Graduation Project

Presentado a

UNIVERSIDAD DE LOS ANDES
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

To obtain the degree of

ELECTRICAL ENGINEER

by

Sebastian Celis Sierra

Big Data Analysis of Power Consumption in the City of Bogota

Sustained the December 13th of 2017 in front of the jury:

- *Advisor:* Luis Felipe Giraldo PhD,Assistant Professor, Universidad de Los Andes
- *CoAdvisor:* Paulo Manuel de Oliveira PhD,Associate Professor, Universidad de Los Andes
- *Evaluation Committee:* Fernando Lozano PhD,Associate Professor, Universidad de Los Andes

Dedicated to Marcelo Sierra.

Acknowledgments

Thanks to Codensa for sharing the data used during the development of this project.
Thanks to Luis Felipe Giraldo for his constant help in this project; his motivation and charisma was useful to finish this work.
Thanks to Paulo Manuel for his assistance during the project, his valuable advice helped focusing the project.
Thanks to my mother and my father for their support.

Contents

*

List of Figures

List of Tables

Abstract

In the present work, a package for R language was developed for clustering data given by smart meters installed across the city of Bogota, by Codensa. Using statistical techniques, it was possible to characterize the consumption of different populations according to their socioeconomic status and locality where they are located;this in order to reveal patterns of consumption among the city, testing the value of the installation of the new 6060 meters located at 11 different localities.

1. Introduction

1.1 Problem Statement

Electrical efficiency is a main topic in the energy policy in the entire world. The main objectives are to guarantee the power flow into the network with the minimum cost and to increase the coverage area of the service. The group of consumers represents a valuable resource to define how the aggregate consumers in a city uses electricity in different periods of the day. Smart Grids incorporates communication technologies in every aspect of:

- Generation of Energy.
- Transmission of Electricity.
- Distribution of Electricity.
- Final User Consumption of Power.

For the case of Bogota, Colombia, with the installation of smart meters, a new problem arose; how to relate the data from the meters to find a correlation between the time of use of electricity and the use of the transportation system.

1.2 Justification

Since the second half 2016, Codensa has started the project of the installation 40,000 smart meters in different locations of Bogota (Puente Aranda, Engativá, Fontibón y Suba), with the purpose of metering the consumers consumption of power in a 24-hour period. The study should find a relation between the amount of energy consumed in specific regions depending of the hour of the day. The number of hours that a consumer spends every day transporting itself is closely linked to the electrical Consumption behavior of the user. This work will find any correlation between the displacement of power Consumption in different regions of the city and the transport time that consumers should spend in other to reach their work or study locations, following this threat the following hypothesis is made: Power peaks in the load curves are more significant in certain locations depending on the time of day. In the morning the consumption will be high in areas of the periphery of the city, with the passage of the hours the curve of power consumption will be greater in the center of Bogota's, finally during the afternoon the power consumption will increase again in the periphery. Previous behavior is linked to transportation of consumers that live far away from their work centers that might be in the center.

1.3 Solution Approach

Using clustering methods in conjunction with traditional statistics methods, a proper study of energy consumption will be done.The energy consumed in the different areas of the city will be evaluated, making a discrimination according to the socioeconomic stratum of the evaluated population.

2. Objectives

2.1 General Objective

Do a clustering study of data from the smart meters installed in Bogota, in order to provide information about the time of use of electricity at various locations of the city. This analysis will reveal patterns of consumption for each group studied, allowing detecting unique electrical behaviors according to the sector and the stratum where the different populations studied live.

2.2 Specific Objectives

1. Validate the hypothesis related to the displacement of the peaks of load curves in specific regions of Bogota during a 24-hour period. The use of the library "ggmap" from the R environment will show in the Bogota city map the most relevant zones of power consumption during a specific hour of the day.
2. Identify the pattern of consumers belonging to a specific area and a specific stratum, with the intention of creating clusters of consumers. This clustering should be made with two different approach:
   (a) Clustering consumers depending their socioeconomic stratum.
   (b) Clustering consumers depending their residence location.
3. Analyze the power consumption of consumers in the interest of detecting anomalous behaviors that might affect the system.
4. Relate each location of study with a power consumption pattern.
5. Propose suggestion of use of smart meters to **CODENSA**.
6. Develop all the software resources necessary to study the data of smart meters and upload them as a free software for anyone interested in the use of it.

3. State of Art

In general, in Colombia there are few studies at a national level aimed at determining the viability of installing smart meters in the Colombian electrical network , which for the present project, will be useful for evaluation and comparison:

1. The study made in 2016 by Eng. Maria Angelica Arroyo at Universidad de los Andes was focused in the implementation of Smart Meters in Cali Colombia ;EMCALI (Empresas Municipales de Cali) has implemented Smart Meters in different parts of Cali, Colombia, this data has been analyzed in order to study the power consumption of strata 1,2 and 3[10]. he previous work done at **Universidad de los Andes**, grouped the power consumption behavior of consumers from stratus 1 and 2 from the city of Cali. Data filtering was applied to clean the data set that was used to obtain the characteristic curves of consumption of each population. This work recognize the importance of measuring the behavior of a group of consumers this in favor of energy saving, exploring different aspects that might be used in the future, using new pricing methodologies[10].

2. The study made in 2013 by Msc Victoria Eugenia Perez in Cali Colombia, this one gave a more technical understating of the incorporation of energy measurement systems in the network, exploring its installation and economical implications. This work explores the technical and economical factors that imply the incorporation of smart meter's into the whole Colombian electrical system, exploring the possibility of installing Communication Equipment in order to measure stratum 1,2 and 3 of the city Cali, covering a large number of substations throughout the urban geography.

3. The study made by Msc. Juan Felipe Valencia in 2014 [8], focused in the application of Big Data as a generator of valuable information in all sectors and especially in the electricity markets. This work contains a bibliographic reference of international experiences of the management of large data of smart meters, exploring the opportunities and use of this systems in residential electricity consumption measurement in Colombia.

4. The study made in 2016 by Jenny Paola Lozano and William Camilo Guzman at **Universidad Distrital Francisco José de Caldas** [7] studied the demand for electricity according to consumption habits presented in the urban residential population of Bogota, by identifying trends and patterns of consumpton from the users of electricity.

4. Theoretical Frame

4.1 Socioeconomic Stratum in Colombia

In Colombia, the houses are classified depending on the physical characteristics of the same and their urban or rural environment.The strata in which houses or properties can be classified are 7 (from 0),the classification in any of the seven strata is an approximation to the hierarchical socioeconomic difference. This classification is made in order, to make a differential charging of home public services allowing assigning subsidies and collecting contributions. The DANE (Departamento Nacional de Estadística) established " Socioeconomic stratification is the mechanism that allows classifying the population in different strata or groups of persons with similar social and economic characteristics, through the examination of the physical characteristics of their homes, the immediate environment and the urban or rural context of the same" [4], governed by the terms of Law 505 of 1999 that established the methodology for stratification:

1. Only homes should be stratified.
2. Stratification is the form to separate rural areas and farms from disperse houses at urban centers.
3. Stratification is mandatory.

4.2 Smart Grids

The traditional power grid consists of power plants that generate bulk electric power, transmission substations collocated at generation plants step up the voltage levels for high-voltage transmission lines which carry electric power over long distances with high efficiency [1]. A transmission system is deployed to carry power closer to the consumers. Before the power is delivered to the consumers, voltage levels are reduced at distribution substations that transfer power to the consumers over feeders. Communication network technology introduced in the latter part of the twentieth century supported the deployment of Supervisory Control and Data Acquisition (SCADA) systems. These SCADA system allowed operations personnel to remotely monitor and control transmission and distribution substation equipment from utility operations centers, enhancing operational efficiency. The need for clean energy with large-scale deployment of renewable sources of energy, peak power reduction for environmental and economic reasons, grid modernization, and consumer participation in energy management are some of the motivations for the development of the Smart Grids [1].

As the Smart Grid evolution continues, many new grid elements and functions will be integrated into the system. Examples including the following:

- Renewable and other alternate sources of energy will be deployed throughout the grid.
- Advanced Metering Infrastructure, also known as “smart meters”, will be deployed at consumer locations. In addition to measuring consumption, smart meters’ measures voltages, power, reactive power, and other quantities.
- SCADA connectivity will be extended beyond substations to support the monitoring and control of reclosers, capacitor banks, and other elements in the distribution grid.

One of the main elements that compose Smart Grids are Smart Meters. These electronic devices are capable of recording electricity consumption of a user, giving this, one can made analysis in order to identify the behavior of the studied population.

