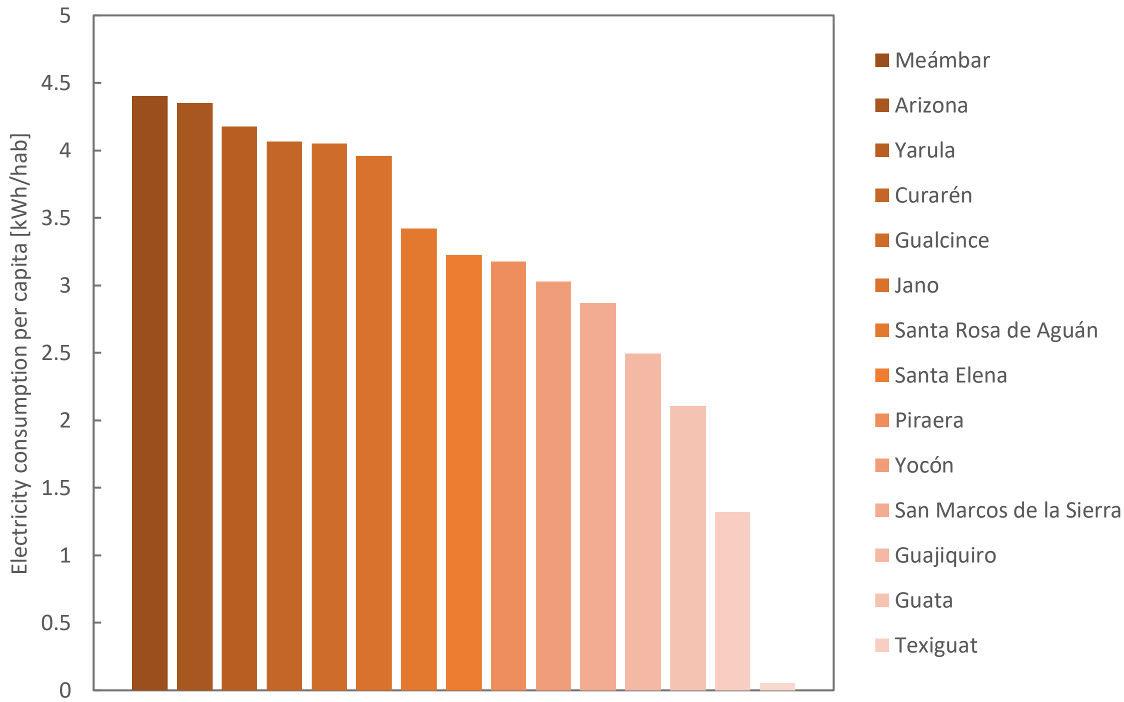


Figure 34: Bottom 15 municipalities with lowest electricity consumption per capita

Source: Own Elaboration with information obtained via email from (EEH , 2022)

Pitifully, in perspective, a consumption of 4.5 kWh per capita in Meámbar (Bottom 15) and a consumption of approximately 170 kWh per capita in San Pedro Sula (Top 15) lays bare that a person in San Pedro Sula consumes 40 times more electricity that in Meámbar. This fact demonstrates once again the inequality of electricity consumption in the national territory.

Normalizing electricity consumption dividing it by the population of the municipality is a great way of getting a value that is comparable from municipality to municipality. Another normalized parameter will be calculated and analyzed for this investigation.

Electricity Density or electricity consumption per area is another normalized form of electricity consumption. How to obtain the parameter is described in Equation 21. In an ideal scenario, where no disparity in electricity consumption distribution exists, all municipalities should have the same value of electricity density. The value of electricity density for each municipality can be found in Annex 1.

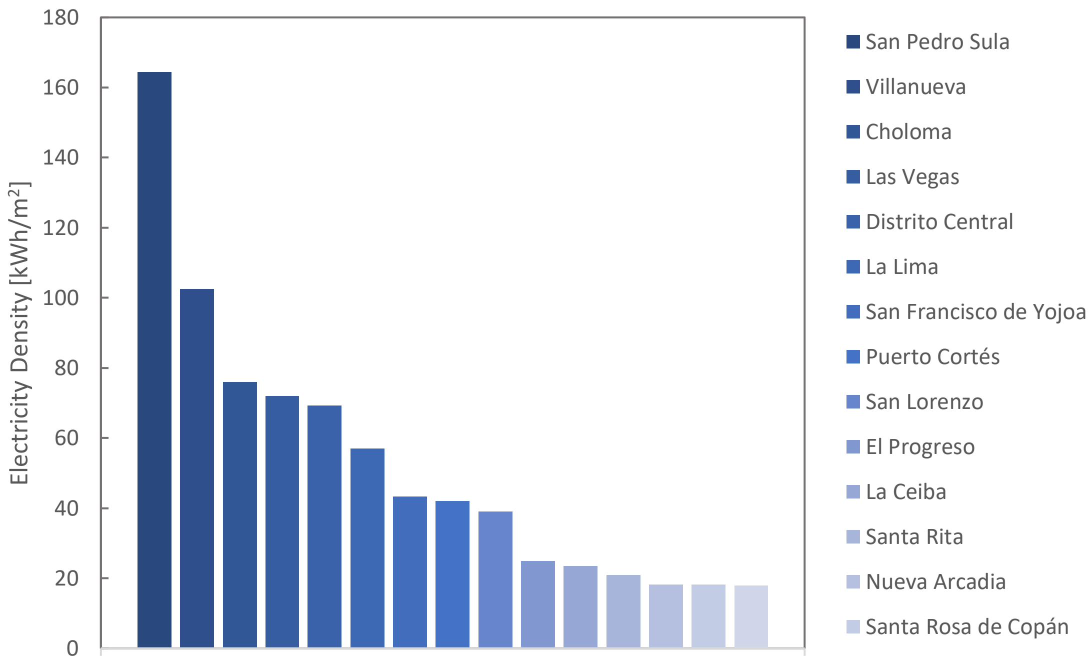


Figure 35: Top 15 municipalities with highest electricity density in Honduras

Source: Own Elaboration with information obtained via email from (EEH , 2022)

In the Top 15, San Pedro Sula is identified as the municipality with the highest consumption per square meter. San Pedro Sula is followed by Villanueva, Choloma and Las Vegas, whose economical focus is the industrial sector. Distrito Central stands out in the 5th place. In the contrary of most municipalities, Santa Rosa de Copán offers the same behavior that in electricity consumption per capita, it maintains its place in 15th place.

While in the Top 15 the range take in values from 160 to 20 kWh per square meter, in the Bottom 15 the values go from 0.2 to 0 kWh per square meter.

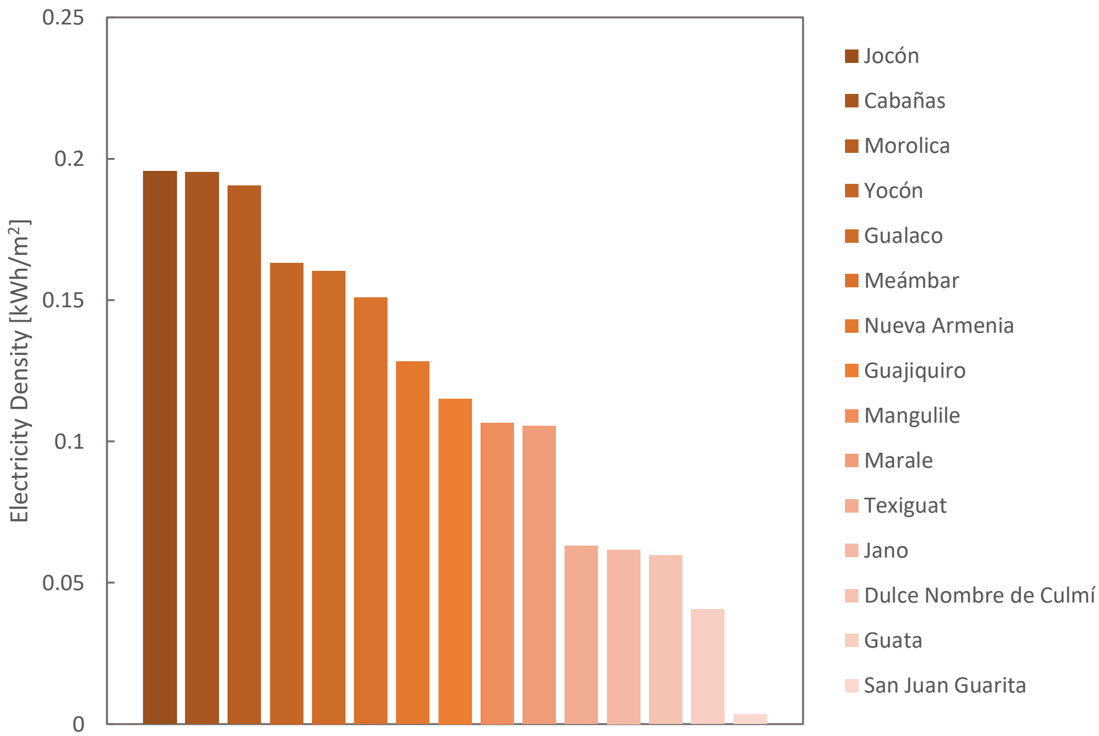


Figure 36: Bottom 15 municipalities with lowest electricity density

Source: Own Elaboration with information obtained via email from (EEH , 2022)

Including both parameters (electricity consumption per capita [kWh/hab] and electricity consumption per area [kWh/m^2]) to identify which municipalities have similar characteristics of electricity consumption can be done through a scatterplot. In the graph below, the y-axis is the electricity consumption per capita, and the x-axis is the electricity density parameter. A Logarithmic Scale is used for both axes to separate better the municipalities.

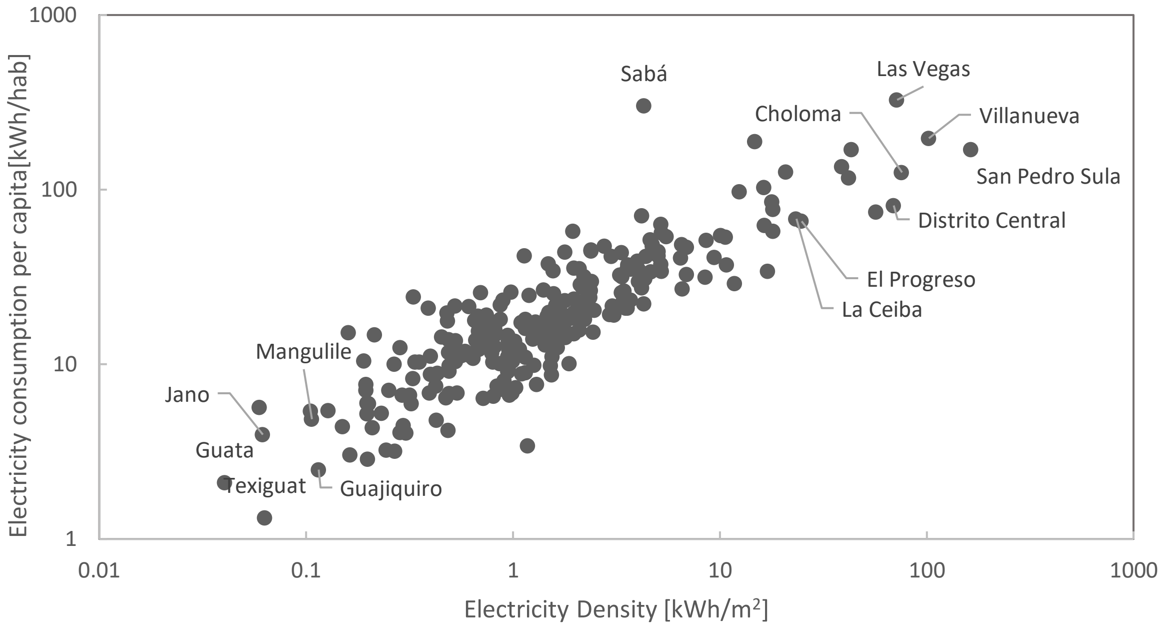


Figure 37: Relationship between electricity consumption per capita and electricity density

Source: Own Elaboration with information obtained via email from (EEH , 2022)

The municipalities are represented by dots in Figure 39, the dots that are left oriented are the municipalities that have minor values of electricity consumption per capita and electricity consumption per square meter such as Texiguat, Guajiquiro, Jano and Mangulile there are various characteristics that these places have in common such as low urbanization level and per capita income.

The dots that are right oriented are the municipalities that have mayor values of electricity consumption per capita and electricity consumption per square meter such as Villanueva, San Pedro Sula, Distrito Central and Las Vegas.

Other municipalities that also have great values of these two parameters are El Progreso and La Ceiba. But there is another interesting fact about these two municipalities: geographically they are ubicated in distinct departments however they have the same characteristics of electricity consumption.

With this example in mind, a conclusion can be reached: electricity consumption behavior can be the same in municipalities that are away from each other and is not necessarily the same value for municipalities in the same department or near each other, that is why disaggregation is necessary.

5.2 INDIVIDUAL CORRELATION ANALYSIS BETWEEN SOCIODEMOGRAPHIC VARIABLES AND ELECTRICITY CONSUMPTION

It is important to understand the sociodemographic characteristics of each municipality to have a wider perspective to analyze electricity consumption data.

5.2.1 POPULATION [INHABITANTS]

In relation to inhabitants, electricity consumption data presents a linear relationship with respect to the variable in the x-axis. Being the coefficient of determination (R^2) greater than 0.7 this relationship is considered significant for the thesis.

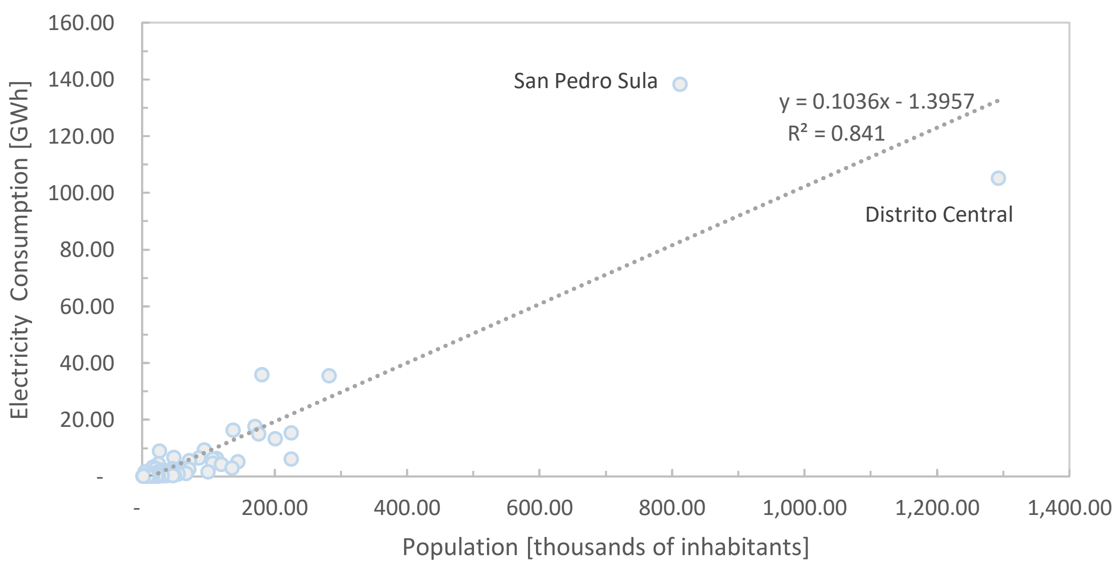


Figure 38: Relationship between electricity consumption and population in municipalities of Honduras

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

In the search of a greater coefficient of determination (R^2) an experiment utilizing only the residential electricity consumption for each municipality but the same number of inhabitants as the graph above was done.