4.3 Smart Meters

Smart meters are electronic devices capable of recording the consumption of electric energy in intervals of hours or less, enabling two-way communication between the meter and the central system [1]. Analytics based on data reported by smart meters can contributed to a wide variety of applications such as demand response, distribution management, asset management, and consumer energy management.
The traditional electric meters at consumer locations are being replaced by smart meters. Periodic measurements of electric quantities collected by the meter are used to support many utility operations and business functions such as periodic monitoring energy consumption, voltage, power, distribution of pricing information, and other consumer-centric functions in addition to billing the consumers.
Smart meters provide periodical "interval measurements". For meaningful operation of many distribution management functions such as demand response and volt, VAR, watt control (VVWC), interval measurements are required once every hour on once every 15 minutes or even at the rate of once every 5 minutes [1]. Depending on the capabilities of the meter, some or all the following measurements can be reported in every interval:

- Cumulative energy consumption (Wh) up to the end of the interval.
- Instantaneous rms voltage measurement at the end of the interval.
- Instantaneous rms current measurement at the end of the interval.
- Instantaneous power measurement at the end of the interval.
- Average reactive power measurement over the interval.
- Instantaneous power factor measurement at the end of the interval.

Using the huge amount of data obtained from these devices, an electrical company can start using big data and clustering analysis to define the kind of consumption behavior that many groups of consumers have.

4.4 Energy and Power

Energy is the capacity of doing work. Energy is power integrated over time [12]:

$$Energy = Power \times time$$

$$Wh = W \times h$$

Where:

- Wh is the energy.
- W is the power.
- h is the time in hours

Power is the rate at which work is done, or energy is transmitted:

$$Power = \frac{Energy}{Time}$$

Load Factor is a term that is often used to describe a load. It is defined as the ratio of the average demand to the maximum demand; it gives an indication of how well the utility's facilities are being utilized [12].

4.5 Cluster Analysis

Cluster analysis is the task of grouping a set of objects in a group that shares similar characteristics between the objects [2]. Cluster must capture the natural structure of the data to obtain meaningful groups. Cluster analysis has long played a key role in a wide variety of fields [2]:

- Statistics.
- Biology.
- Machine Learning.
- Pattern recognition.
- Data Mining.

Classes share common characteristics, so dividing objects in groups provide an abstraction from individual data objects to the clusters in which those data objects reside. These cluster prototypes can be used as the basis for several data analysis or data processing techniques.
This method was selected because of the need of the study to classify the different consumers of the study without the need of using a really complex algorithm ;other algorithms that use neural networks, Bayesian networks and decision trees can also be applied to solve this problematic.

4.6 Boxplots

To understand each of the data groups of this work, boxplots will be used to represent the dispersion of the same. A Boxplot is a method for graphically depicting groups of numerical data through their quartiles. Boxplots for this proyect will show the mean (using a red dot), the median (using a line), the maximum and minimum reach by each group.

5. Residential Energy Consumption in Colombia

UPME has developed different energetic characterization studies in different sectors such as: residential, transportation, industrial and commerce, establishing the different electrical habits of each group.
The energy consumption is measurement using specialized equipment that takes measurements for a period of time. Characterization of residential energy consumption using the data obtained from the measurements, refers [5] to:

1. Obtaining the mean of energy consumed for a month (W/month).
2. Identifying the final use of the energy.
3. Identifying the electronic equipment that every house has (kind, amount of power consumed, time of use).
4. Geographical location.
5. Identifying if a location is rural or urban.
6. Temperature of the location.

The residential sector is formed by the houses classified in economical strata from 1 to 6 according to their location, city, access roads and public services. For doing a proper characterization of this group, it would be necessary to divide each group depending of the stratum and the geographical region, after this, obtaining the average power consumed should be done. **UPME** has also identified the home appliances most common in Colombia geography , so any proper study should include the effect of the following elements [5]:

1. Illumination.
2. Color Television.
3. Cooler 9 ft.
4. Iron.
5. Blender.
6. Cellphones.
7. Washing Machine.
8. Computer.
9. DVD.
10. Sound System.
11. Fan (for zones under 1000 AMSL)

6. Smart Grids in Colombia

Multiple objectives have been proposed by **UPME** for the Colombian Energetic future for the year 2050, this objectives are the principal motivation for the implementation of smart grids in the whole Colombian geography. The strategical energy objectives proposed are[6]:

1. Universal Access to Electricity: Smart Grids permit to augment the amount of distributed generation using microgrids, in favor of not interconnected zones.
2. A Productive and Efficient Country: Achieve a reliable and efficient energy supply, minimizing the risk of power cuts and ensuring continuity of the supply.
3. A Competitive Country:The implementation of new technologies will improve the country's competitiveness and obtain benefits from the technological, electrical and industrial sectors.
4. An Efficient Country: Promote the efficient management of demand in all productive sectors, and incorporate the implementation of renewable energies through the development of new technologies.

The development of the four objectives, indirectly achieves a fifth objective, called **"Un Pais de Oportunidades/ Progreso Social"**; with the labor of the four objectives, it would be possible to promote the development of social welfare through the promotion and implementation of the set of smart grid technologies.

6.1 Advanced Measurement Infrastructure (AMI)

The application of smart grids need information of the state of the network, the offer and the consumers [6]. The measurement infrastructure, in conjunction with an appropriate communication network, this will provide intelligent networks with the necessary information to make different decisions required for proper operation.
This technology will include all the parts of the systems, from the generation centrals, to the meters located at any client home. The features considered for this technology, according to **UPME** are [6]:

- Reading and Remote Operation: This will contribute to save costs of operation where the displacement of personnel will not be need.
- Limitation of Power Remotely: It allows the operators of the system to reduce operating costs, and consumers to reduce the value of their invoice, since they will be able to make changes in contracted power in an agile manner.
- Detection of meter manipulation and notification to the company: This will be a valuable tool for reduction the amount of non-technical loses.
- Information to the User: Smart Meters will allow the consumer to have information of the electrical behavior of his locality in real time. The user will know their consumption profile, and will be able to calculate the savings that would involve changing their habits.
- Hour Pricing: The user will be allowed to know the price energy for a 24 hour period, bringing him the possibility of changing his habits in order to consume energy at the hours of lower price.
- Measurement of Distributed Generation:The connection to distributed generation in residential areas will require that smart meters have both the capacity to measure incoming and outgoing energy, in this way, consumers will act as consumers and generators.
- Active Load Management: It will be possible to connect or disconnect manageable cargo at the most convenient times according to the demand curve.

7. Methodology

Using the data given by Codensa a code project should be done using the R environment. The software should be capable of filtering the multiple data given, to standardize the data into a form that can be useful for a statistical analysis. After the filtering procedure, a clustering analysis should be done in order to identify the behavior of multiple groups of clients. An outliers removal system should be implemented, in order to eliminate measurements that differ to much with the mean of the centroid. After the clustering analysis is done, one should have done the statistical analysis that will relate the location of a possible client with the way it consumes power during a period of time. For this, the use of mapping tools will create figures in the space of Bogota that will relate the power consumption of each locality with each other; two ways are proposed:

1. Relate the mean consumption of an specific hour of day with other localities of the city.

2. Relate the mean consumption of an specific 24 hours signals with other localities of the city, in which lines will connect each locality if any kind of correlation exist.

The statistical analysis will be focused in identifying the different relations between groups that exists in terms of consumption of energy.
The methodology used in the chronological order to complete the different tasks proposed is exposed in this chapter. The next diagram, will explain in general terms, the methodology used:

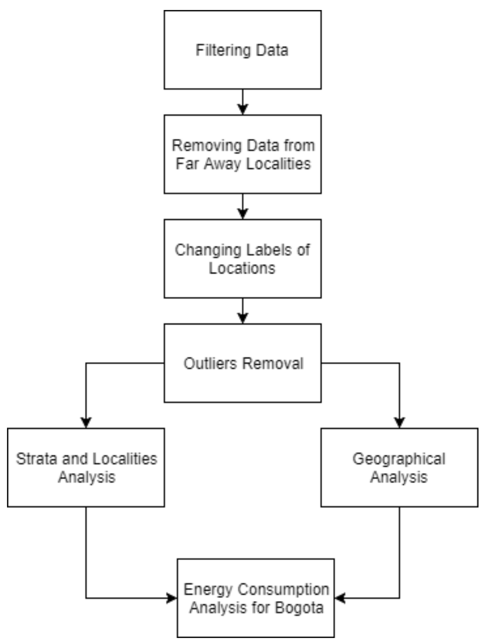

Figure 7.1: Methodology Used

7.1 Libraries Used

At first, for developing the multiples functions that were be used in the project, the following libraries of R were used depending the necessity of the function:

- **matrixStats**: Matrix's statistics.
- **corrgram**: Visualization of data in correlation matrix's.
- **fpc**: Clustering.

- **Hmisc**: Data Analysis.
- **reshape2**: Melt and Case Vectors.
- **plotly**: Data Visualization.
- **ggplot2**: Data Visualization.
- **ggmap**: Maps Creation.
- **leaflet**: Maps Creation.
- **timeDate**: Time and Date Data.
- **lubridate**: Time and Date Data.
- **zoo**: Time Series.
- **cluster**: Clustering.
- **factoextra**: Clustering.
- **NbClust**: Clustering.
- **plyr**: Cominbing Data.
- **reshape**: Melt and Case Vector.
- **GMD**:Generalized Minimum Distance of distributions.
- **htmltools**: Tools for HTML.

The following scheme shows how each library was used:

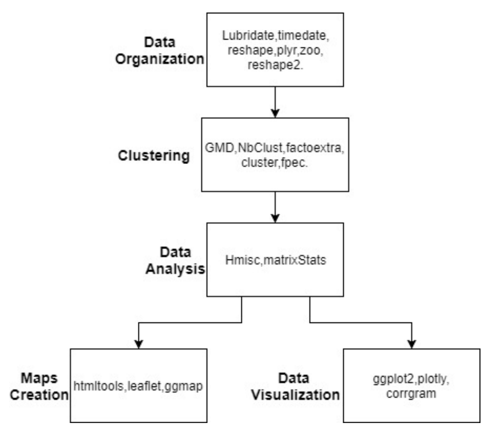

Figure 7.2: Use of each library in the study.

7.2 Functions Created

- Filter Data: Function used to filter the data given. Multiple data comes with mistakes from the measurement, resulting in incomplete or redundant data.
- Clients Month: Used for obtaining the tags of the clients measurement in a month.

- k.elbow: Used for optimizing the number of clusters.
- Cluster Mean: Used for starting a k-means clustering procedure.
- Cluster Organize: Used for organizing the data-frames that will be needed to start the clustering algorithm.
- Separate Localities: Automatically separates a month in individual groups that represents every location that was measure during that period of time.
- Separate strata: Automatically separates a month in individual groups that represents every stratum that was measure during that period of time.
- Weekdays Analysis: K-means analysis for the seven days of the week, separating the data set into groups that represent:
  - Monday.
  - Tuesday.
  - Wednesday.
  - Thursdays.
  - Friday.
  - Saturday.
  - Sunday.
- Map Visualize:Function used to creates maps using leaflet.
- Cluster Images: Used for saving jpg files of the clusters analyzed.
- Cluster Outliers: Used for identifying the clients that represents outliers from the data-set.
- Cluster Outliers Removal: Used for removing all outliers from the k cluster.
- Change Localities: Used for changing all labels of localities that might generate problems at the time of obtaining the coordinates of an specific place.
- Remove Unwanted Locations: Used for removing all data from the following locations:
  1. Rodamontal.
  2. Madrid Cundinamarca.
  3. Zipaquira Veredal.
- Assign Random Coordinates: Used for assigning random coordinates to each client.
- Localities Names: Used for obtaining all locations from a data set.
- Correlation: Used for obtaining the correlation matrix of a system.

7.3 Removing Non-Valid Data

Data from regions that cannot be located in a specific zone of the map should be eliminated, for this specific case, all data from **Rodamontal** was discarded for this reason. Rodamontal, is consider a rural zone near to *Cogua, Cundinamarca* but the specific coordinates of the place are not given. **Madrid Cundinamarca** and **Zipaquira Rural** will not be taken into account because of their remoteness.

7.4 Changing Labels of Locations

In order to use the function *geocode* that is used to extract the coordinates from Google Maps from specific locations, to find the coordinates of the different locations, the labels of **Kennedy** and **Suba** should be change to avoid getting the coordinates of places located in North America. Also, to avoid further mistakes, every location label will be changed in order to specify if it's a locality of Bogota, or a town of Cundinamarca.

7.5 Filtering Algorithm

All data sets for every month are full of incorrect measurements that might be obtained by bad practices and bad planning of the operators of the measurement system. Duplicate measures and incomplete data are the most common fails find in the data set. The following algorithm was implemented in order to clean the data before manipulating it:

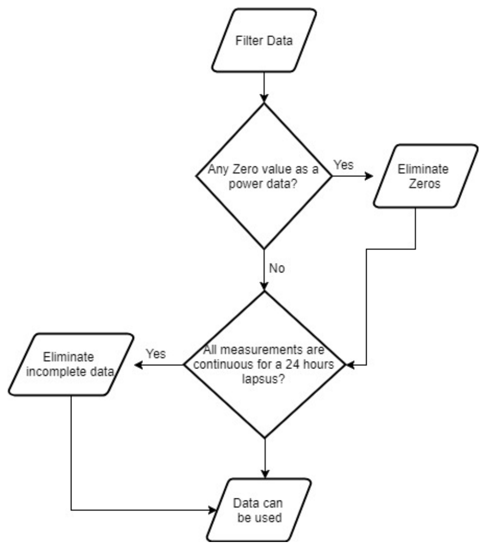

Figure 7.3: Data Filtering

The amount of time that this process requires, may vary depending of the length of the data-set to be clean. For each month, the following amount of clean data is available:

- November 2016: 626160
- December 2016: 904440
- January 2017: 231432
- February 2017: 707712

October data will not be taken into account, the data-set only has 133584 observations.

7.6 Oultiers Removal

The identification of outliers is a fundamental point of this degree project, therefore, specialized functions were developed in the recognition of customers with curves of consumption 24 hours abnormal enough to be considered outliers.It was possible to identify the different consumers that are considered as outliers in a data-set;a 24-hour signal is considered an outlier as long as the distance of the signal compared to the centroid of the cluster is too much. To take an example, for the case of the localities, the complete grouping of 24-hour signals for Fontibon would look like the following:

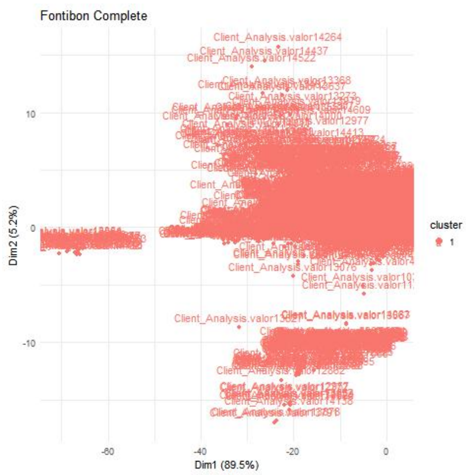

Figure 7.4: Visualization of all Data for Fontibon

In the figure 7.4, it's possible to visualize all outliers that exist in the data-set for the 4 months of study. In order to give to **CODENSA** valuable information of the most problematic clients, the program would identify the tag of the clients that represents outliers, for example, the client that has more curves that represent outliers for the data-set is the one with the tag **1819206**, with 30 curves that represent critical outliers.Each client that represents an outlier, for the Fontibon study case, each client included in the outliers list, will have an average of 10 days of anomalous energy consumption. After deleting the outliers, the form of the data for Fontibon will be:

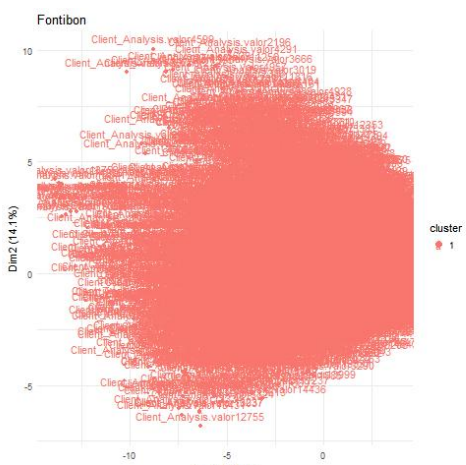

Figure 7.5: Visualization of Fontibon after Outliers Removal

For each locality, the percentage of a total of 165524 hours period signals consider outliers is shown for each locality:

Table 7.1: Percentage of Outliers for the whole data set for each Locality

Locality	**# Outliers**
Puente Aranda	14.84%
Suba	27.81%
Engativa	35.42%
Fontibon	10.95%
Cogua	4.31%
Kennedy	0.1691%
Cogua Veredal	1.66%
Zipaquira	4.28%
Rincon Santo	0.2718%
San Isidro	0.211%
El Olivo	0.0543%

By not having an equal amount of data by location, it will not be possible to reach a conclusion about which location has the highest number of customers with an anomalous behavior. Suba, Puente Aranda, Engativa and Fontibon, are the localities with the greatest amount of data, this will explain the great amount of outliers found. The outliers removal technique, will be applied for all data-sets studied in this project; this methodology will be proposed, in order to provide CODENSA with more information about consumers that do not adapt to the power curve proposed for each stratum and specific location.
For the case of the strata, the following table represents the percentage of the total amount outliers (20186) for the 4 months period for each stratum:

Table 7.2: Percentage of Outliers for the whole data set for each Stratum

Stratum	**# Outliers**
S0	33%
S1	0.96%
S2	11.08%
S3	50.6%
S4	1.42%
S5	0.04%
S6	2.90%

As in the previous case, the data for each stratum are not the same, so we can not say which economic stratum has the highest number of consumers with undue behavior.Stratum 3, having the most data, will be the one with the highest number of outliers.