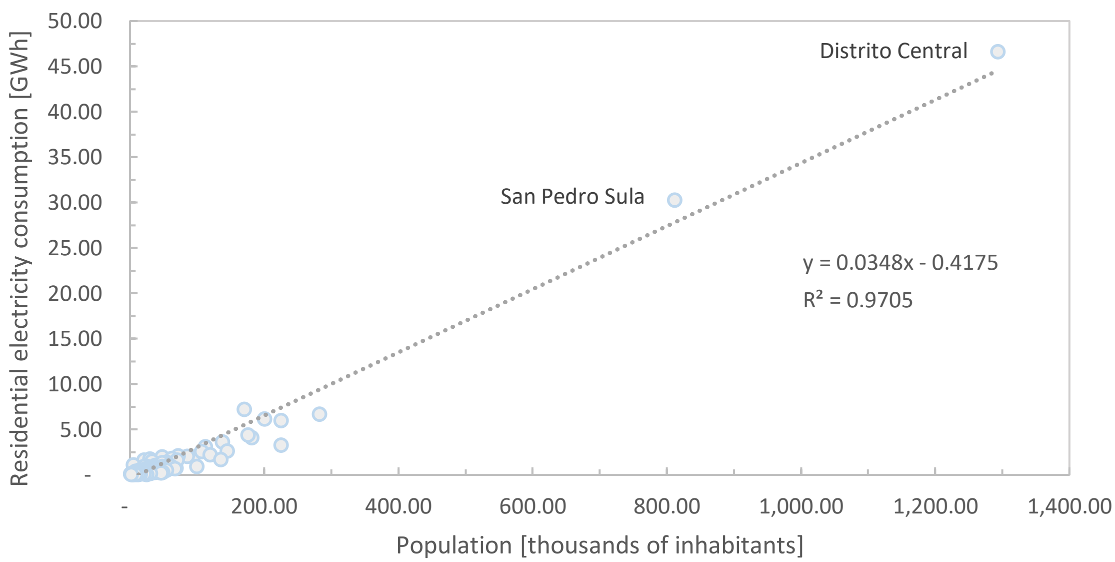


Figure 39: Relationship between residential electricity consumption and population in municipalities of Honduras

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

The graph above determines that population has a greater coefficient of determination in relation to residential sector consumption of electricity where the value for R^2 is 0.97. This means that population has a higher influence in residential consumption than overall consumption in the municipalities of Honduras.

5.2.2 HUMAN DEVELOPMENT INDEX [HDI]

It was already understood in Chapter two that human development seems to be affected positively by electricity consumption for many countries in the world. That is why is necessary to understand how the relation is between HDI and electricity consumption in Honduras. The best fitted function for these two variables is a logarithmic function even though the R^2 coefficient is less than 0.7 the relationship is significant.

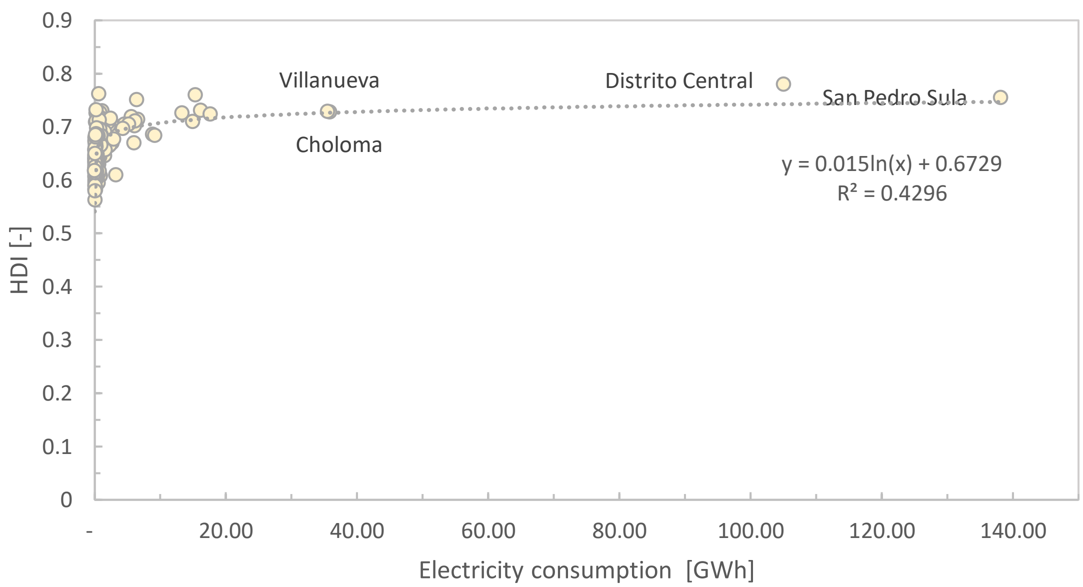


Figure 40: Relationship between HDI and Electricity consumption

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

In Figure 42, only four municipalities were distinguished: San Pedro Sula, Distrito Central, Villanueva and Choloma. To achieve a better understanding of the municipalities that are alike based on these two parameters, a logarithmic scale was applied to the x-axis, allowing the opportunity of creating clusters.

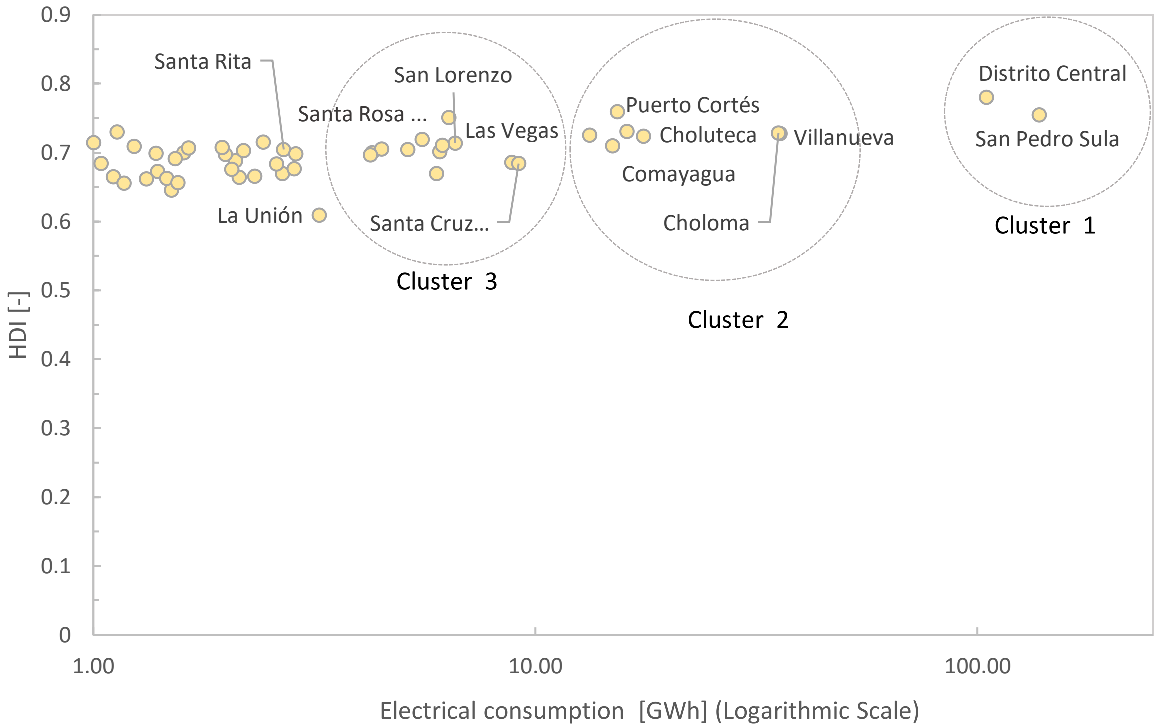


Figure 41: Formation of clusters by similar parameters of HDI and electricity consumption

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

In Figure 43, clusters can be identified:

- Cluster 1 contains the main cities of Honduras, San Pedro Sula which is the industrial capital and Tegucigalpa in Distrito Central which is the capital city of Honduras.
- Cluster 2 represents municipalities with manufacturing industrial activity such as Villanueva and Choloma or municipalities such as Comayagua and Puerto Cortés where commerce is highly representative of the economy.
- Cluster 3 includes San Lorenzo and Santa Rosa de Copán that are touristic attractions in Honduras. Also includes Las Vegas, municipality that we are discussed as the mining capital.

Concluding that industrial and touristic focused municipalities in Honduras have greater HDI and greater electricity consumption.

Figure 4 showed the behavior of 30 countries with respect to their respective Human Development Index. The same behavior is demonstrated with the 259 municipalities of Honduras with respect to their HDI values.

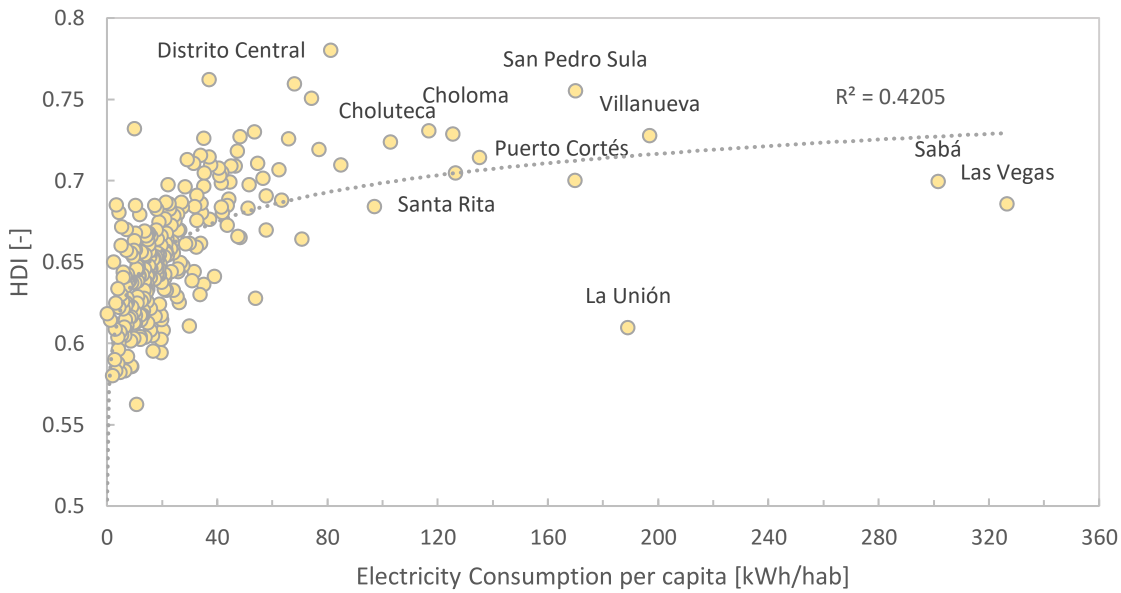


Figure 42: Relationship between HDI and Electricity Consumption per capita

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

For greater values than 100 kWh/hab, the HDI axis does not represent values less than 0.7. Except for the municipalities of Sabá, Las Vegas and La Unión, which present low Human Development Index values when electricity consumption is high. These three municipalities were already studied in Section 5.1.1. to understand this arbitrary behavior.

5.2.3 PER CAPITA INCOME [L.]

There is no function that describes the behavior of electricity consumption based on per capita income, given that for a value of x a range of y-values are responses. In the contrary, in a linear/ logarithmic function for a value of x just one value of y should be a valid response.

Even though a low value of R^2 of 0.3361 is the best fitted possible function, there is a qualitatively explainable relationship between these parameters. The higher the per capita electricity consumption, the wider the range in which electricity consumption can vary. For instance, when the per capita income is L.1,000 then electricity consumption can vary from 0 to 350 kWh/hab. While the per capita income is L. 200 the electricity consumption can only vary from 0 to 50 kWh.

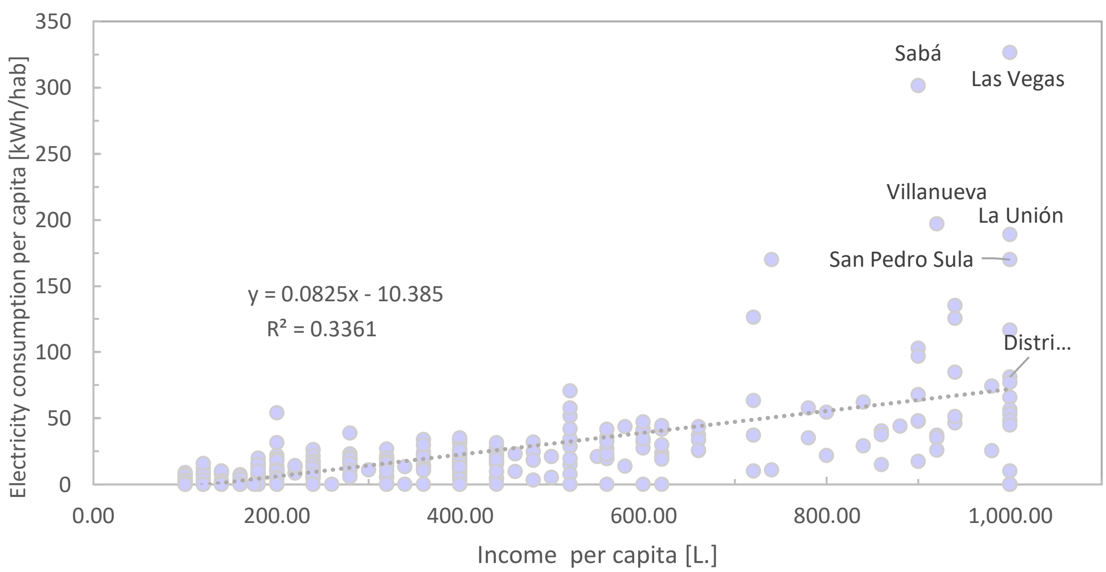


Figure 43: Relationship between Electricity Consumption per capita and Income per capita

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

The more money a municipality is willing to give to their population, the better chances that this population consumes more electricity per capita.

5.2.4 Urbanization Level [%]

Urbanization level dictates how many households are in the urban area of a municipality. The trend for urbanization is similar to the graph above (per capita income). The R^2 is low but a qualitative analysis yields a conclusion: The more urbanized a municipal population is, the greater the chances that they consume more electricity. It is also clear to notice that the imaginary asymptote of 20 GWh is only crossed by municipalities that present 80% to 100% of urbanization level; this happens only for four municipalities, San Pedro Sula, Distrito Central, Villanueva and Choloma.

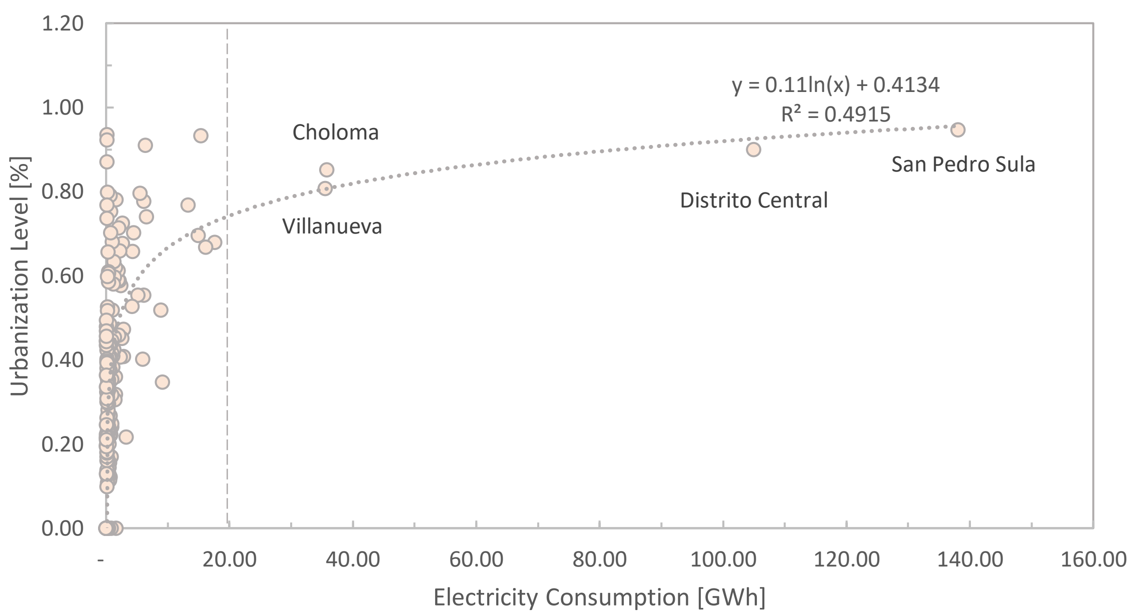


Figure 44: Relationship between Electricity Consumption and Urbanization Level

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

Another scatterplot might be helpful to understand the behavior of electrical demand with respect to urbanization level in municipalities. For this case a new value was obtained, the urban population.

Where urban population (UP) is defined as:

$$UP = Population\ [inhabitants] * Urbanization\ Level\ [\%]$$

Equation 25: Definition of urban population

Source: Own deduction

A linear relationship between urban population and electricity consumption may be seen on figure 47, also representing a high coefficient of determination.