7.7 Elbow Method

The oldest method for determining the true number of clusters that exists in a data set is called the elbow visual method. The idea is to start with 2 centroids ($K = 2$), and keep increasing it in each step by 1, calculating the clusters and the cost that comes with the training. At some value for K, the cost drops dramatically, this is K value desired[9].

7.8 K-Means Algorithm

K- means is one of the simplest unsupervised learning algorithm that solve the clustering problem, following a simple way to classify a given data set through a certain number of clusters[9].

The algorithm is based in defining the k centroids of each cluster. All the centroids will be placed as far as possible from each other; the next step is to associate each data point with the nearest centroid. When every point is associated with a center, the next step will be to re-calculate k new centroids as centers of the clusters of the previous step. The loop will continue until centroids do not move any more. Finally, the aim of this algorithm is to minimize the square error function. The objective function is:

$$W(S,C) = \sum_{k=1}^{K} \sum_{i \in S_k} ||y_i - c_k||^2$$

Where S is a K cluster partition of the entity set represented by the vectors $y_i(i \in 1)$ in the M-dimensional feature space, consisting of the non-empty non-ovelapping Clusters S_k, each with a centroid $c_k(k = 1, 2, ..., K)$

7.9 Correlation Coefficient

The correlation coefficient of two variables in a data set is equal to their co-variance divided by the product of their individual standard deviations. This result is a normalized measurement that states how the two are linearly related.
The sample correlation coefficient is defined by the following formula, where s_x and s_y are the sample standard deviations, and s_{xy} is the sample co-variance:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Similarly, the population correlation coefficient is defined as follows, σ_x and σ_y are the population standard deviations, and σ_{xy} is the population co-variance:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

7.10 Coefficient of Variation

In statistics, the coefficient of variation also called relative standard deviation, is a standardized measure of dispersion of a probability distribution or frequency. It is often expressed as a percentage, and is defined as the ratio of the standard deviation σ to the mean μ.

7.11 Rate of Change

The rate of change is the speed at which a variable changes over a specific period of time. The Rate of Change can generally be expressed as a ratio between a change in one variable relative to a corresponding change in another. Generally the rate of change for a variable between the points B and A, will be:

$$ROC = \frac{X_B - X_A}{X_A}$$

8. Results and Analysis

Using the data that has been already cleaned and adapted for the requirements of the functions that will be used, multiple results will be obtained using traditional statistical methods combined with new forms of representing data visually. The data was filtered as it was mentioned in the Chapter Methodology in the Section 7.5, where all incomplete and mistaken curves of data were eliminated ; after this, data corresponding to the locations of Rodamontal, Madrid Cundinamarca and Rural Zipaquira was eliminated. The amount of smart meters installed and studied were **6060**, the percentage of meters installed per stratum is the following:

Table 8.1: Percentage of Smart Meters per Stratum

0	7.60%
1	2.44%
2	17.44%
3	67.90%
4	3.844%
5	0.06%
6	0.693%

As shown in the previous table, 67.90% of the data belong to stratum 3; in order to carry out an appropriate characterization of the load in the city, an adequate separation of strata and localities will be carried out, identifying the average energy consumed in a period of 24 hours and the load factor for each case. It is known that the behavior between working days and holidays is different, therefore, knowing Sundays and holidays in Colombia for the dates studied, two large groups of data were classified, Business Days and Holidays. The mean total energy and the mean load factors were found for each specific case, the variance of all this data be visualized in Appendix A

Table 8.2: Total Energy Mean WH for Business Days.

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	9602.34	No Data	No Data	4241.031	No Data	No Data	No Data
Suba	12325.27	No Data	4343.192	3696.298	No Data	3451.306	12437.44
Engativa	15563.57	No Data	No Data	5761.858	No Data	No Data	No Data
Fontibon	33412.53	No Data	No Data	3240.624	4234.496	No Data	No Data
Cogua	7718.571	3890.835	3676.276	3175.858	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	3142.725	No Data	No Data
Cogua Veredal	No Data	1690.857	3329.569	2606.893	No Data	No Data	No Data
Zipaquira	7121.463	3584.52	2929.28	4154.738	10885.25	No Data	No Data
Rincon Santo	No Data	No Data	3040.079	No Data	No Data	No Data	No Data
San Isidro	No Data	2641.4	3587.765	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	2970.571	No Data	No Data	No Data	No Data

Table 8.3: Load Factor Mean for Business Days.

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	0.459	No Data	No Data	0.3871	No Data	No Data	No Data
Suba	0.5323	No Data	0.392494	0.3756	No Data	0.4426	0.5463
Engativa	0.5962	No Data	No Data	0.4242	No Data	No Data	No Data
Fontibon	0.82508	No Data	No Data	0.41186	0.45605	No Data	No Data
Cogua	0.540288	0.32978	0.329786	0.342715	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	0.45774	No Data	No Data
Cogua Veredal	No Data	0.247434	0.316791	0.32395	No Data	No Data	No Data
Zipaquira	0.511934	0.320164	0.363785	0.360274	0.443727	No Data	No Data
Rincon Santo	No Data	No Data	0.271021	No Data	No Data	No Data	No Data
San Isidro	No Data	0.306081	0.274333	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	0.331027	No Data	No Data	No Data	No Data

As evidenced in the above tables, the highest energy consumption corresponds to stratum 0, being particular the case of the Fontibon Locality, where the consumption is extremely high, having the highest load factor.In general, the lower load factors are evident for strata 1, 2 and 3, with a lower consumption compared to other strata. Fontibon was already mentioned as the location with the highest energy consumption, however, the localities of Suba, Engativa and Puente Aranda, also have high consumption rates compared to other localities. Something interesting noted in the classification, was to see how the areas where there is a greater presence of stratum 1 and 2, will be the ones with the lowest energy consumption.

Table 8.4: Total Energy Mean WH for Holidays.

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	6528.318	No Data	No Data	4429.64	No Data	No Data	No Data
Suba	10824.4	No Data	4590.715	3701.585	No Data	4300	10143.25
Engativa	10296.97	No Data	No Data	5536.356	No Data	No Data	No Data
Fontibon	27198	No Data	No Data	3277.058	4106.442	No Data	No Data
Cogua	7250.625	4004.301	3542.559	3175.858	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	2999.8	No Data	No Data
Cogua Veredal	No Data	1652.5	3250.414	2428	No Data	No Data	No Data
Zipaquira	6101.789	3256.205	2785.035	4149.947	6481	No Data	No Data
Rincon Santo	No Data	No Data	2955.707	No Data	No Data	No Data	No Data
San Isidro	No Data	No Data	3650.727	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	2408.783	No Data	No Data	No Data	No Data

Table 8.5: Load Factor Mean for Holidays.

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	0.383149	No Data	No Data	0.401917	No Data	No Data	No Data
Suba	0.58535	No Data	0.400532	0.38764	No Data	0.417923	0.528537
Engativa	0.811107	No Data	No Data	0.424092	No Data	No Data	No Data
Fontibon	0.812531	No Data	No Data	0.418921	0.455122	No Data	No Data
Cogua	0.579233	0.342405	0.338681	0.343277	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	0.510578	No Data	No Data
Cogua Veredal	No Data	0.39003	0.319344	0.235861	No Data	No Data	No Data
Zipaquira	0.492683	0.336608	0.374358	0.360044	0.42149	No Data	No Data
Rincon Santo	No Data	No Data	0.31088	No Data	No Data	No Data	No Data
San Isidro	No Data	No Data	0.311084	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	0.311331	No Data	No Data	No Data	No Data

For the holidays, consumption patterns similar to working days are evident, in some cases, the average of electric power used in a 24-hour period is even higher compared to working days. However, the general trend of these days, is a decrease in the burden on households, evidencing load factors a little lower than during the working days.
In order to validate the averages obtained previously, CODENSA consumption tables were searched for as reference [3]; this in order to understand the circuits that feed the different consumer populations that have average consumption and a defined economic stratum. In order to validate the data, it was necessary to obtain the amount of population that inhabits the different localities of the city by stratum, which is how the localities are subdivided; this is done in order to know in some way, the transformers that are responsible for supplying the loads to these populations according to their size and average consumption in kW.
In [3], the size of the population will be an important value to dimension the distribution transformer to be installed in a new work depending of the stratum of the population. With this purpose in mind, we proceeded to obtain the number of clients by location and stratum; the highest concentration of population occurred in stratum 3 in general terms, followed by stratum 2.

Table 8.6: Amount of Clients Per Locality and Stratum

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	50	No Data	No Data	758	No Data	No Data	No Data
Suba	84	No Data	521	1169	No Data	4	42
Engativa	162	No Data	No Data	1596	No Data	No Data	No Data
Fontibon	101	No Data	No Data	477	125	No Data	No Data
Cogua	10	49	184	63	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	86	No Data	No Data
Cogua Veredal	No Data	10	85	2	No Data	No Data	No Data
Zipaquira	52	94	217	52	3	No Data	No Data
Rincon Santo	No Data	No Data	34	No Data	No Data	No Data	No Data
San Isidro	No Data	3	16	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	11	No Data	No Data	No Data	No Data

The average loads in KVA were obtained with the tables 8.2 8.4;The following table is a summary of the previous tables, allowing us to know the consumption depending on the location and stratum studied.

Table 8.7: KVA Consumption for Each Stratum and Region

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	0.40	No Data	No Data	0.18	No Data	No Data	No Data
Suba	0.51	No Data	0.18	0.15	No Data	0.14	0.52
Engativa	0.65	No Data	No Data	0.24	No Data	No Data	No Data
Fontibon	1.39	No Data	No Data	0.14	0.18	No Data	No Data
Cogua	0.32	0.16	0.15	0.13	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	0.13	No Data	No Data
Cogua Veredal	No Data	0.07	0.14	0.11	No Data	No Data	No Data
Zipaquira	0.30	0.15	0.12	0.17	0.45	No Data	No Data
Rincon Santo	No Data	No Data	0.13	No Data	No Data	No Data	No Data
San Isidro	No Data	0.11	0.15	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	0.12	No Data	No Data	No Data	No Data

Assuming that there are no charges installed in common areas whose use is not frequent or loads different from common services such as commercial or residential, the transformer will normally be selected depending on the average load in KVA and the number of customers studied; it will be assumed that each transformer will only supply residential units of the same stratum and of the same locality.

Table 8.8: Transformers Used for Supply Power for each Locality Depending the Stratum

	0	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	45 KVA	No Data	No Data	500 KVA	No Data	No Data	No Data
Suba	75 KVA	No Data	400 KVA	500 KVA	No Data	15 KVA	75 KVA
Engativa	225 KVA	No Data	No Data	2*630 KVA	No Data	No Data	No Data
Fontibon	225 KVA	No Data	No Data	400 KVA	112.5 KVA	No Data	No Data
Cogua	15 KVA	30 KVA	30 KVA	112.5 KVA	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	75 KVA	No Data	No Data
Cogua Veredal	No Data	15 KVA	45 KVA	15 KVA	No Data	No Data	No Data
Zipaquira	30 KVA	45 KVA	112.5 KVA	45 KVA	15 KVA	No Data	No Data
Rincon Santo	No Data	No Data	30 KVA	No Data	No Data	No Data	No Data
San Isidro	No Data	15KVA	15 KVA	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	15 KVA	No Data	No Data	No Data	No Data

Taking into account the reference, the size of the sample, the average consumption, the location and the stratum, we proceeded to develop the table 8.8.The effect of the population on the size of the transformer is evidenced, where the areas with the highest population density, in turn, will have a transformer with a higher load. In general terms, Fontibon will be the location with the largest average transformers, followed by Engativa. In case of having all the measurements of all the households of each locality and each stratum, it will be possible to know the different circuits in a rigorous way, which are responsible for feeding all these thousands of users.

In order to better understand the behavior of each stratum and locality, the means for power consumption were taken in a period of 24 hours and they were superimposed depending on whether one is comparing strata or localities.

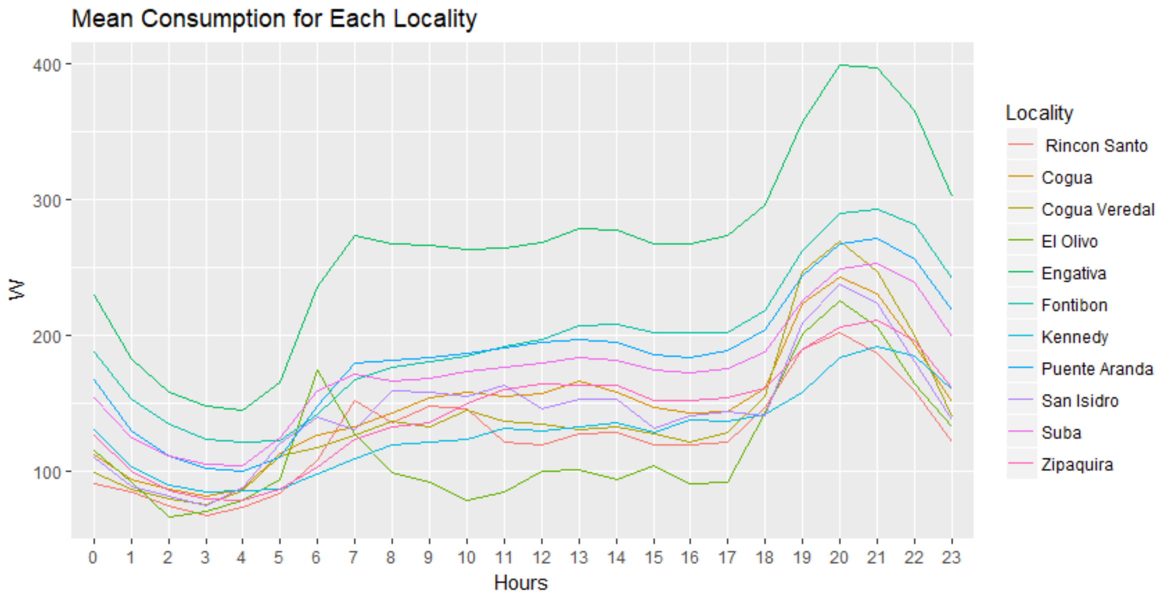

Figure 8.1: Mean Consumption for all Locations

In the figure 8.1 is easy to see how Fontibon is the locality with the greatest mean consumption of power, having a similar behaviour with any other curves. One first approach for understanding the way each locality behaves, will be obtaining the rates of changes of this curves, to establish wish locality in mean wakes up first during the morning. In the Appendix B, the rates of changes for the mean consumption of each locality is shown for the morning hours, to find the hour where the majority of people wake up in the morning for each zone; rate changes helps in the identification of changes in curves, if in a discrete interval, the curve augment, the rate of change will be positive, if the curve decrease, the rate of change will be negative, the magnitude of the rate of change will depend of the difference between the value in the two intervals. A greater rate of change will tell that there was a change in the curve that should be consider. Using the data from tables B.1 and B.2 its possible to determine the periods where each population wakes up. For the majority of localities, the peak of maximum rate change is around 6 AM, this will be an exception for the Cogua and Cogua Veredal, where the rate change has its maximum around 5 AM, for Fontibon and Rincon Santo, the maximum will happen around 7 AM. Fontibon and Rincon Santo are located near by in the west of the city, with a similar distance to the center of the city in comparison to Kennedy and Puente Aranda. Another determinant that could validate the hypothesis would be to analyze the hour of consumption increase in the afternoon hours, however, as seen in the graph, the consumption in the different localities tend to be similar.