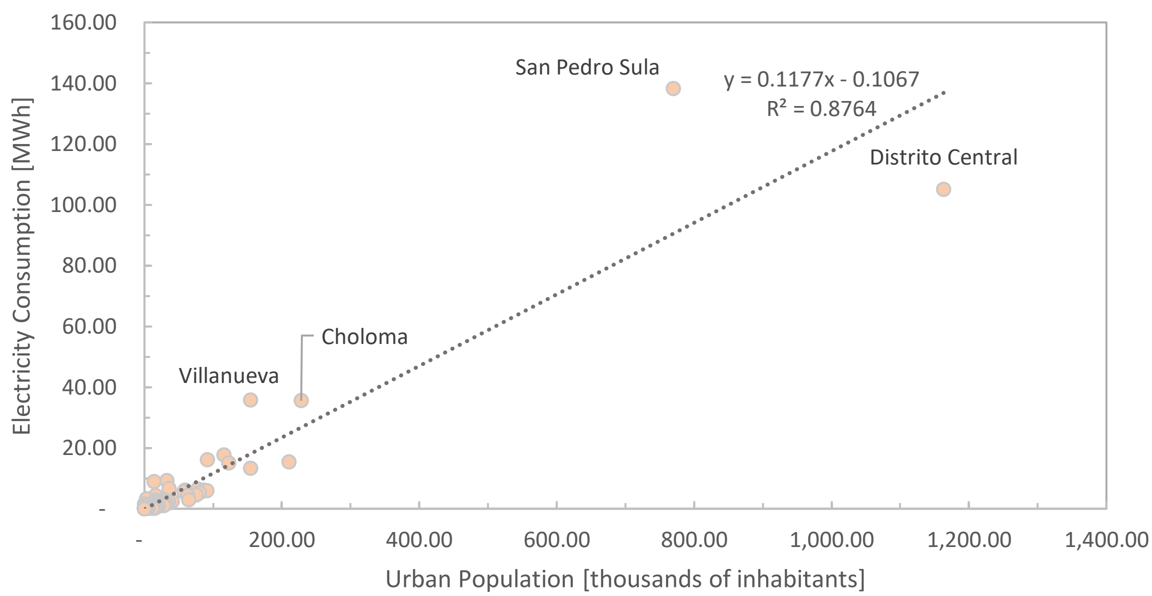


Figure 45: Relationship between electricity consumption and urban population

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (Empresa Nacional de Energía Eléctrica, 2019) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

5.2.5 ELECTRICAL COVERAGE [%]

It is natural to think that electrical coverage has an influence on electrical consumption, nevertheless this investigation concludes in the contrary. The maximum coefficient of determination (R^2) achieved was 0.10 and it was only when a 6^{th} degree polynomial was utilized.

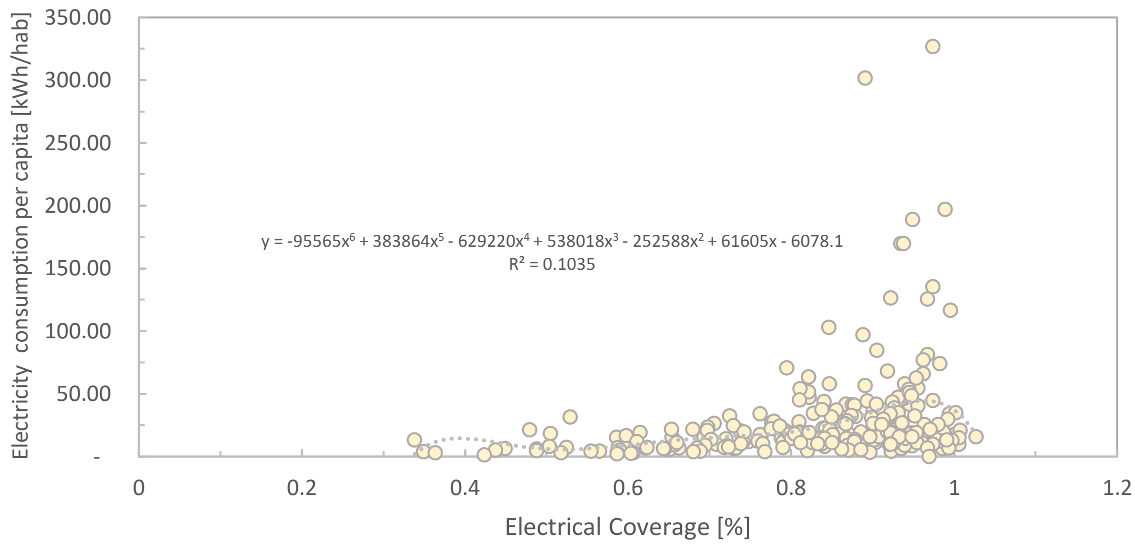


Figure 46: Relationship between electricity consumption per capita and Electrical Coverage

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Empresa Nacional de Energía Eléctrica, 2019)

Realizing that in the previous graph the relationship was difficult to achieve, the axes were shifted, obtaining the graph below where a better R^2 is achieved, double the first coefficient of determination.

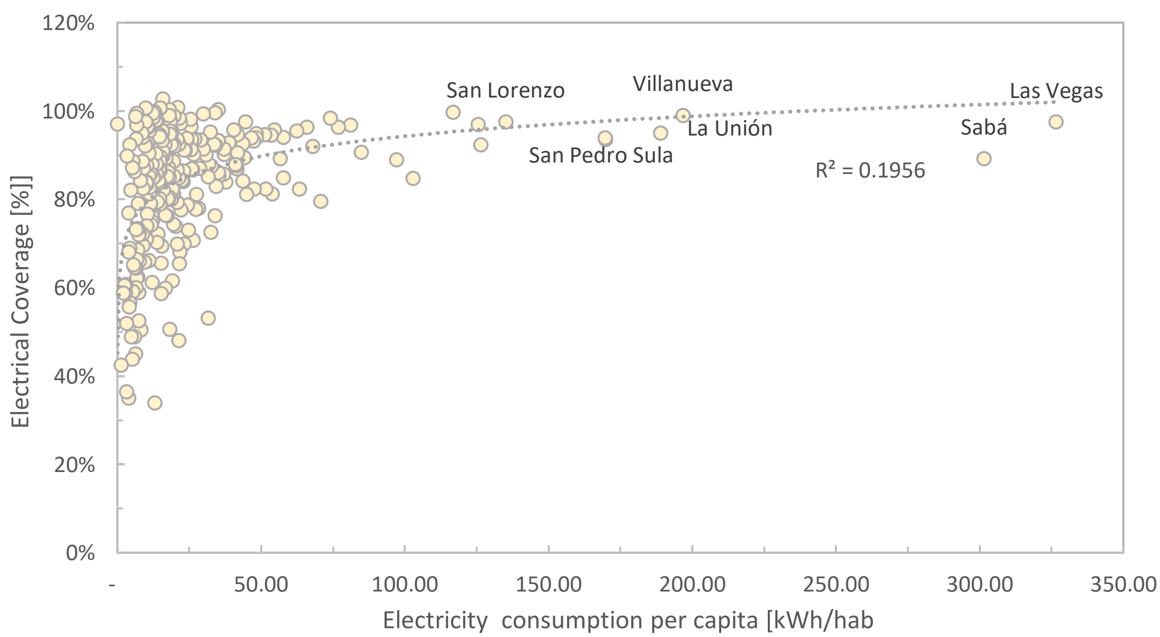


Figure 47: Relationship between Electricity Consumption per capita and Electrical Coverage (rotated axis)

Source: Own elaboration with information obtained via email from (EEH , 2022) and (Empresa Nacional de Energía Eléctrica, 2019)

The coefficient of determination is too low when comparing them with the population, per capita income, HDI and the other parameters.

Another scatterplot might be helpful to understand the behavior of electrical demand with respect to electrical coverage. In this case a new value was obtained, the covered population. Where covered population (CP) is defined as:

$$CP = Population[inhabitants] * Electrical\ coverage\ [\%]$$

Equation 26: Definition of covered population

Source: Own deduction

In Figure 50 a linear relationship is noticeable, and a high coefficient of determination is obtained.

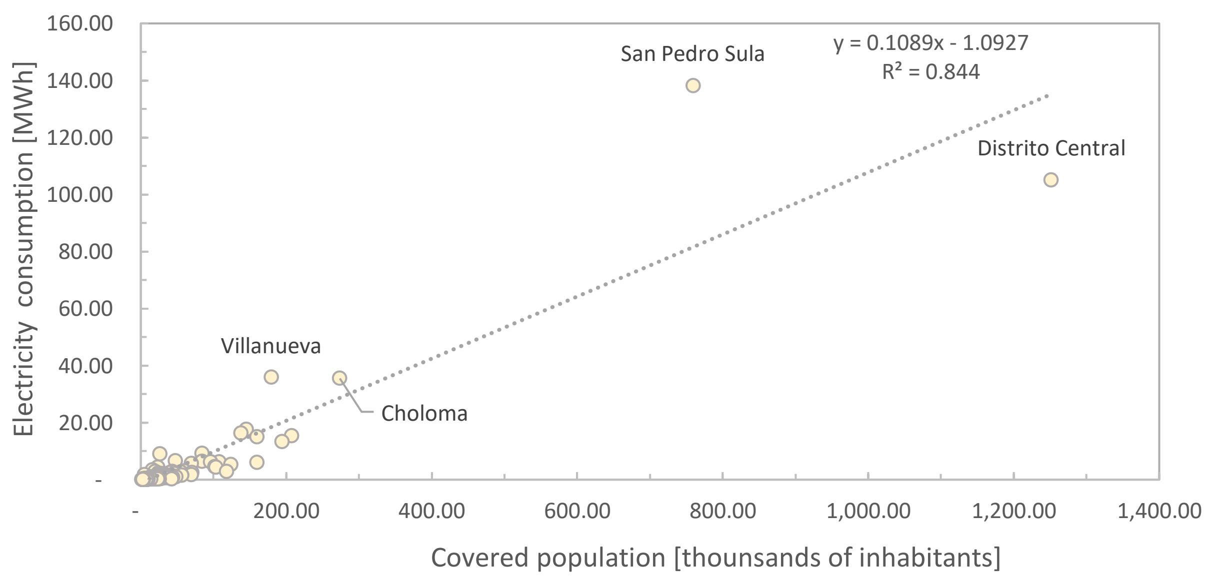


Figure 48: Relationship between electrical consumption and covered population

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (Empresa Nacional de Energía Eléctrica, 2019) and (Secretaría de Gobernación, Justicia y Descentralización, 2020)

5.2.6 INDUSTRIALIZATION LEVEL [INHABITANTS]

The INE in 2013, conducted a population census. One of the variables recollected then and interesting for this study is the value of those who live in a municipality and work in industrial or commercial sector. These values were tabulated for each municipality and utilized to understand the possible relationship between electricity consumption with respect to industrial and commercial occupation.

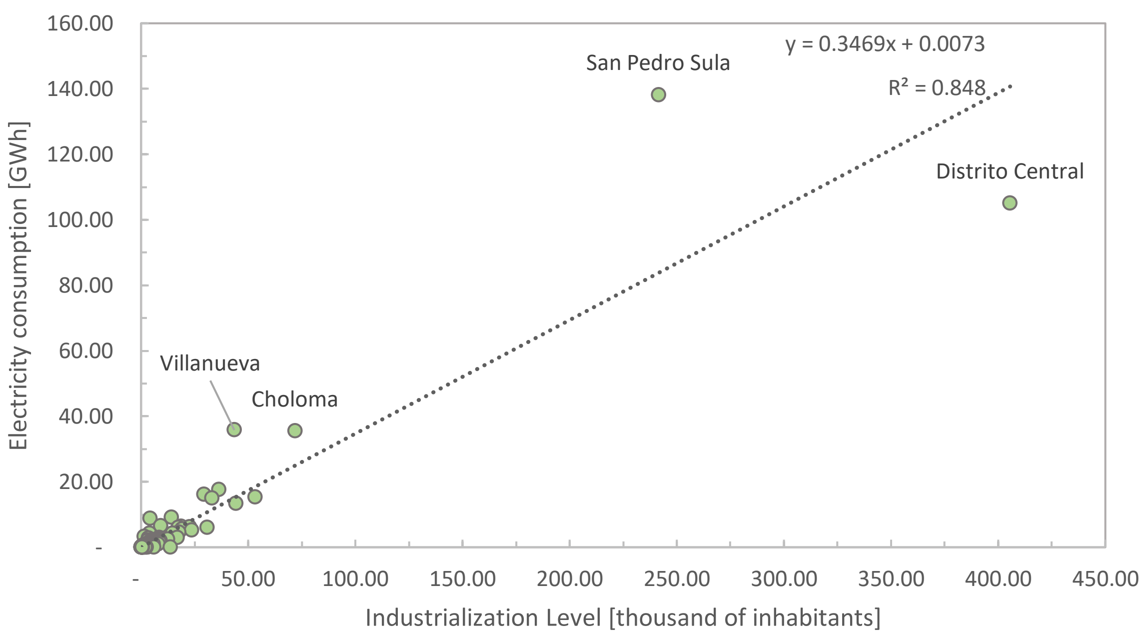


Figure 49: Relationship between electricity consumption and industrial/commercial occupation

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

The variables exhibit a linear relationship with a high coefficient of determination which is good for the linear regression model that is going to be achieved in section 5.3 of this thesis.

5.2.7 ETHNIC POPULATION [INHABITANTS]

In Honduras, 10 percent of the total population has an ethnic background. In this study of electric energy consumption, it is hypothesized that the municipality´s use of electricity will decrease as the ethnic diversity of the population increases.

Typically, the localities with the least amount of development in industrial and commercial activity are those where Indigenous people reside. The living circumstances, in favor of honoring their culture and beliefs, occasionally are basic, no luxuries or amenities in most of the cases.

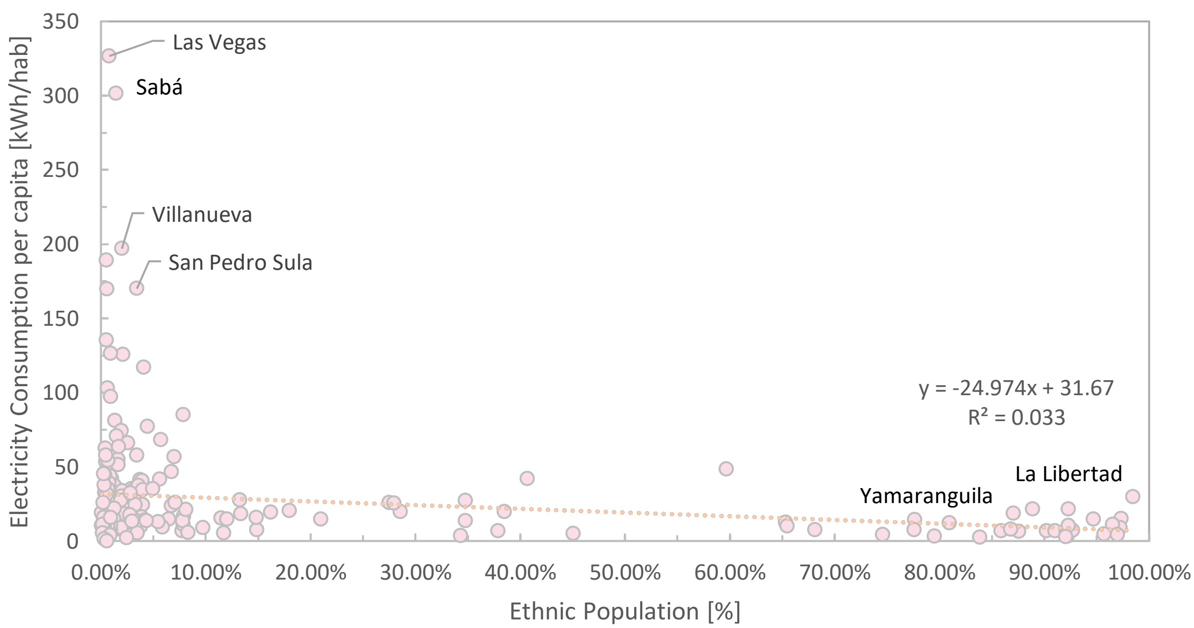


Figure 50: Relationship between ethnic population and electricity consumption

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

Because the coefficient of determination is low for a linear regression model in Figure 52 the relationship between the percentage of ethnic population and electricity consumption per capita is nonlinear. However, there is a relation between these parameters, the greater the percentage of ethnic population in a municipality the minor the ranges in which electricity consumption can vary, so the initial hypothesis is valid.

For instance, the municipalities that present a 60% to 100% of ethnic population such as Yamaranguila in Intibucá and La Libertad in Comayagua demonstrate to vary from 0 to 50 kWh per inhabitant, whereas in San Pedro Sula or Villanueva that demonstrate value of less than 10% of ethnic population that consumption of electricity ranges from 0 to 250 kWh per inhabitant.