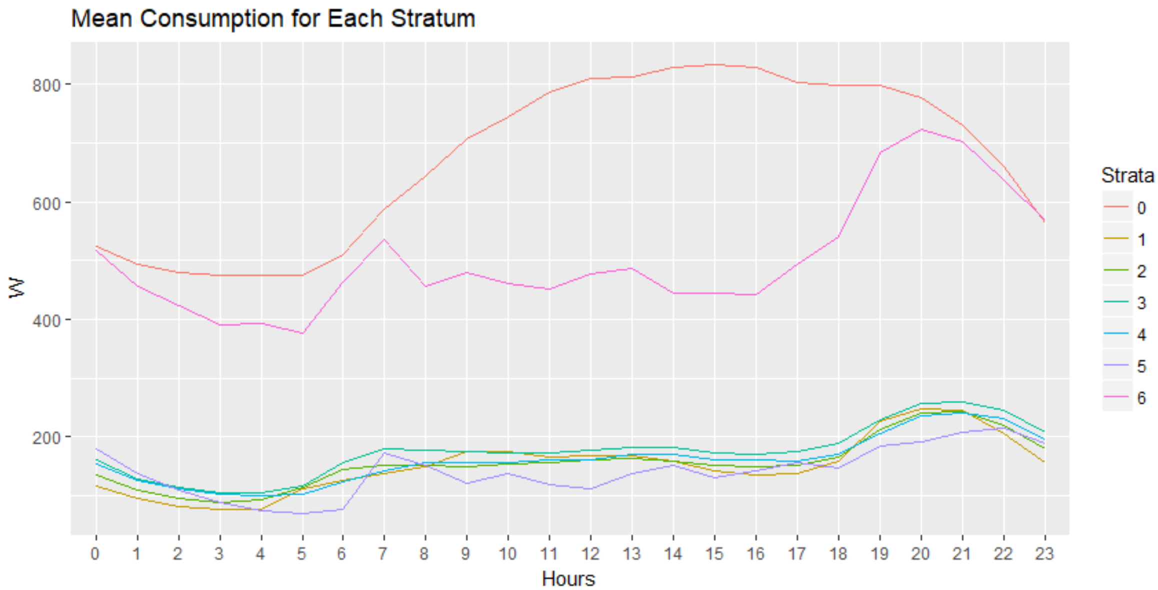

Figure 8.2: Mean Consumption for all Strata

In the figure 8.2, it's possible to see the behavior of the different strata, if any want to investigate thoroughly in each of these groups you can see the Appendix C. Stratum 0 will make a difference, having the highest consumption in comparison with all the strata, in addition to having a consumption curve unique in its type, which should be explained more thoroughly using other methodologies such as surveys of the analyzed populations. In the case of the rates of changes (tables B.3 B.4), it was possible the time of the day where each stratum wakes up; the first stratum to wake up is stratum 1 around 5 AM, Strata 2,3,4,6 wakes up around 6 AM, and finally strata 0 and 5 wakes up around 7 AM. In the case of stratum 5, only the behavior of 4 consumers is being studied, so this curve should be revised when more data is available. In the case of stratum 0, there is evidence of abnormal energy consumption, in addition to a late start of activity compared to the other strata, the consumption curve seems commercial, as if it were linked to clandestine workshops or any other direct residential activity; As stated previously, it will be necessary to conduct surveys that validate these hypotheses.

8.1 Economical Stratum Analysis

To visualize and record the way different strata from the city consume energy along a 24 hours interval, it was need to join the data from the months given (November 2016, December 2016, January 2017, February 2017), after this, using the function **Separate Stratum** seven different groups were created with the total amount of 24 hours curves (102906), where each stratum represents a percentage of all the amount of curves:

1. Stratum 0: 7.13%
2. Stratum 1: 1.61%
3. Stratum 2: 15.58%
4. Stratum 3: 70.76%
5. Stratum 4: 3.95%
6. Stratum 5: 0.05%

7. Stratum 6: 0.91%

As it can be simple seen, the vast majority of data was taken for the stratum 3 clients, while the stratum 5 only records a few measurements of 4 meters.

Each number of the table 8.10 represents one locality, as is shown in the following:

Table 8.9: Localities

Puente Aranda	1
Suba	2
Engativa	3
Fontibon	4
Cogua	5
Kennedy	7
Cogua Veredal	7
Zipaquira	8
Rincon Santo	9
San Isidro	10
El Olivo	11

Table 8.10: Correlation Matrix of Mean Energy Use for a 24 hour period

Locality	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	1.000	0.992	0.992	0.975	0.950	0.966	0.895	0.980	0.916	0.910	0.732
2	0.992	1.000	0.995	0.967	0.963	0.958	0.922	0.965	0.920	0.930	0.791
3	0.992	0.995	1.000	0.959	0.961	0.947	0.920	0.960	0.943	0.932	0.785
4	0.975	0.967	0.959	1.000	0.920	0.993	0.888	0.965	0.848	0.855	0.737
5	0.950	0.963	0.961	0.920	1.000	0.894	0.967	0.941	0.947	0.977	0.779
6	0.966	0.958	0.947	0.993	0.894	1.000	0.859	0.956	0.816	0.840	0.711
7	0.895	0.922	0.920	0.888	0.967	0.859	1.000	0.861	0.927	0.948	0.861
8	0.980	0.965	0.960	0.965	0.941	0.956	0.861	1.000	0.867	0.890	0.642
9	0.916	0.920	0.943	0.848	0.947	0.816	0.927	0.867	1.000	0.937	0.757
10	0.910	0.930	0.932	0.855	0.977	0.840	0.948	0.890	0.937	1.000	0.763
11	0.732	0.791	0.785	0.737	0.779	0.711	0.861	0.642	0.757	0.763	1.000

From the table 8.10, it's possible to identify the correlation matrix for the means of each locality studied. The great majority of localities tend to consume energy in a similar manner during the day, this will give a strong correlation between localities that are in Bogota and have a great amount of data to be evaluated. Suba, Fontibon, Puente Aranda, Cogua and Engativa, have a strong correlation between each other, this happen, even if Cogua is a town near Bogota. Rincon Santo, San Isidro and El Olivo, are the places that have the lower correlations between any other localities, this might be because of the lack enough information from this zones to do a proper load characterization. From this results, it's valid to say, that people from Bogota tend to have a similar routine that is independent from the place their live. The hipothesis could not be validated with the data given, it's necessary to have data from industrial areas to compare the different curves.

8.1.1 Characteristics of Energy Mean for Each Stratum

Table 8.11: Mean Energy Total Consumption Characteristics for All Strata by User

Characteristic	S0	S1	S2	S3	S4	S5	S6
Average Demand W	672.364	151.837	156.078	175.550	160.754	142.033	502.093
Maximum Demand W	833.335	247.383	242.073	260.188	241.410	214.700	724.040
Hour of Maximum Demand	15:00	20:00	21:00	21:00	21:00	22:00	20:00
Load Factor	0.807	0.614	0.645	0.675	0.666	0.662	0.693

This table represents the information showed in the graphs, where it can be displayed the weird pattern of the stratum 0, this will have the greatest load factor, average demand and maximum demand, also the hour of the maximum demand is in the afternoon, being atypical with respect to the other groups. All other strata tend to behave in similar ways, the greatest difference will be the stratum 6, that has the greatest peak of consumption, this can be easily explained with the fact that stratum 6 group is the one with more economical incomes, this will refer in a bigger energetic consumption. Evaluating the Load Factor, the stratum 0 is the one with the greatest.
The Table 8.11 reveals how every stratum of the city and it's surroundings tends to consume energy, being the Stratum 0 an abnormality. Purchasing power is linked to the consumption of electricity, so the consumption curve of the population with the lowest income in Bogota represents an anomaly, several hypotheses can be made to try to explain this phenomenon:

1. It is possible that the houses are engaged in illegal commercial activities (furnaces, workshops, plastics factories, foundries, mining machinery, etc.)
2. The lowest cost that these populations must pay to consume the service.
3. This stratum represents a population with a high unemployment rate.
4. It is possible that families are more numerous, so there will be an increase in the consumption of basic services.

8.2 Geographical Analysis

Using the functions of the libraries **leaflet** and **ggmap**, it was possible to visualize groups of consumers that are located in different places of the city. Using the function **Assign Random Coordenates**, an scenery of a map with consumers that have an specific energy consumption can be created.

Figure 8.3: Clients/Regions Representations using *Leaflet*

This tool will be used to relate the geographical coordinates of each user with their energetic pattern. Using this toolbox created in R for this specific purpose, the hypothesis will be tried to be validated. The average consumption by localities on different days of the week in the Appendix F was studied, showing how all regions tend to behave similarly regardless of the day of the week, the time and the month; something that goes hand in hand with overlapping power curves for locations previously studied in this document.

In the following section, a graphical analysis will be implemented using the libraries **ggmap** and **leaflet**, in which the localities mean consume for an specific hour of a weekday will be studied. Also, a correlation analysis between a 24 hours for different localities will be done; in this specific case, the correlation between localities will be represented using lines with different thickness, a greater thickness indicates a greater correlation between means. Also, using pie-charts, the correlation matrix will be represented, giving a new form of understating the data.The maps represent the areas of urban Bogota, this is because if regions such as Cogua or Zipaquira are represented, the maps would be too extensive. Taking the 4 months studied, we look for patterns that allow identifying the behavior of the individuals that move in the city. In general terms, it was found that the localities have similar consumption trends regardless of their distance to the center.

8.2.1 November

For the month of November, there is a high correlation between the consumption averages of 24 hours per locality, this is visually evident with the thickness of the lines that connect to each locality.In general for the different areas of Bogota, consumption will tend to be similar between zones. There will be a high correlation between the Bogota Urbana in comparison with the adjacent rural areas.

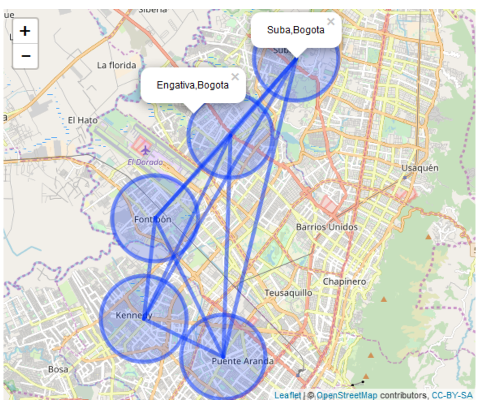

Figure 8.4: Map of Correlations Between Mean Consumption for Mondays Nov/16 for Locations of Study

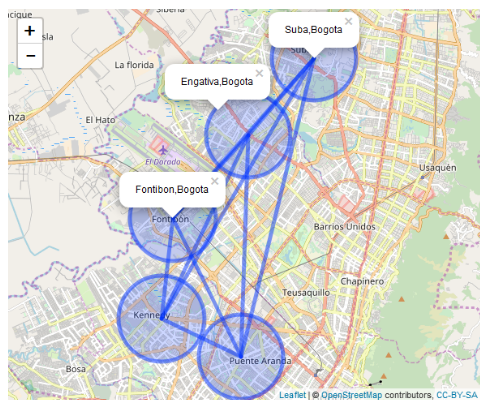

Figure 8.5: Map of Correlations Between Mean Consumption for Sundays Nov/16 for Locations of Study

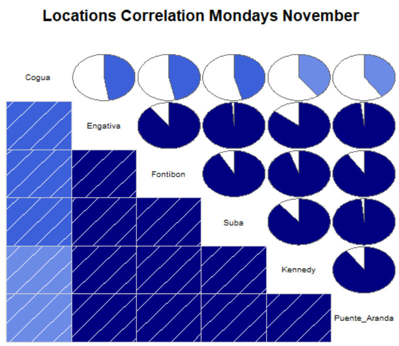

Figure 8.6: Correlations Visualized using Pie Chart Between Mean Consumption for Mondays Nov/16 for Locations of Study

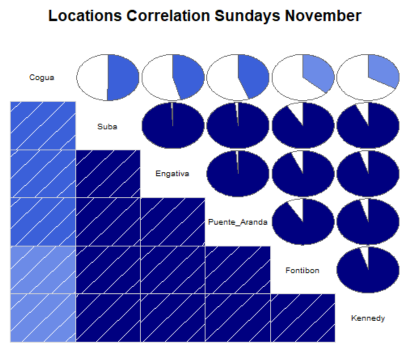

Figure 8.7: Correlations Visualized using Pie Chart Between Mean Consumption for Sundays Nov/16 for Locations of Study

There is a high correlation between all the localities, except for any couple with Cogua. This is due to the fact that Cogua represents a semi-urban population, with consumption patterns different from those seen in the city.

Table 8.12: Correlation between the 24 Hours Mean for each Locality Mondays Nov/16

Localities	Puente_Aranda	Suba	Engativa	Fontibon	Cogua	Kennedy
Puente_Aranda	1.000	0.986	0.982	0.913	0.402	0.908
Suba	0.986	1.000	0.987	0.920	0.458	0.890
Engativa	0.982	0.987	1.000	0.897	0.476	0.859
Fontibon	0.913	0.920	0.897	1.000	0.466	0.950
Cogua	0.402	0.458	0.476	0.466	1.000	0.398
Kennedy	0.908	0.890	0.859	0.950	0.398	1.000

Table 8.13: Correlation between the 24 Hours Mean for each Locality Sundays Nov/16

Localities	Puente_Aranda	Suba	Engativa	Fontibon	Cogua	Kennedy
Puente_Aranda	1.000	0.986	0.988	0.915	0.443	0.955
Suba	0.986	1.000	0.991	0.913	0.503	0.933
Engativa	0.988	0.991	1.000	0.937	0.460	0.954
Fontibon	0.915	0.913	0.937	1.000	0.369	0.958
Cogua	0.443	0.503	0.460	0.369	1.000	0.330
Kennedy	0.955	0.933	0.954	0.958	0.330	1.000

8.2.2 December

In the month of December, as in the previous month, there is a high correlation between the localities of the city. The correlation between Cogua and the other localities increased, this might be because of the holidays season, where people of the town tend to change their habits, consuming more energy that in other months.

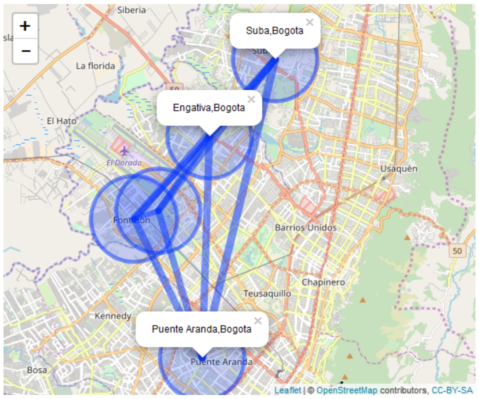

Figure 8.8: Map of Correlations Between Mean Consumption for Tuesdays Dec/16 for Locations of Study

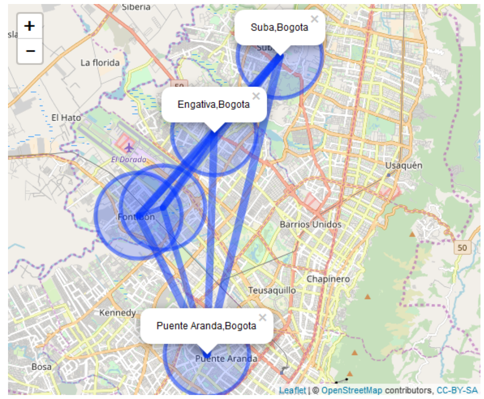

Figure 8.9: Map of Correlations Between Mean Consumption for Saturdays Dec/16 for Locations of Study

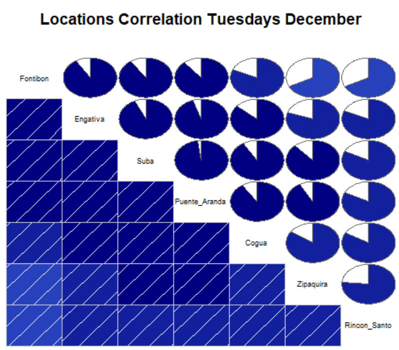

Figure 8.10: Correlations Visualized using Pie Chart Between Mean Consumption for Tuesdays Dec/16 for Locations of Study

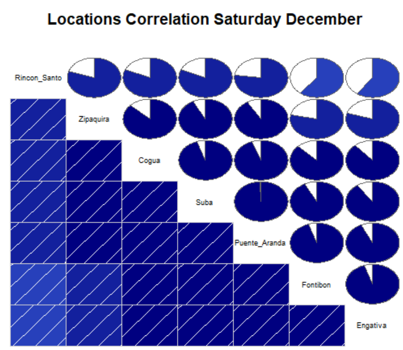

Figure 8.11: Correlations Visualized using Pie Chart Between Mean Consumption for Saturdays Dec/16 for Locations of Study

It is evidenced for the month of December, as for the month of January, a high correlation between all the zones, excepting the majority of the couples with Fontibon and Rincon Santo, as it is seen in the graphs 8.10 y 8.11.

Table 8.14: Correlation between the 24 Hours Mean for each Locality Tuesdays Dec/16

Localities	Fontibon	Puente_Aranda	Suba	Cogua	Rincon_Santo	Zipaquira	Engativa
Fontibon	1.000	0.888	0.904	0.815	0.676	0.673	0.912
Puente_Aranda	0.888	1.000	0.977	0.904	0.824	0.917	0.948
Suba	0.904	0.977	1.000	0.913	0.821	0.879	0.932
Cogua	0.815	0.904	0.913	1.000	0.834	0.846	0.861
Rincon_Santo	0.676	0.824	0.821	0.834	1.000	0.762	0.820
Zipaquira	0.673	0.917	0.879	0.846	0.762	1.000	0.801
Engativa	0.912	0.948	0.932	0.861	0.820	0.801	1.000

Table 8.15: Correlation between the 24 Hours Mean for each Locality Saturdays Dec/16

Localities	Fontibon	Puente_Aranda	Suba	Cogua	Rincon_Santo	Zipaquira	Engativa
Fontibon	1.000	0.948	0.925	0.874	0.608	0.784	0.949
Puente_Aranda	0.948	1.000	0.991	0.945	0.767	0.909	0.930
Suba	0.925	0.991	1.000	0.948	0.815	0.924	0.902
Cogua	0.874	0.945	0.948	1.000	0.816	0.870	0.886
Rincon_Santo	0.608	0.767	0.815	0.816	1.000	0.803	0.601
Zipaquira	0.784	0.909	0.924	0.870	0.803	1.000	0.798
Engativa	0.949	0.930	0.902	0.886	0.601	0.798	1.000

In general, the average for all localities tends to be similar, even comparing the average of the localities of Bogota with the measure given for Zipaquira.