5.2.8 Electricity Consumption for Cooking

The quantity of families that use electricity as a cooking source instead of wood or propane gas is one of the values recollected and interesting for this investigation. These results were tabulated for each municipality by the author and employed to determine whether there might be a correlation between electricity consumption and electricity use as a heat source for cooking purposes.

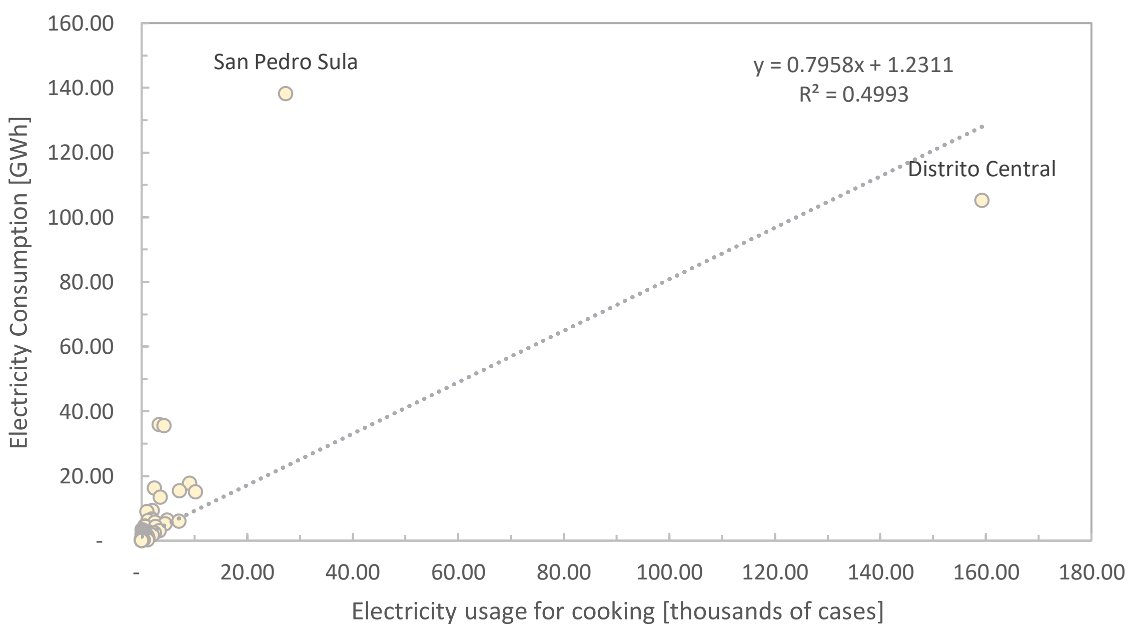


Figure 51: Relationship between electricity consumption and electricity usage for cooking

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

Considering Figure 53 the correlation between electricity usage for cooking and electricity consumption is not linear due to the low coefficient of determination. Also, there is no qualitatively explainable relationship noted.

5.2.9 ELECTRICITY CONSUMPTION FOR LIGHTNING

The quantity of households that utilized public service of lightning is important data for this investigation.

These results were tabulated for each municipality by the author and employed to determine whether there might be a correlation between electricity consumption and electricity use for lightning.

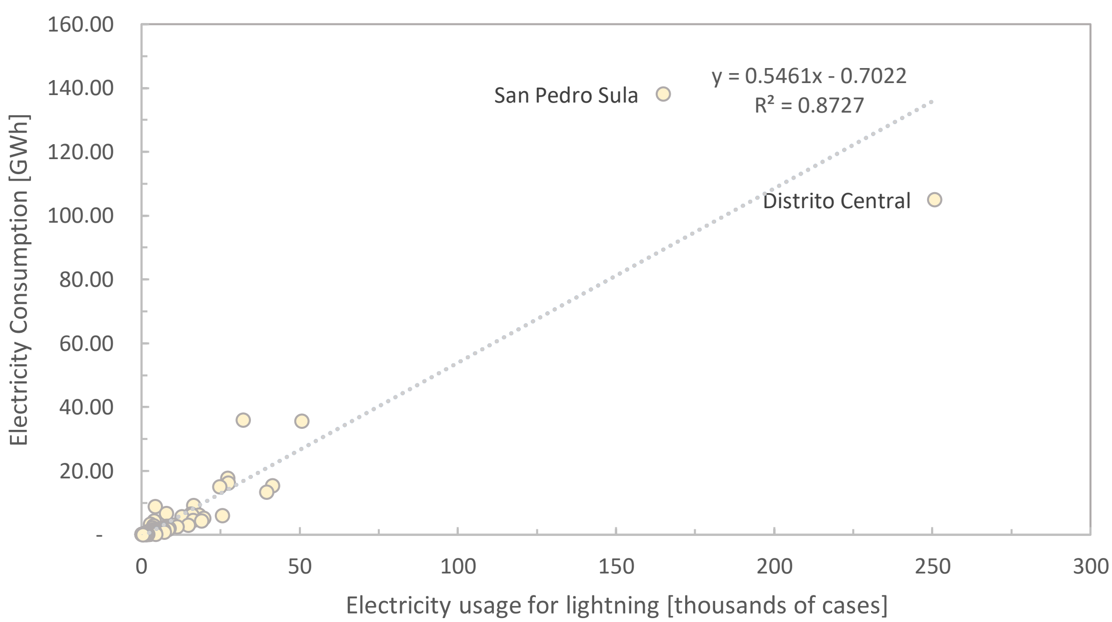


Figure 52: Relationship between electricity consumption and electricity usage of lightning

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

The electricity consumption in the municipality has a linear correlation with electricity usage for lightning, the coefficient of determination is high (greater than 0.7) indicating that the independent variables explain well the dependent variable.

5.2.10 POVERTY LEVEL BASED ON UNSATISFIED BASIC NEEDS

The poverty level in each municipality is one of the variables recollected back then and interesting for this investigation. These values were tabulated by the author and employed to determine whether there might be a correlation between electricity consumption and the poverty level in a region.

As a hypothesis, if the level of poverty is high for the population, then the consumption of electricity will be low.

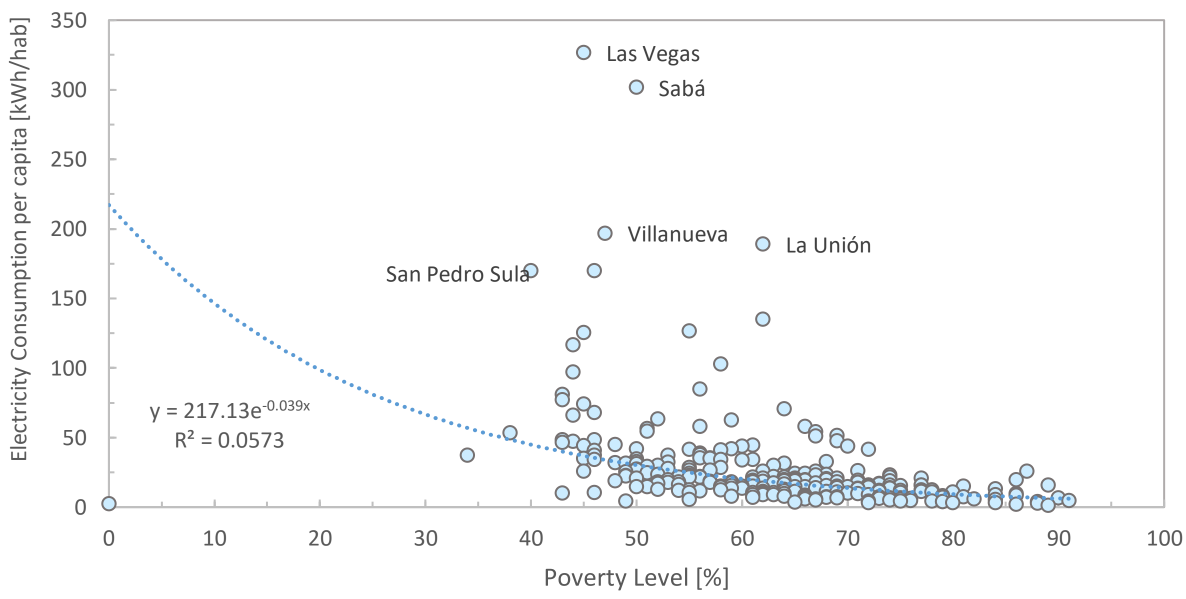


Figure 53: Exponential relationship between Poverty Level and Electricity Consumption

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

The coefficient of determination with an exponential analysis is low. A higher coefficient of determination can be reach with a linear relationship. However, the coefficient is still not valid for this thesis (less than 0.7).

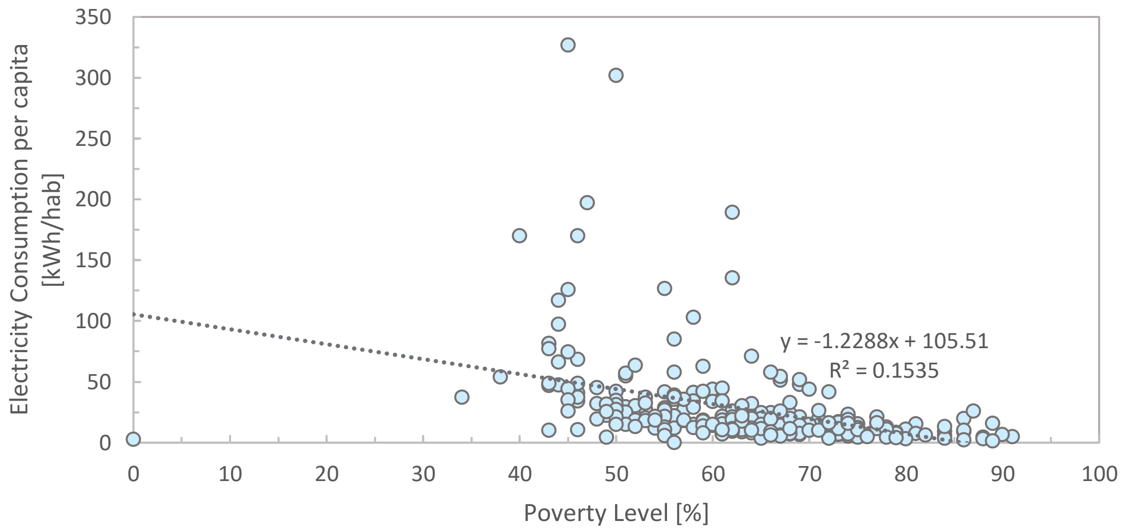


Figure 54: Linear relationship between Poverty Level and Electricity Consumption

Source: Own elaboration with information obtained via email from (EEH , 2022) and information from (INE, 2013)

The hypothesis is partly true because the higher the poverty level, the lower the ranges of electricity consumption. But the hypothesis could not be totally accepted because the R square is not high so the relationship between these parameters is nonlinear nor exponential. Another function must be adapted to this behavior.

5.3 REGRESSION MODEL PROCESS

The process to encounter the best fitted regression model started in RStudio where the 10 independent variables recollected were considered to create a multiple regression model and a quantile regression model to predict the dependent variable. This dependent variable is the electricity consumption for the 39 municipalities that were not in the sample.

5.3.1 MULTIPLE LINEAR REGRESSION MODEL

In RStudio, the characteristics tabulated were named as variables (y and x's). Where the y represents the independent variables, and the x represents dependent variables.

Table 4: Variables names in RStudio

Characteristics	Variable assigned
Electricity Consumption [MWh]	Y
Population [inhabitants]	X_1
Urbanization Level [%]	X_2
HDI [-]	X_3
Income per capita [L.]	X_4
Industrial Occupation [Inhabitants]	X_5
Ethnic Population [%]	X_6
Poverty Level [-]	X_7
Electricity Usage for cooking [%]	X_8
Electricity Usage for lightning [%]	X_9
Electrical coverage [%]	X_{10}

Source: Own Elaboration

Utilizing the library "readxl" the tabulated excel data was available to the program RStudio and with the function "linear model" (represented in the compiler as" lm ()") a model was predicted. The information obtained from this first experiment and the process of iteration is detailed below.

After simulating the electricity consumption dependable of the ten variables recollected by the author, R gives as an output the ANOVA analysis of the regression model. This ANOVA analysis is achieved utilizing the function “summary ()” in the compiler.

Three parameters need to be verified in accordance with this thesis' methodology.

1. **The coefficient of determination** for a linear model including the total features tabulated (10 sociodemographic variables) is valid. This acceptance is because the requirement was met, the R^2 for this model is 0.8607 which is greater than 0.7.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.648e+03  9.688e+03  -0.893   0.3729
x1           1.934e-02  1.827e-02   1.059   0.2908
x2           1.652e+03  1.528e+03   1.081   0.2807
x3           1.321e+04  1.366e+04   0.967   0.3344
x4           3.240e+00  1.619e+00   2.002   0.0464 *
x5           2.942e-01  6.020e-02   4.887 1.84e-06 ***
x6          -5.597e+01  1.074e+03  -0.052   0.9585
x7          -3.030e+00  2.968e+01  -0.102   0.9188
x8          -2.660e+04  5.386e+03  -4.938 1.45e-06 ***
x9           5.238e+02  2.441e+03   0.215   0.8303
x10         -6.511e+02  2.990e+03  -0.218   0.8278
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4234 on 248 degrees of freedom
Multiple R-squared:  0.8661,	Adjusted R-squared:  0.8607
F-statistic: 160.5 on 10 and 248 DF,  p-value: < 2.2e-16
```

Figure 55: ANOVA analysis in RStudio (R^2)

Source: Own Elaboration through RStudio

2. Following the methodology, another requirement must be satisfied, the **p-value of the general model** must less than 0.05 meaning that the overall model is significant. This linear model of 10 variables independent meets the requirement because the value of general p-value is less than 2.2e-16.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -8.648e+03  9.688e+03  -0.893   0.3729    
x1           1.934e-02  1.827e-02   1.059   0.2908    
x2           1.652e+03  1.528e+03   1.081   0.2807    
x3           1.321e+04  1.366e+04   0.967   0.3344    
x4           3.240e+00  1.619e+00   2.002   0.0464 *  
x5           2.942e-01  6.020e-02   4.887 1.84e-06 ***
x6          -5.597e+01  1.074e+03  -0.052   0.9585    
x7          -3.030e+00  2.968e+01  -0.102   0.9188    
x8          -2.660e+04  5.386e+03  -4.938 1.45e-06 ***
x9           5.238e+02  2.441e+03   0.215   0.8303    
x10         -6.511e+02  2.990e+03  -0.218   0.8278    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4234 on 248 degrees of freedom
Multiple R-squared:  0.8661,	Adjusted R-squared:  0.8607 
F-statistic: 160.5 on 10 and 248 DF,  p-value: < 2.2e-16
```

Figure 56: ANOVA analysis in RStudio (general p-value)

Source: Own Elaboration through RStudio

3. Another condition for the ANOVA analysis is to determine that the **slope coefficients** are significant (greater that 0 in absolute value).


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -8.648e+03  9.688e+03  -0.893   0.3729    
x1           1.934e-02  1.827e-02   1.059   0.2908    
x2           1.652e+03  1.528e+03   1.081   0.2807    
x3           1.321e+04  1.366e+04   0.967   0.3344    
x4           3.240e+00  1.619e+00   2.002   0.0464 *  
x5           2.942e-01  6.020e-02   4.887 1.84e-06 ***
x6          -5.597e+01  1.074e+03  -0.052   0.9585    
x7          -3.030e+00  2.968e+01  -0.102   0.9188    
x8          -2.660e+04  5.386e+03  -4.938 1.45e-06 ***
x9           5.238e+02  2.441e+03   0.215   0.8303    
x10         -6.511e+02  2.990e+03  -0.218   0.8278    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4234 on 248 degrees of freedom
Multiple R-squared:  0.8661,	Adjusted R-squared:  0.8607 
F-statistic: 160.5 on 10 and 248 DF,  p-value: < 2.2e-16
```


Figure 57: ANOVA analysis in RStudio (Slope coefficients)

Source: Own Elaboration through RStudio

This requirement is met by every independent variable of this model, comparing the “Estimate” column as absolute values to zero, all result significant.

4. The last parameter to be checked is **the p-value for each independent variable.** Each p-value must be less than 0.05 to be valid for this thesis methodology. This happens only with the variable x_4, x_5 and x_8 for this first model of the regression function.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.648e+03  9.688e+03  -0.893   0.3729
x1           1.934e-02  1.827e-02   1.059   0.2908
x2           1.652e+03  1.528e+03   1.081   0.2807
x3           1.321e+04  1.366e+04   0.967   0.3344
x4           3.240e+00  1.619e+00   2.002   0.0464 *
x5           2.942e-01  6.020e-02   4.887 1.84e-06 ***
x6          -5.597e+01  1.074e+03  -0.052   0.9585
x7          -3.030e+00  2.968e+01  -0.102   0.9188
x8          -2.660e+04  5.386e+03  -4.938 1.45e-06 ***
x9           5.238e+02  2.441e+03   0.215   0.8303
x10         -6.511e+02  2.990e+03  -0.218   0.8278
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4234 on 248 degrees of freedom
Multiple R-squared:  0.8661,    Adjusted R-squared:  0.8607
F-statistic: 160.5 on 10 and 248 DF,  p-value: < 2.2e-16
```

Figure 58: ANOVA analysis in RStudio (each p-value)

Source: Own Elaboration through RStudio

Using the function “forest model ()” and transforming the output into a portable network graphics document, the subsequent table is achieved.

Variable	N	Estimate	p
x1	259	19.34 (-16.65, 55.33)	0.29
x2	259	1652283.37 (-1357559.47, 4662126.22)	0.28
x3	259	13208758.89 (-13690190.50, 40107708.28)	0.33
x4	259	3240.29 (51.78, 6428.80)	0.05
x5	259	294.21 (175.63, 412.78)	<0.001
x6	259	-55967.55 (-2170322.21, 2058387.10)	0.96
x7	259	-3030.07 (-61478.08, 55417.94)	0.92
x8	259	-26597191.79 (-37204829.93, -15989553.65)	<0.001
x9	259	523826.47 (-4283946.86, 5331599.81)	0.83
x10	259	-651142.60 (-6539624.52, 5237339.32)	0.83

Figure 59:Summary of linear regression model with 10 independent variables

Source: Own elaboration through RStudio

From the model of ten variables already discussed, a fast way to discard variables that are not significant to the model based on the p-value is utilizing the function in R called "ols_step_both_p" which gives as an output a set of variables that work together as a model where all variables are significant.

In Figure 62 the result is shown where x_4, x_5 x_8 and x_2 are the only variables remaining.

Stepwise Selection Summary

Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	x5	addition	0.848	0.848	26.1200	5088.0402	4428.5432
2	x8	addition	0.853	0.851	20.0020	5082.4512	4372.6505
3	x4	addition	0.862	0.861	3.8010	5066.5816	4232.6533
4	x2	addition	0.865	0.863	1.6050	5064.2804	4205.9076

Figure 60: Output of the function ols_step_both_p ()

Source: Own elaboration through RStudio

Opting now only for income per capita (x_4), industrialization level (x_5), electrical coverage (x_8) and urbanization level (x_2) as the independent variables, a new linear model is created with the function "lm ()" obtaining the subsequent results.

```
                 1.                                   2.
Coefficients:
              Estimate  Std. Error t value  Pr(>|t|)
(Intercept) -9.086e+02  5.245e+02  -1.732  0.08444 .
x8          -2.490e+04  4.950e+03  -5.031  9.25e-07 ***
x2           2.751e+03  1.334e+03   2.062  0.04019 *
x4           4.116e+00  1.376e+00   2.991  0.00305 **
x5           3.574e-01  1.001e-02  35.686  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4206 on 254 degrees of freedom
Multiple R-squared:  0.8647,    Adjusted R-squared:  0.8626
F-statistic: 405.9 on 4 and 254 DF,  p-value: < 2.2e-16
```
3 and 4

Figure 61: ANOVA analysis of final linear regression model

Source: Own elaboration through RStudio

Where:

1. The slope coefficients are not insignificant-
2. The p-value for each variable is less than 0.05.
3. The adjusted R-squared or coefficient of determination (R^2) is greater than 0.7
4. P-value for the model is less than 0.05.

With these four characteristics it is concluded that the model is valid. Being defined by the next equation.

$$y = -908.6 + 0.357x_5 + 4.11x_4 + 2751x_2 - 24900x_8$$

Equation 27: Final multiple linear regression model

Source: Own deduction through RStudio

Where:

- y is electricity consumption [MWh]
- x_4 is income per capita [L.]
- x_5 is industrialization level [habitants that work in commercial and industrial sector]
- x_2 is urbanization level [%]
- x_8 is the electrical coverage [%]

More information about the model can be achieved in the subsequent table.

Variable	N	Estimate		p
x8	259		-24903.22 (-34651.94, -15154.50)	<0.001
x2	259		2750.63 (124.09, 5377.18)	0.040
x4	259		4.12 (1.41, 6.83)	0.003
x5	259		0.36 (0.34, 0.38)	<0.001

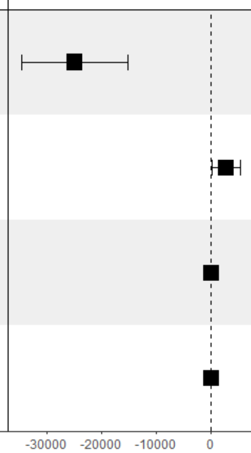


Table 5: Final multiple linear regression model parameters

Source: Own elaboration with information from final regression model obtained through RStudio

The code for the multiple linear regression can be found in Annex 2.

5.3.2 QUANTILE REGRESSION MODEL

A multiple linear model was already obtained to predict the consumption of electricity based on the industrialization level, per capita income, urbanization level, and electrical coverage per municipality. The function is described in section 5.3.1 as Equation 33. However, a challenge is faced. Utilizing this function, negative numbers appear as an answer of electricity consumption *(y)*, which is impossible because ranges of electricity consumption can only be positive.

Because of that matter, quantile regression is taken into consideration. Quantile regression gives a result of the dependent variable as a range based on the probability of the value. For instance, applying quantile regression to the model discussed in section 5.3.1 will entail generating a range of values for Electricity Consumption per each municipality.

The accumulative probability of a value happening is also called quantile level. A quantile regression indicates a function per each quantile studied. For this study the quantile levels considered will be when τ is equal to 0.05 and when τ is equal to 0.95. The value τ was defined in Chapter 3 Section 3.3.4 as well as this regression method.

Table 6: Example of ranges of electricity consumption per municipality

Municipality: XX	$\boldsymbol{\tau} = 0.05$	$\boldsymbol{\tau} = 0.95$
Electricity Consumption [MWh]	-40	500

Source: Own elaboration

The ranges achieved will vary from the value that has 5% of probability of occurrence to the value that has 95% of accumulative probability of occurrence. Understanding by this, that the real value of consumption can vary between these two values. This range is not exempt of negative values but for this thesis, negative values will be considered equal to zero.

Table 7: Example of ranges of electricity consumption when negatives are discarded

Municipality: XX	$\boldsymbol{\tau} = 0.05$	$\boldsymbol{\tau} = 0.95$
Electricity Consumption [MWh]	0 (before -40)	500

Source: Own elaboration

The quantile regression will be done in RStudio. The library “quantreg” must be downloaded. Inside this library, the function “regression quantile” found in the compiler as rq () should be used to enter the data tabulated in excel. Utilizing the library “readxl” the tabulated excel data will be available. The code for the quantile regression in R can be found in Annex 3.

The results of the quantile regression based on four independent variables are shown below:

```
tau: [1] 0.05

Coefficients:
            Value       Std. Error  t value     Pr(>|t|)
(Intercept)  -106.13946   35.95524   -2.95199    0.00345
x4             -0.02702    0.13792   -0.19591    0.84483
x5              0.20918    0.04377    4.77884    0.00000
x8          -2422.07463 1615.47148   -1.49930    0.13504
x2             66.02189  179.75329    0.36729    0.71371
```

Figure 62: ANOVA Analysis for quantile regression where $\tau = 0.05$

Source: Own elaboration through RStudio

For both scenarios, when the quantile level is equal to 0.05 and 0.95, the independent variables x_4, x_8 and x_2 show p-values that exceed the 0.05 value, this means a low significance for the model.

```
tau: [1] 0.95

Coefficients:
            Value      Std. Error t value    Pr(>|t|)
(Intercept) -352.74763  124.71515   -2.82843    0.00505
x4             1.67610    0.46340    3.61696    0.00036
x5             0.56641    0.00344  164.68945    0.00000
x8           834.68022 2790.85685    0.29908    0.76513
x2          -105.95291  336.60040   -0.31477    0.75319
```

Figure 63: ANOVA Analysis for quantile regression where τ= 0.95

Source: Own elaboration through RStudio

Just the variable of industrialization level (x_5) maintains a value valid as a p-value in both scenarios. That is why only this variable will be considered for the quantile regression model.

Table 8: Quantile regression model for tau equal to 0.05 and 0.95

Quantile regression model	$\tau = 0.05$	$\tau = 0.95$
Intercept	-212.44977	-31.28594
Coefficient	0.20019	0.65464

Source: Own elaboration with information from simulations of RStudio

The respective functions can be obtained through Table 8

$$y = -212.44977 + 0.20019 \cdot x_5$$

Equation 28: Quantile Regression for tau equal to 0.05

Source: Own deduction from information and library of RStudio

$$y = -31.28594 + 0.65464 \cdot x_5$$

Equation 29: Quantile Regression for tau equal to 0.95

Source: Own deduction from information and library of RStudio

In both cases:

- y is electricity consumption [MWh]
- x_5 is industrialization level [habitants that work in commercial and industrial sector]

For this research, the maximum electrical demand will be considered per each municipality. For planning generation, the worst-case scenario (maximum demand) must be contemplated for this reason the quantile regression for tau equal to 0.95 is considered as the final quantile regression.

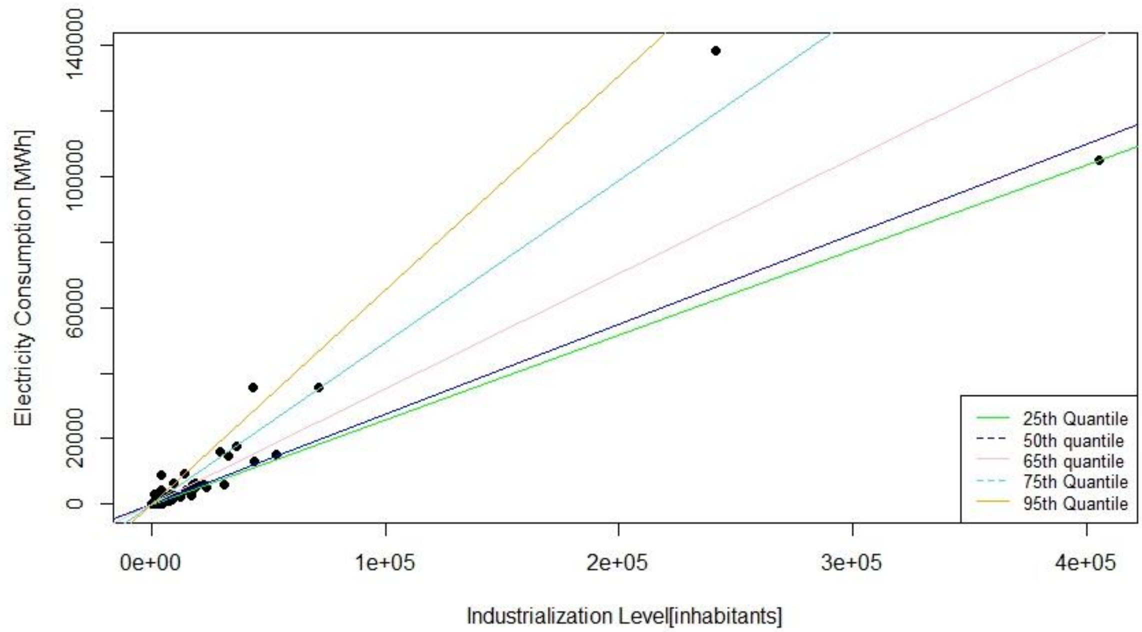


Figure 64: Quantile linear regression model for electricity consumption

Source: Own Elaboration in RStudio

A comparison between the multiple regression model and the quantile univariable regression model for the data obtained for the 39 municipalities remaining is shown in Annex 4.

5.4 GLOBAL DATA VISUALIZATION RESULTS

In this section a departmental analysis will be carried on. The total electricity consumption, per capita electricity consumption, and electricity density will be calculated for each department as it was done in the first section per municipality. Finally, it is studied what percentage of departmental consumption represent each municipal consumption. The figures shown in this section are available in Tableau Public (Munguía Deras, 2022). In this document only the analysis of Atlántida will be carried out, if the reader would like to investigate further the departmental analysis of the 17 remaining departments this could be achieved accessing to Annex 10.

5.4.1 ATLÁNTIDA DEPARTMENT

Atlántida is in the north of Honduras. It is composed by eight municipalities.

Table 9: Electricity consumption characteristics of Atlántida

Atlántida		
Electricity Consumption [MWh]	**Per capita electricity consumption [kWh/hab]**	**Electricity Density [kWh/m^2]**
25,216.56	51.10	5.86

Source: Own elaboration with information obtained via email from (EEH , 2022)

With the cartographic tool created in Tableau electrical consumption is easy to track, the choropleth map of electricity consumption for Atlántida is shown below.

Figure 65: Choropleth map of electricity consumption in Atlántida

Source: Own elaboration with information obtained via email from (EEH , 2022)

In Atlántida, La Ceiba is the municipality with the highest electricity consumption, accounting for 61% of the department's total consumption. Arizona, in the other hand, barely registers in terms of overall consumption.

Table 10: Percentages of electricity consumption for Atlántida municipalities

Municipality	Percentage
La Ceiba	60.9%
Tela	24.1%
La Masica	4.1%
El Porvenir	4.0%
San Francisco	3.1%
Jutiapa	2.4%
Esparta	0.8%
Arizona	0.4%

Source: Own elaboration with information obtained via email from (EEH , 2022)

Figure 66: Choropleth map of electricity density in Atlántida

Source: Own elaboration with information obtained via email from (EEH , 2022)

The only municipality in Atlántida in which the range of electricity density is between 10 and 50 kWh/m2 is La Ceiba.

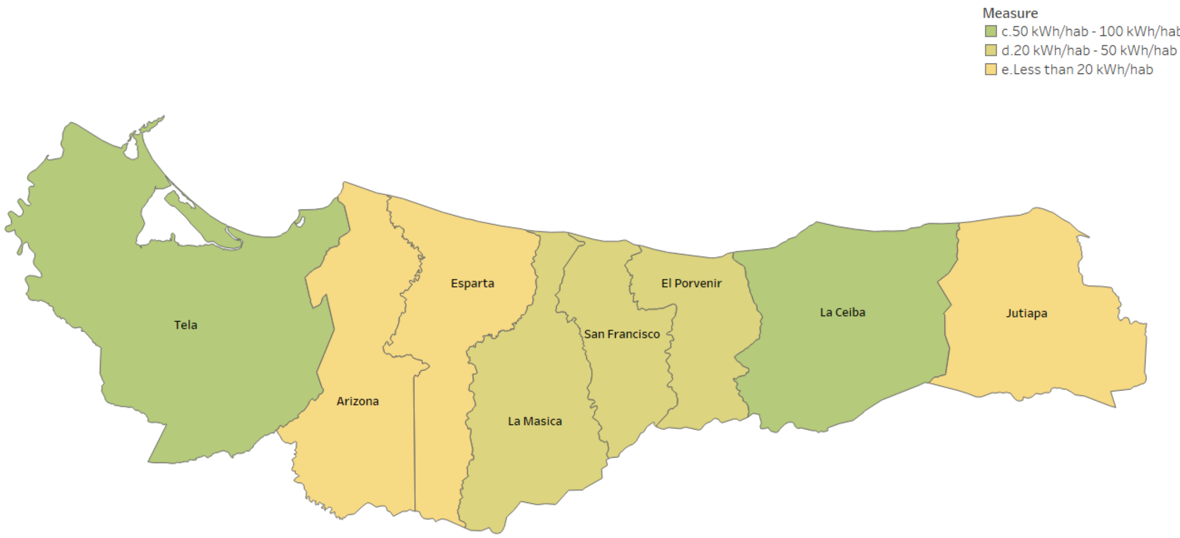


Figure 67: Choropleth map of electricity consumption per capita in Atlántida

Source: Own elaboration with information obtained via email from (EEH , 2022)

La Ceiba has a greater total electricity consumption; however, in comparison between municipalities for the electricity consumption per capita La Ceiba and Tela exhibit similar behavior, both comprehend ranges of 50 to 100 kWh per capita.

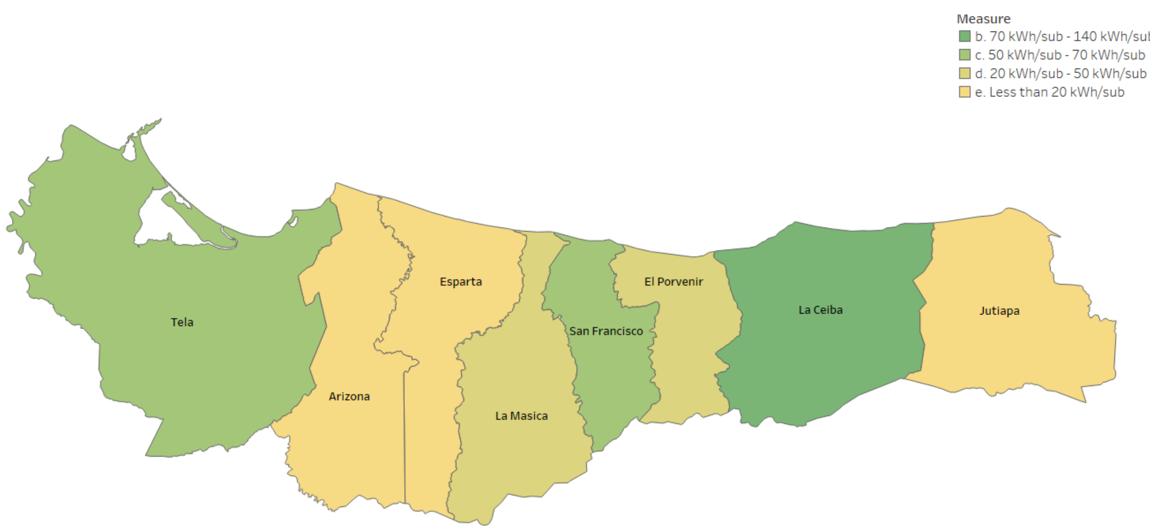


Figure 68: Choropleth Map of electricity consumption per subscriber in Atlántida

Source: Own elaboration with information obtained via email from (EEH , 2022)

San Francisco shows a higher consumption per subscriber in comparison to its value of electricity consumption per capita.

5.5 SECTORIAL ANALYSIS

In this section a multiple regression model is obtained for residential, commercial, and industrial consumption. This regression analysis is not done with the purpose of finding a function that predicts the demand but to understand which variables might affect the residential consumption.

Besides the multiple regression model, an analysis is done for each main sector: residential, commercial, and industrial.

- In the residential sector, there is a discussion about consumption per household and how much of the mean income of a household is utilized to cover electricity consumption needs in each municipality.
- In the commercial sector, there is a discussion about which type of commerce has greater influence in the electricity consumption at commercial level.
- In the industrial sector, there is a discussion about which industries have greater influence in the electricity consumption at industrial level.

5.5.1 Residential Sector

In Annex 5 the table where residential electricity consumption is tabulated for each one of the 259 municipalities is found this information was obtained from EEH in 2022. The analysis of the residential sector will consider two parts: the regression analysis and the consumption analysis.

a) Multiple regression model for residential consumption of electricity

For the multiple regression model for residential consumption the same methodology explained in section 5.4.1 was applied.

First, 10 independent variables were considered to train the model in RStudio, the dependent variable to be predicted is the electricity consumption in the residential sector.

Table 11: Variables considered for residential consumption model

Characteristics	Variable assigned
Residential Electricity Consumption [kWh]	Y
Population [inhabitant]	X_1
Income [lempiras]	X_2
Ethnic Population [inhabitants]	X_3
Urbanization Level [inhabitants]	X_4
Industrial Level [inhabitants]	X_5
HDI [-]	X_6
Electricity Usage for illumination [cases]	X_7
Electricity Usage for cooking [cases]	X_8
Electricity coverage [inhabitants]	X_9
Households with electrical coverage [cases]	X_{10}

Source: Own elaboration with information obtained via email from (EEH , 2022)

Now that the 10 variables are recognized, in RStudio the "lm ()" function is called to introduce the data to the program and the" summary ()" function is called to obtain information about the model.

```
                1.                                      2.
                 Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)     3.240e+05  3.140e+05    1.032   0.30305
X1              4.090e+00  3.057e+00    1.338   0.18206
X2              8.720e-03  2.767e-03    3.151   0.00183 **
X3             -5.510e-01  3.250e+00   -0.170   0.86551
X4                     NA         NA       NA        NA
X5              1.019e+02  1.135e+01    8.977   < 2e-16 ***
X6             -4.318e+05  4.840e+05   -0.892   0.37311
X7             -1.008e+01  2.554e+01   -0.395   0.69340
X8              6.865e+00  5.988e+00    1.146   0.25269
X9             -3.041e+01  4.506e+00   -6.750  1.03e-10 ***
X10             1.019e+02  1.205e+01    8.461  2.29e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 244200 on 249 degrees of freedom
  (39 observations deleted due to missingness)
3. Multiple R-squared:  0.9954,    Adjusted R-squared:  0.9953
F-statistic:  6012 on 9 and 249 DF,  p-value: < 2.2e-16
```

Figure 69: First simulation of residential consumption regression model

Source: Own elaboration through RStudio

Figure 138 provides information about the model of residential consumption; in rectangle one: the coefficient values are significant for all the dependent variables, in rectangle two: the p-values are only less than 0.05 for the variables X_2, X_5, X_9 and X_{10}. In rectangle 3 the coefficient of determination shown called Adjusted-R^2 is greater than 0.7.

From the model of ten variables already discussed, a fast way to discard variables that are not significant to the model based on the p-value is utilizing the function in R called "ols_step_both_p" which gives as an output a set of variables that work together as a model where all variables are significant.

Figure 138 reveals that only X_{10}, X_9 and X_5 are significant for the model.

```
                           Stepwise Selection Summary
---------------------------------------------------------------------
                     Added/                    Adj.
Step    Variable    Removed     R-Square     R-Square      C(p)
---------------------------------------------------------------------
   1       X2       addition       0.988        0.988     351.5230
   2       X5       addition       0.992        0.992     156.0210
   3      X10       addition       0.993        0.993     115.2790
   4       X2       removal        0.993        0.993     113.5850
   5       X9       addition       0.995        0.995      10.2450
---------------------------------------------------------------------
```

Figure 70: Output of the function ols_step_both_p () for the residential sector

Source: Own elaboration through RStudio

Opting now only for electricity coverage (x_9), industrial level (x_5), and households with electrical coverage (x_{10}) as the independent variables, a new linear model is created with the function "lm ()" obtaining the subsequent results.

```
                  1.                           2.
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29652.506  21004.839   1.412    0.159
X10           103.006      7.572  13.603   <2e-16 ***
X5            114.722      4.098  27.998   <2e-16 ***
X9            -22.842      2.253 -10.140   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 247200 on 255 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9951
F-statistic: 1.76e+04 on 3 and 255 DF,  p-value: < 2.2e-16
3.
```

Figure 71: ANOVA analysis of final regression model for residential sector

Source: Own elaboration through RStudio

In Figure 140, the best fitted model obtained through RStudio is shown, is important to note that:

1. The estimated coefficients are significant.

2. The p-values are less than 0.05, meaning that each variable is significant in the model.

3. The coefficient of determination is greater than 0.7, therefore acceptable.

Finally, the followed-up equation is obtained:

$$y = 29652.51 + 103.01x_{10} - 22.84x_9 + 114.72x_5$$

Equation 30: Final regression model to predict residential electricity consumption

Source: Own elaboration with information obtained from RStudio *b) Residential consumption analysis*

An analysis of consumption per household will be described in this section. Likewise, an economic analysis of the budget that Honduran families must manage to pay for the electrical services is conducted.

To obtain the electrical consumption per household (EPCH), it was required the residential consumption for each municipality which was already obtained in Subsection A of the 5.5.1 Section and to know the number of households that are covered electrically in each municipality, which is the variable X_{10} introduced in in Subsection A. This data was found in (Empresa Nacional de Energía Eléctrica, 2019).

To obtain the electricity consumption per household for each municipality the subsequent equation must be utilized:

$$\mathrm{ECPH} = \frac{\text{Residential electricity consumption [kWh]}}{\text{Households with electrical coverage [cases]}}$$

Equation 31: Definition of electricity consumption per household

Source: Own deduction

The electricity consumption per household in a month was obtained for every municipality, In Annex 6 this data can be found.

For the original data (259 municipalities) Santa Rita in Yoro is the municipality that consumes the most per household. In Santa Rita, on average, a house consumes 258 kWh per month.

Considering the 30 municipalities that consume the most at the household level, like in Figure 140, only Santa Rita, Bonito Oriental, and Valle de Ángeles exceed the consumption of 200 kWh per household while the rest maintain a consumption between 180 to 110 kWh in a month of electricity consumption.

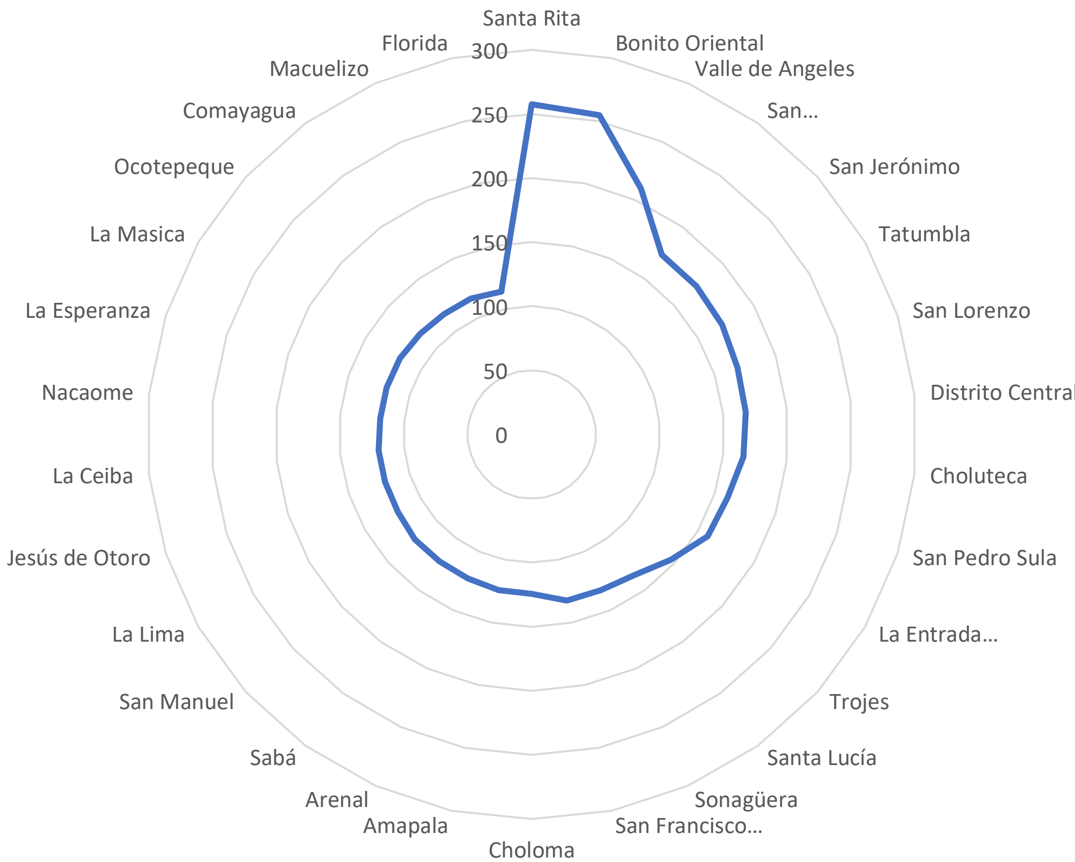


Figure 72: Radar Graph of electricity consumption per household in Honduras' municipalities considering the top 30 consumers

Source: Own elaboration with information obtained via email from (EEH , 2022)

A comparison is made between the lowest consumption per household in Arizona and the greatest consumption of electricity per household in Santa Rita. In between the important municipalities of San Pedro Sula and Distrito Central are shown.

The consumption in Santa Rita is 11,727 times higher per household than in Arizona, this again, shows how inequal electricity consumption is distributed in the national territory.

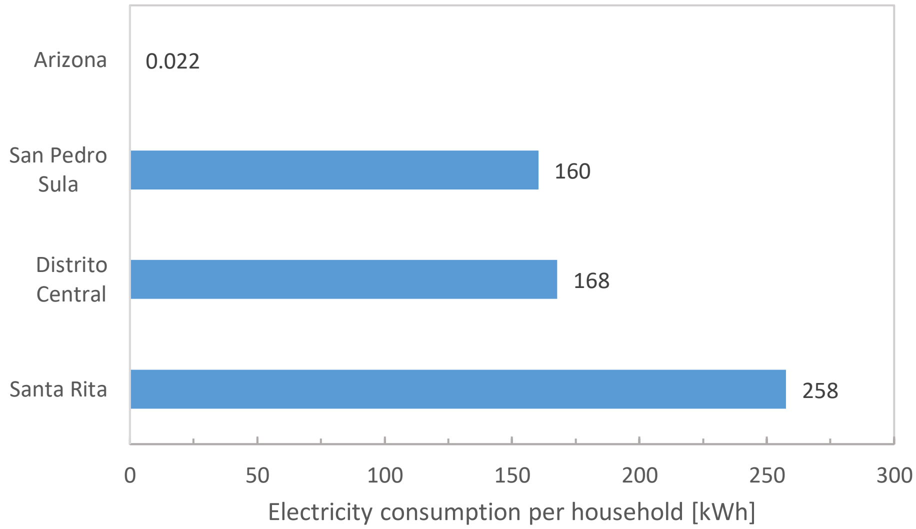


Figure 73: Comparison between consumptions in municipalities

Source: Own elaboration with information obtained via email from (EEH , 2022)

Those who have large consumptions will also be the ones who will pay more for electricity because in Honduras currently a payment of fixed rate electricity service is contemplated. For the residential sector, the first 50 kWh consumed are fixed in certain tariff (cheaper than the normal tariff), if the consumption is greater than 50 kWh then the tariff changes. In Table 46 this is better explained.

Table 12: Residential Tariff for June-August trimester

Residential Sector	Electricity Price [$/kWh]
First 50 kWh consumed	0.1792
The following consumption of kWh	0.2331

Source: Own elaboration based on (CREE, 2022)

To clarify this concept an example is postulated. A household where 150 kWh were consumed is going to pay:

$$Total\ payment = \left(50kWh \cdot \frac{0.1792\ \$}{kWh}\right) + \left((150\ kWh - 50\ kWh) \cdot \frac{0.2331\ \$}{kWh}\right)$$

$$Total\ payment = \$\ 31.85$$

Because of this tariff, the consumption is proportional to the budget destined to the electricity consumption and has the intercept that represents the change in tariff in the limit of 50 kWh.

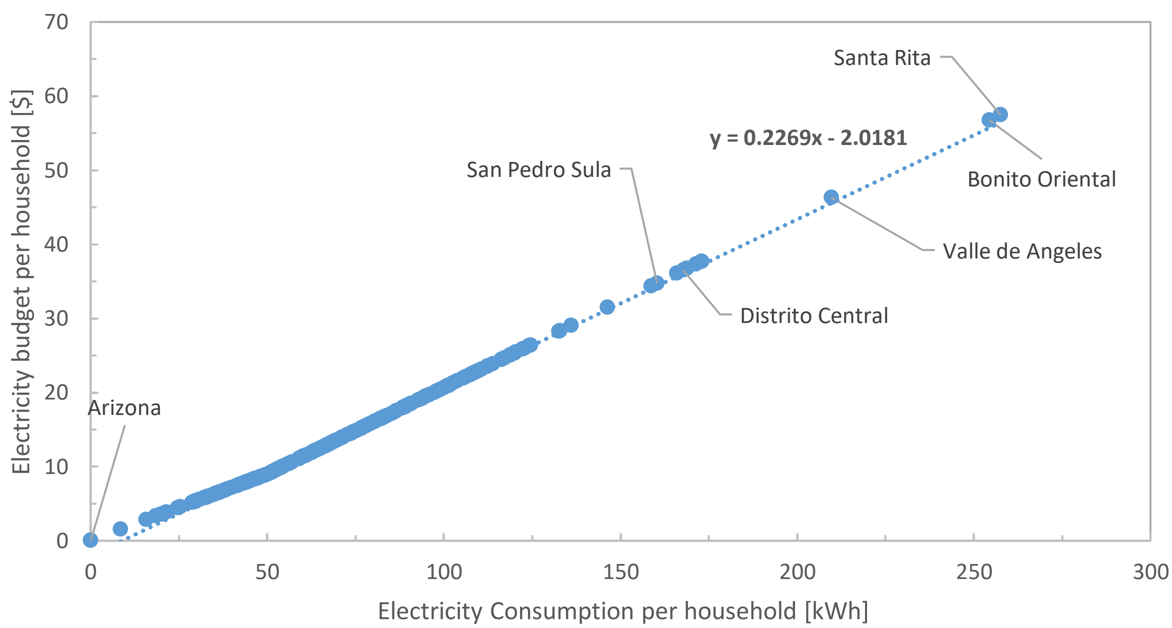


Figure 74: Relationship between electricity consumption and electricity budget

Source: Own elaboration with information obtained via email from (EEH , 2022)

In Arizona, 0.0039 dollars is the budget destined to electricity consumption per household whereas in Distrito Central 36 dollars must be pay every month in average per household. In Santa Rita, Bonito oriental and Valle de Ángeles more than 45 dollars are destined to pay the electricity distribution company.

A great question arrives from this variability of budgets in distinct regions of Honduras, how much of the salary or income of a family represents the electrical bill.

According to (Cámara de Comercio de Cortés, 2021) the minimum wage varies depending on the sector of the economy in which one works. If an inhabitant performs in activities such as agriculture or hunting and fishing the minimum wage is L. 7,033.88, on the other hand, if an inhabitant works in financial establishments and real estate the minimum wage goes up to L.13,346.47.

Worst-case scenario with a wage of L. 7,033.88 in San Pedro Sula or Distrito Central, almost the 15% of the salary of an inhabitant (assuming just one member of the family works in the household) is destined to the electrical services budget. In the best-case scenario, with a wage of L.13,346.47 only 6% to 7% is destined to the electricity consumption payment.

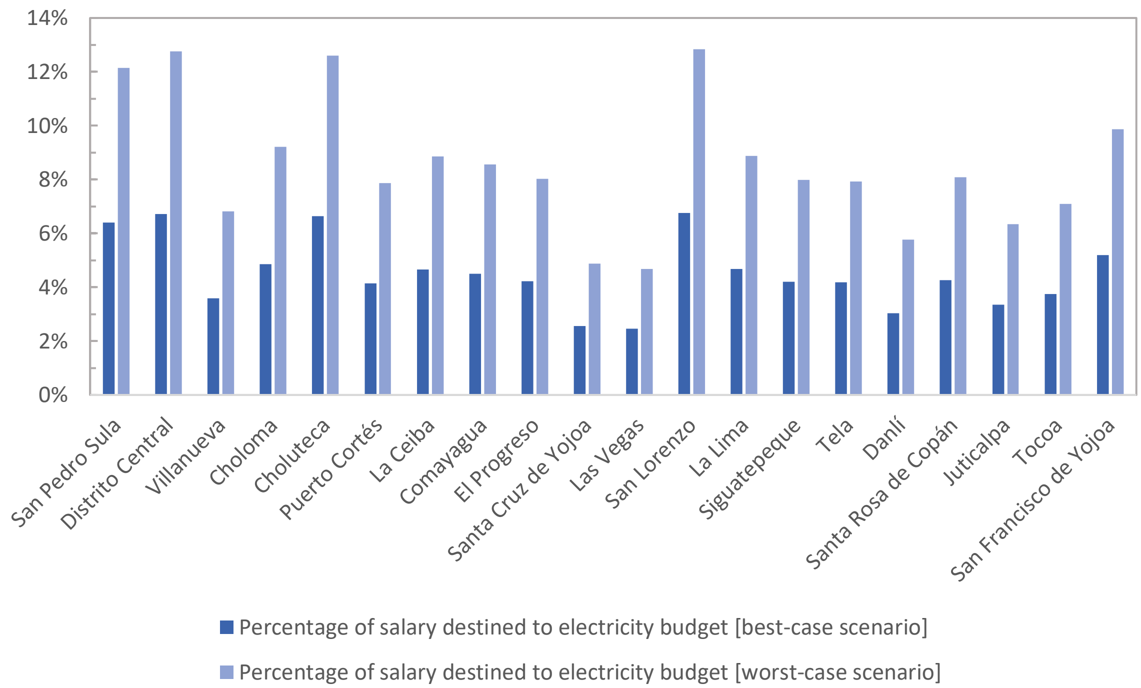


Figure 75: Comparison of best-case and worst-case scenario based on electricity budget

Source: Own elaboration with information obtained via email from (EEH , 2022)

The percentage of salary destined to the electricity consumption payment can be found in Annex 7 for each municipality.

It is crucial to mention that 58% of jobs in Honduras are informal and no one guarantees the minimum wage under these conditions (Banco Mundial, 2020). The poverty level reached in 2020, 70% of the population in Honduras, meaning that this percentage of the population cannot even afford to meet their basic needs let alone electricity service.

Even though the worst-case scenario in this investigation is presented based on the minimum wage, there are worst scenarios for Honduran families that live in poverty and work informal jobs.

5.5.2 COMMERCIAL SECTOR

As well as it was done for the residential sector, a regression model will be obtained for commercial sector and an analysis of the consumption will be done. In Annex 5 the table where commercial electricity consumption is tabulated for each one of the 259 municipalities is found this information was obtained from EEH in 2022.

a) Multiple regression model for commercial consumption of electricity

For the multiple regression model for commercial consumption the same methodology explained in section 5.4.1 was applied.

First, eight independent variables were considered to train the model in RStudio, the dependent variable to be predicted is the electricity consumption in the commercial sector.

Table 13: Variables considered for commercial consumption model

Characteristics	**Variable assigned**
Commercial Electricity Consumption [kWh]	Y
Inhabitants working on wholesale and retail trade, repair of motor vehicles and motorcycles	X_1
Inhabitants working on transport and storage	X_2
Inhabitants working on Lodging and catering (hotel and restaurant)	X_3
Inhabitants working on information and communications	X_4
Inhabitants working on financial and insurance	X_5
Inhabitants working on real estate	X_6

Inhabitants working in professional, scientific, and technical activities	X_7
Inhabitants working on administrative and support services activities	X_8

Source: Own elaboration with information obtained from (INE, 2013)

All the variables named in Table 47 are part of the variable "number of inhabitants working per economical sector [cases]" presented in 5.3. These variables were obtained from the INE (2013) and each one represents the value of inhabitants that work in the distinct businesses of the sector. The values for these eight variables can be found in Annex 8.

For this thesis, it is assumed that a greater quantity of inhabitants working in certain type of business represents that there is a greater quantity of this business in the municipality. This asseveration is made due to the lack of data in the country. The ideal scenario would be to explore the number of businesses of a certain type in the commercial sector and to determined variables based on this quantity (not in the quantity of inhabitants working in the business).

The regression model to determine a function of commercial consumption where there are eight independent variables is obtained through RStudio, as well as the residential regression model and the general regression model. The subsequent results are obtained:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     -49249      24618  -2.000 0.046529 *
x1            50821456    4988409  10.188  < 2e-16 ***
x2            -9126702    3936045  -2.319 0.021216 *
x3            32547836    7339944   4.434 1.38e-05 ***
x4          -138778293    7034686 -19.728  < 2e-16 ***
x5            -9081941   15572244  -0.583 0.560275
x6           -32036062    8806426  -3.638 0.000334 ***
x7           140624384   13331372  10.548  < 2e-16 ***
x8             4180194   25452708   0.164 0.869680
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 350600 on 250 degrees of freedom
Multiple R-squared:  0.9905,    Adjusted R-squared:  0.9902
F-statistic:  3249 on 8 and 250 DF,  p-value: < 2.2e-16
```

Figure 76: First simulation of regression model for electricity consumption in the commercial sector

Source: Own elaboration through RStudio

For the variables x_8 and x_5 the p-value is not acceptable in Figure 146, therefore they are discarded from the regression model.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -48609      24519  -1.983   0.0485 *  
x1            51213540    4558613  11.234  < 2e-16 ***
x2            -8696546    3416993  -2.545   0.0115 *  
x3            31818783    6179289   5.149 5.27e-07 ***
x4          -140311194    5172707 -27.125  < 2e-16 ***
x6           -31455581    7272385  -4.325 2.20e-05 ***
x7           133748031    7492898  17.850  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 349400 on 252 degrees of freedom
Multiple R-squared:  0.9905,    Adjusted R-squared:  0.9902
F-statistic:  4359 on 6 and 252 DF,  p-value: < 2.2e-16
```

Figure 77: Final regression model for commercial electricity consumption

Source: Own elaboration through RStudio

For the model shown in Figure 146 and 147 the independent variables were normalized being divided by the maximum value of the row. With the purpose of obtaining representative slope coefficients and understand which independent variables has greater influence in the final consumption.

In Figure 147, it can be distinguished that X_7, X_1 and X_3, in the respective order have estimated coefficients of 133,748,031 and 31,818,783 and 51,213,540. Having businesses that work in scientific areas the greatest influence in the consumption of electrical energy.

Obtaining the final equation that describes the behavior of the demand:

$$y = -48609 + 51{,}213{,}540 \cdot x_1 - 8{,}696{,}546 \cdot x_2 + 31{,}818{,}783 x_3 - 140{,}311{,}194 x_4 - 31455581 x_6 + 133748031 x_7$$

Equation 32: Regression model for commercial electricity consumption

Source: Own elaboration through RStudio

b) Commercial consumption analysis

It was already discussed in the subsection a, that X_7, X_1 and X_3 have positive influence in the total value of electricity consumption; this shows that the businesses of wholesale and retail trade, repair of motor vehicles and motorcycles (X_1), hotels and restaurants (X_3) and businesses that work in scientific areas (X_7) have great influence in the final consumption of the commercial sector.

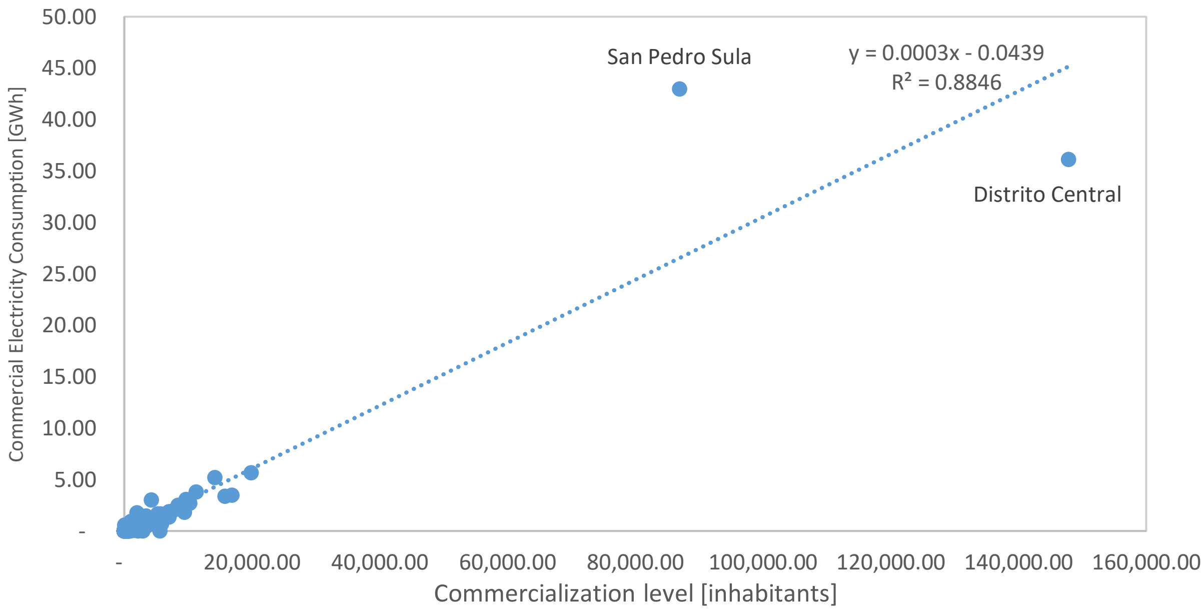


Figure 78: Relationship between inhabitants that work in the commercial sector versus electricity consumption in the commercial sector

Source: Own elaboration from information obtained via email from (EEH , 2022)

A relationship between the sum of the eight variables presented in Table 47 and the consumption of electricity in the sector is shown in Figure 148. The relationship is linear and shows a high coefficient of determination. Which indicates the greater amount of people working in a sector, the greater consumption this sector could have.

In Annex 5, the data of residential consumption per municipality as well as the data of commercial consumption per municipality can be found.

5.5.3 Industrial Sector

As well as it was done for the residential and commercial sector, a regression model will be obtained for the industrial sector and an analysis of the consumption will be done. In Annex 5, the data of residential and commercial consumption per municipality (259) as well as the data of industrial consumption can be found.

a) Multiple regression model for industrial consumption of electricity

For the multiple regression model for industrial consumption the same methodology explained in section 5.4.1 was applied.

First, eight independent variables were considered to train the model in RStudio, the dependent variable to be predicted is the electricity consumption in the industrial sector.

Table 14: Variables considered for industrial consumption model

Characteristics	**Variable assigned**
Commercial Electricity Consumption [kWh]	Y
Inhabitants working on agriculture, forestry, and fishing	X_1
Inhabitants working on mining	X_2
Inhabitants working on manufacturing industries	X_3
Inhabitants working on supplying of electricity, gas, steam, and air conditioning	X_4
Inhabitants working on water supply, sewage disposal, waste management and decontamination	X_5
Inhabitants working on construction	X_6

Source: Own elaboration with information obtained from (INE, 2013)

All the variables named in Table 48 are part of the variable “number of inhabitants working per economical sector [cases]” presented in 5.3. These variables were obtained from the INE (2013) and each one represents the value of inhabitants that work in the distinct businesses of the industrial sector. The values for these six variables can be found in Annex 9.

The regression model to determine a function of industrial consumption where there are six independent variables is obtained through RStudio, as well as the residential regression model and the commercial model. The subsequent results are obtained:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     42656     144787   0.295  0.76853
x1           -2088180    1227651  -1.701  0.09019 .
x2            3509429    1134671   3.093  0.00221 **
x3           68940775    2018421  34.156  < 2e-16 ***
x4           27877627   30557292   0.912  0.36248
x5          -31518876   16054622  -1.963  0.05072 .
x6          -23019596   16113243  -1.429  0.15435
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1616000 on 252 degrees of freedom
Multiple R-squared:  0.8762,    Adjusted R-squared:  0.8733
F-statistic: 297.4 on 6 and 252 DF,  p-value: < 2.2e-16
```

Figure 79: First simulation of regression model for electricity consumption in the industrial sector

Source: Own elaboration through RStudio

The Adjusted R^2 is valid as well as the estimated coefficients. On the contrary, the p-values (marked with the blue box) are not all valid. Those that are not acceptable should be discarded. A fast way to discard variables that are not significant to the model based on the p-value is utilizing the function in R called “ols_step_both_p” which gives as an output a set of variables that work together as a model where all variables are significant.

```
                              Stepwise Selection Summary
-------------------------------------------------------------------------
                      Added/                      Adj.
Step    Variable      Removed     R-Square     R-Square       C(p)
-------------------------------------------------------------------------
   1       x3         addition       0.742        0.741     270.4120
   2       x6         addition       0.869        0.868      13.6110
   3       x2         addition       0.874        0.872       6.2200
-------------------------------------------------------------------------
```

Figure 80: Output of the function ols_step_both_p () for the industrial sector

Source: Own elaboration through RStudio

The significant variables for this model are inhabitants working on mining (X_2), inhabitants working on manufacturing industries (X_3) and inhabitants working on construction (X_6) according to the output of the "ols_step_both_p" function.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -57946     105482  -0.549  0.58325    
x3           69144361    2012525  34.357  < 2e-16 ***
x2            3452642    1131559   3.051  0.00252 ** 
x6          -46304105    2877411 -16.092  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1623000 on 255 degrees of freedom
Multiple R-squared:  0.8737,    Adjusted R-squared:  0.8722 
F-statistic: 587.9 on 3 and 255 DF,  p-value: < 2.2e-16
```

Figure 81: Final simulation of regression model for electricity consumption in the industrial sector

Source: Own elaboration through RStudio

In Figure 149, all the parameters are valid. Obtaining the final equation that describes the behavior of the consumption of industrial electricity:

$$y = -57{,}946 + 69{,}144{,}361x_3 + 3{,}452{,}642x_2 - 4{,}6304{,}105x_6$$

Equation 33: Final regression model to predict industrial electricity consumption

Based on the estimated coefficients, X_3 which represents the inhabitants working on manufacturing industries has the highest impact in the dependent variable because it presents the highest coefficient. The manufacturing industries are followed by the mining activities which also have a positive a big value in the total consumption of the sector.

b) Industrial consumption analysis

The variables that had a positive impact in consumption were manufacturing industries and mining, the sum of the workers that belong to the two sectors is in Figure 150 represented by the x-axis and the electricity consumption of the industrial sector is represented by the y-axis.

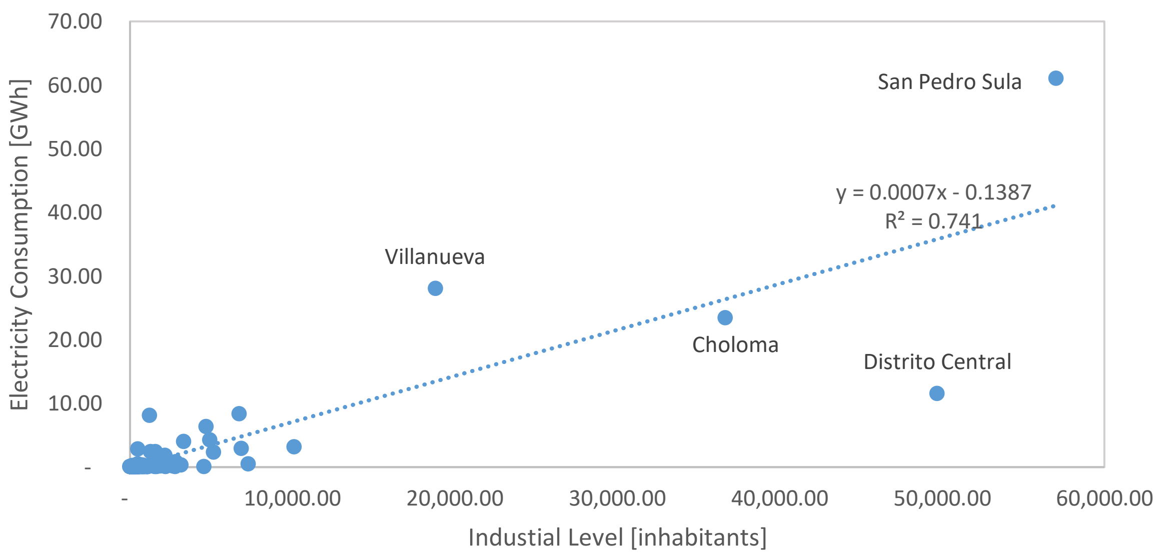


Figure 82:Relationship between inhabitants that work in the industrial sector versus electricity consumption in the industrial sector

Source: Own elaboration from information obtained via email from (EEH , 2022)

VI. Conclusions

A cartographic interactive tool to locate the electricity consumption of March 2022 per each municipality in Honduras was developed in Tableau. Besides, an analysis of the correlation between sociodemographic variables and electricity consumption in Honduras was achieved in Excel, as well as a regression model to acquire electricity consumption per municipality based on significant sociodemographic variable. These processes revealed that the consumption of electricity in Honduras is unequal, only 18 municipalities of the 259 studied represent 80% of the electrical consumption of the country, i.e., only 7% of the municipalities consumed 80% of the electrical resources. A little less than half of the population uses 80% of the electricity, while the other half uses only 20%. Deeper analysis of the electricity consumption in Honduras revealed the following results:

1. The highest consumers of electricity are San Pedro Sula, Central District, Villanueva and Choloma. Per capita electricity consumption is led by Las Vegas, Sabá, Villanueva and Colón. Electricity consumption per square meter is led by San Pedro Sula, Villanueva and Choloma. The municipality of Sabá in Colón, the municipality of La Unión in Copán and the municipality of Las Vegas in Santa Bárbara stand out as exceptional cases in this investigation. This is due to its high electricity consumption capita and per area, however, it has been distinguished that these high rates are due to the industrial sector and are not related to the population or the area.

2. A pattern is recognized when analyzing departmental consumption, in half of the departments there is a municipality inside of the department that accounts for more than 50% of the departmental consumption. For example, in Cortés, San Pedro Sula represents 55% of the electricity consumption of the department.

3. The lack of municipalities that show dark color in the interactive choropleth map tool in Tableau to track monthly electricity consumption, proves the inequality in the electricity consumption in Honduras.

4. When analyzing the variables proposed for the multiple regression model, it is concluded that only income per capita (x4), industrialization level (x5), electrical coverage (x8) and urbanization level (x2) are significant when determining electrical energy consumption.

5. The residential sector in Honduras is the one that consumes the most electricity in the country. Santa Rita is recognized as the department in which the greater consumption per household is achieved. In the commercial sector, the businesses that refer to the scientific area are the ones that consume the most. Whereas, in the industrial area the manufacturing industries have the greatest impact in the total consumption of the sector.

The main limitation of this work is the lack of data of monthly consumption, only the consumption for March 2022 is presented in the Tableau tool. The data of other months of 2022 exists but it is not of public access. The author explored only the consumption for March 2022 provided by the electrical utility company (EEH) via email because of the time limitation to achieve results. The consideration of more data in the data visualization tool, normalization of independent variables and use a logarithmic function to obtain only positive values as electricity consumption could be an objective for fellow researchers.

Despite this limitation, we believe that this work can serve as a guideline for further investigations. Other investigators could contact EEH to map the monthly consumption for the whole year of 2022 or even electricity consumption of ten year back to explore tendencies.

VII. RECOMMENDATIONS

- Explore the social and economic conditions of the outliers such as the municipalities of Sabá, La Vegas and La Unión to recognize the cause of the deviation.
- Opt to decentralize industries (having multiple facilities that cover large areas distributed close to customers) to improve the distribution of electricity demand in the national territory.
- Accelerate the industrialization level and urbanization level in various municipalities to increase the electricity consumption of the region and with it increase the human development index.
- Improve the actual legislation in concern to the topic of distributed generation to activate investment in residential and commercial sector to install distributed energy systems such as photovoltaic technology.

VIII. APPLICABILITY

The data visualization tool in Tableau can be applied so that the public in general, but above all, decision makers can inspect the consumption of their municipality and, with respect to it, make energy planning decisions focused on the decentralization of energy generation.

The consumption per sector is available in the geospatial information tool in Tableau. When the draft of law: "Law for the Rational and Efficient Use of Energy" becomes public in La Gaceta the top priority must be that the sector that consumes the most in each municipality must be the one that makes the greatest efforts towards energy efficiency.

The regression model studied reveals which variables are significant to determined electricity consumption. Knowing that income per capita, industrialization level, electrical coverage and urbanization level are the main variables when studying electricity, municipalities could try to increase these variables and therefore increase electricity consumption in the region which might bring human and economic development to the municipal territory.

IX. REFERENCES

Banco Mundial. (2020). Retrieved from Los empleos informales y la baja participación de mujeres persisten en el mercado laboral de Honduras: https://www.bancomundial.org/es/news/press-release/2020/02/17/diagnostico-del-trabajo-en-honduras#:~:text=El%20estudio%20hall%C3%B3%20que%20un,y%20la%20industria%20(11%20%25).

Empresa Nacional de Energía Eléctrica. (2020). *PROYECCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA DELSISTEMA INTERCONECTADO NACIONAL DE HONDURAS (2021-2035)*. Retrieved from http://www.enee.hn/Subgerencia%20planificacion/2021/DCTO%20_PROY_DEM_ELECT_2021_2035_TRES_ESCENARIOS%20FINAL%20(09-11-20).pdf

Santamaria et al . (2014). Regulation on Distributed Generation: Latin American Case. doi:10.1109/tdc-la.2014.6955274

United Nations Development Programme. (2007). *Human Report Development 2007-2008*. New York: Palgrave Macmillan. Retrieved from https://hdr.undp.org/system/files/documents//human-development-report-20072008-english.2008-english

(UNAM), U. A. (2019). Retrieved from Relación y Correlación lineal simple: http://asesorias.cuautitlan2.unam.mx/Laboratoriovirtualdeestadistica/CARPETA%203%20INFERENCIA_ESTADISTICA/DOC_%20INFERENCIA/TEMA%204/09%20REGRESION%20Y%20CORRELACION%20LINEAL%20SIMPLE.pdf

Ackermann, T. A. (2001). Distributed generation: a definition. *ELSEVIER*, 10. doi:10.1016/s0378-7796(01)00101-8

Arto, I., Capellán-Pérez, I., Lago, R., Bueno, G., & Bermejo, R. (2016). The energy requirements of a developed world. *Energy for Sustainable Development*. doi:https://doi.org/10.1016/j.esd.2016.04.001

Balamurugan, K. S. (2012). Impact of Distributed Generation on Power Distribution Systems. *Energy Procedia*. doi:10.1016/j.egypro.2012.07.013

Bhatia, S. (2014). *Advanced Renewable Energy Systems Part 1*. Retrieved from https://doi.org/10.1201/b18242

Cámara de Comercio de Cortés. (2021). Retrieved from https://www.ccichonduras.org/spanish/comunicados/2021/Junio/TABLA%20AJUSTE%20SALARIO%20MINIMO.pdf

CREE. (2022). Retrieved from https://www.cree.gob.hn/tarifas-vigentes-enee/

Dirección General de Eléctricidad y Mercados. (2019). *Informe Estadístico Anual del Subsector Eléctrico*. Retrieved from https://portalunico.iaip.gob.hn/portal/ver_documento.php?uid=ODQ1NzYwODkzNDc2MzQ4NzEyNDYxOTg3MjM0Mg==

Dye, S. (2020, February 12). *Quantile Regression*. Retrieved from Towards Data Science: https://towardsdatascience.com/quantile-regression-ff2343c4a03

EEH . (2022). *Electrical Consumption for March 2022*.

Empresa Nacional de Energía Eléctrica. (2019). *COBERTURA DEL SERVICIO DE ENERGÍA ELÉCTRICA EN HONDURAS*.

Esteves, G., Bastos, B., Cyrino, F., Calilid, R., & Souza, R. (2015). Long Term Electricity Forecast: A Systematic Review. *ELSEVIER*, 10. doi:10.1016/j.procs.2015.07.041

FHIA. (2019). Retrieved from http://www.fhia.org.hn/html/index.html

Figueroa Rivera, M. Á. (2012). Crecimiento de la generación distribuida en Honduras. In *El regulador ante los nuevos desafíos de la energía en Iberoamérica* (pp. 119-124). Retrieved from https://catalogobiblioteca.cnmc.es/ARTI/BRARTI100002597/BRARTI100002597_G1/09.generacion_distribuida_figueroa.PDF

Flores, W. (2021). Energy System Observatory of Honduras. doi:10.1109/EPEC52095.2021.9621596

Fromm, R., & Perez, J. (Abril de 2009). *Fundación Hondureña de Investigación Agrícola (FHIA)*. Obtenido de www.fhia.org.hn/descargas/microhidrocentrales/guia_microcentrales.pdf

Fumo et al. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews (ELSEVIER)*, 332-343. doi:10.1016/j.rser.2015.03.035

Gobierno de la república de Honduras. (2019). *Shapefile de la division politica municipal de Honduras*. Retrieved from https://datos.gob.hn/en/dataset/limites-municipales/resource/0a086bb7-1e3e-4cb0-b7d8-e21a8417dc5f

Haerer, D. (2015). Employment trends in the U.S. Electricity Sector, 2008–2012. *Energy policy*.

INE. (2013). Retrieved from http://181.115.7.199/binhnd/RpWebEngine.exe/Portal?BASE=CPVHND2013NAC&lang=ESP

Jacobo Campo, V. S. (2013). *THE RELATIONSHIP BETWEEN ENERGY CONSUMPTION AND GDP: EVIDENCE FROM A PANEL OF 10 LATIN AMERICAN COUNTRIES*. Retrieved from https://www.scielo.cl/pdf/laje/v50n2/art04.pdf

K. Balamurugana, D. S. (2011). Impact of Distributed Generation on Power Distribution. *PV Asia Pacific Conference 2011*. Retrieved from https://core.ac.uk/download/pdf/82039853.pdf

Leung, C. S., & Meisen, P. (2005, July). *How electricity consumption affects social and economic development by comparing low, medium and high human development countries*. Retrieved from http://www.geni.org/globalenergy/issues/global/qualityoflife/HDI-vs-Electricity-Consumption-2005-07-18.pdf

Mazur, A. (2011). Does increasing energy or electricity consumption improve quality of life in industrial nations? *Energy Policy*, 5. doi:https://doi.org/10.1016/j.enpol.2011.02.024

Medina, E. (2011). *Factores determinantes de la demanda eléctrica de los hogares en españa: una aproximación mediante regresión cuantílica.* Valladolid, España.

Mohamed, Z., & Bodger, P. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy*. doi:https://doi.org/10.1016/j.energy.2004.08.012

Moral-Carcedo et al, J. (2005). *Modelling the non-linear response of Spanish electricity demand to temperature variations.* doi:10.1016/j.eneco.2005.01.003

Munasinghe, M. (1990). *INTEGRATED NATIONAL ENERGY PLANNING (INEP) IN DEVELOPING COUNTRIES.* Retrieved from 10.1016/b978-0-408-05634-2.50012-2

Munguía Deras, G. (2022, March). *Honduras Electricity Demand Observation Map v1*. Retrieved from https://tabsoft.co/3KpyjMZ

Ng, A. (2012). *Supervised Machine Learning: Regression ad Classification.* Retrieved from Coursera: https://www.coursera.org/learn/machine-learning/home/info

OLADE. (2017). *Energy Planning Manual 2017.* Retrieved from https://biblioteca.olade.org/opac-tmpl/Documentos/old0379.pdf

Operador del Sistema. (2022). *Plan Indicativo de Expansión de la Generación 2022-2031*. Tegucigalpa.

Pasternak, A. D. (2000). *Global Energy Futures and Human Development: A Framework for Analysis.* Retrieved from http://www.geni.org/globalenergy/issues/global/qualityoflife/HDI-and-electricity-consumption.pdf

Pineda, G. (2022). *Observatorio de la Energía UNITEC HN*. Retrieved from https://public.tableau.com/app/profile/gp17447/viz/EstadsticasElectricidadHonduras/EstadsticasENEE

Quintero, J. P. (2008). GENERACIÓN DISTRIBUIDA: DEMOCRATIZACIÓN DE LA ENERGÍA ELÉCTRICA. *Criterio Libre* .

Quintero, P. (2008). Distributed generation: democratization of electrical energy. *Criterio Libre*.

Reza Torres, E. (2015, July). *Modelado del consumo de energía eléctrica residencial respecto de las dinámicas sociodemográficas en los municipios del Estado de Jalisco utilizando la técnica de regresión lineal múltiple.* Retrieved from https://rei.iteso.mx/handle/11117/3081

Richardson, R. (2015). *Business Applications of Multiple Regression.* 222 East 46th Street, New York, NY 10017: Business Expert Press.

Rodriguez , R., Yao, Y., & SAS Institute Inc. (2017). *Five Things You Should Know about Quantile Regression*. Retrieved from https://support.sas.com/resources/papers/proceedings17/SAS0525-2017.pdf

Santamaría et al, F. (2016). A proposed index to evaluate the state of legislation fostering distributed generation in Latin America and the Caribbean. doi:10.1016/j.tej.2016.02.007

Saxum. (2012). Retrieved from https://www.usofenergy.com/overview/

Secretaría de Energía. (2020). *Balance Energético Nacional 2020.* Retrieved from file:///C:/Users/50498/Downloads/BEN2020%20(1).pdf

Secretaría de Gobernación, Justicia y Descentralización. (2020). *Categorización Municipal*. Retrieved from https://www.sgjd.gob.hn/biblioteca-virtual/sgd/categorizacion-municipal

Secretaría de Gobernación, Justicia y Descentralización. (2020). *Categorización Municipal 1992 - 2020*. Retrieved from (Secretaría de Gobernación, Justicia y Descentralización, 2020)

Torres, E. R. (2015). *Modelado del consumo de energía eléctrica residencial respecto de las dinámicas sociodemográficas en los municipios del Estado de Jalisco utilizando la técnica de regresión lineal múltiple.*

Vezzoli, C. (2018). *Designing Sustainable Energy for All.*

Villatoro Flores, H. F. (2015). Decentralised electricity generation system based on local renewable energy sources in the Honduran rural residential sector. doi:10.1007/s10098-015-1067-x

W. C. Flores et al. (2021). Energy System Observatory of Honduras. *IEEE Electrical Power and Energy Conference*. doi:10.1109/EPEC52095.2021.9621596

Walpole et al. (2012). *Probability and Statistics for Engineers and Scientist.* Pearson/Prentice Hall.

World Bank Open Data. (2020). *Honduras - Place Explorer - Data Commons*. Retrieved from https://datacommons.org/place/country/HND?utm_medium=explore&mprop=amount&popt=Consumption&cpv=consumedThing%2CEnergy&hl=es