8.2.3 January

In the month of January it is when the utility of the implemented tool becomes evident. Because in this case there is a lower correlation between the 24-hour averages of the study locations (which are few for this month), there is evidence of thick and thin lines on the maps.

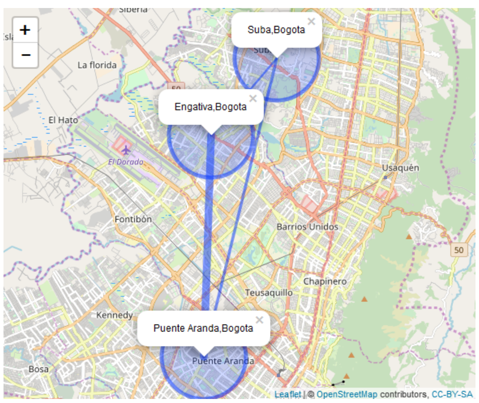

Figure 8.12: Map of Correlations Between Mean Consumption for Wednesdays Jan/17 for Locations of Study

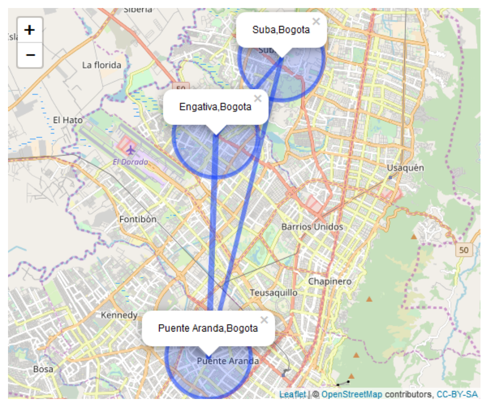

Figure 8.13: Map of Correlations Between Mean Consumption for Thursdays Jan/17 for Locations of Study

In general, the energy behavior of Suba differs greatly with that of the other locations for this specific month; even Zipaquira, which is a neighboring municipality, has a much higher correlation in comparison.

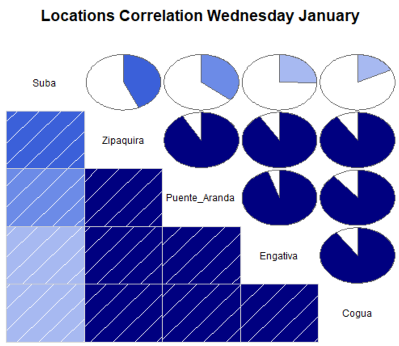

Figure 8.14: Correlations Visualized using Pie Chart Between Mean Consumption for Wednesdays Jan/17 for Locations of Study

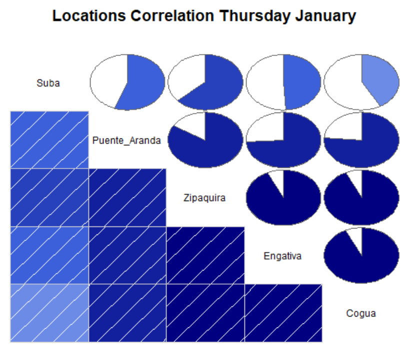

Figure 8.15: Correlations Visualized using Pie Chart Between Mean Consumption for Thursdays Jan/17 for Locations of Study

Table 8.16: Correlation between the 24 Hours Mean for each Locality Wednesdays Jan/17

Localities	Suba	Puente_Aranda	Zipaquira	Cogua	Engativa
Suba	1.000	0.360	0.432	0.176	0.255
Puente_Aranda	0.360	1.000	0.914	0.892	0.950
Zipaquira	0.432	0.914	1.000	0.905	0.911
Cogua	0.176	0.892	0.905	1.000	0.904
Engativa	0.255	0.950	0.911	0.904	1.000

Table 8.17: Correlation between the 24 Hours Mean for each Locality Thursdays Jan/17

Localities	Suba	Puente_Aranda	Zipaquira	Cogua	Engativa
Suba	1.000	0.555	0.635	0.416	0.491
Puente_Aranda	0.555	1.000	0.836	0.764	0.741
Zipaquira	0.635	0.836	1.000	0.931	0.930
Cogua	0.416	0.764	0.931	1.000	0.931
Engativa	0.491	0.741	0.930	0.931	1.000

8.2.4 February

In the month of February, an increase in energy consumption was seen, which was generalized for all the localities and neighboring towns of Bogota. The correlation between means of different localities is shown next:

Figure 8.16: Map of Correlations Between Mean Consumption for Fridays Feb/17 for Locations of Study

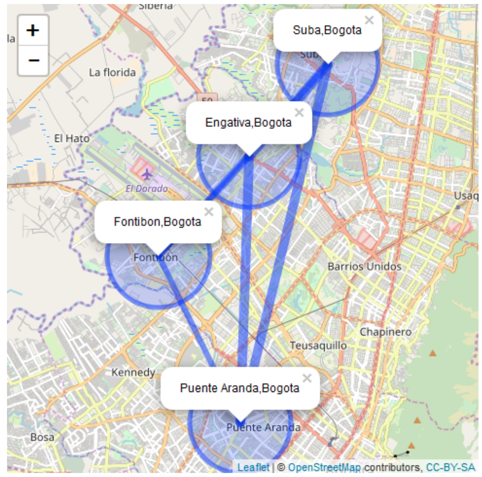

Figure 8.17: Map of Correlations Between Mean Consumption for Sundays Feb/17 for Locations of Study

The behavior of the energy curves was affected during this month, so the consumption between most of the locations was similar for most of the cases.The maps reveal the little correlation between Fontibón and Puente Aranda, something that should be studied a little more thoroughly; a first approximation to explain this low correlation, could occur due to the distances between both locations,however, in past cases, the correlation was high regardless of the distance where Fontibon didn't had an increase in the consumption as the other localities.

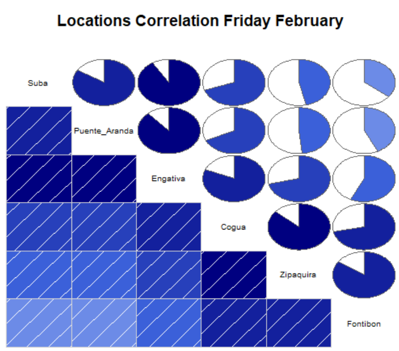

Figure 8.18: Correlations Visualized using Pie Chart Between Mean Consumption for Fridays Feb/17 for Locations of Study

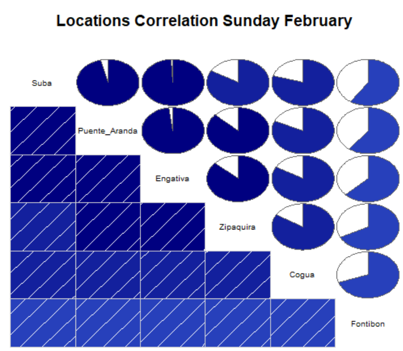

Figure 8.19: Correlations Visualized using Pie Chart Between Mean Consumption for Sundays Feb/17 for Locations of Study

Table 8.18: Correlation between the 24 Hours Mean for each Locality Fridays Feb/17

Localities	Puente_Aranda	Suba	Cogua	Zipaquira	Engativa	Fontibon
Puente_Aranda	1.000	0.841	0.680	0.484	0.884	0.427
Suba	0.841	1.000	0.696	0.464	0.910	0.353
Cogua	0.680	0.696	1.000	0.863	0.815	0.717
Zipaquira	0.484	0.464	0.863	1.000	0.714	0.848
Engativa	0.884	0.910	0.815	0.714	1.000	0.571
Fontibon	0.427	0.353	0.717	0.848	0.571	1.000

Table 8.19: Correlation between the 24 Hours Mean for each Locality Sundays Feb/17

Localities	Puente_Aranda	Suba	Cogua	Zipaquira	Engativa	Fontibon
Puente_Aranda	1.000	0.962	0.819	0.868	0.980	0.603
Suba	0.962	1.000	0.795	0.835	0.991	0.595
Cogua	0.819	0.795	1.000	0.834	0.836	0.697
Zipaquira	0.868	0.835	0.834	1.000	0.866	0.673
Engativa	0.980	0.991	0.836	0.866	1.000	0.628
Fontibon	0.603	0.595	0.697	0.673	0.628	1.000

8.3 Hypothesis Validation

In none of the cases studied, there was evidence of a transfer of cargo between localities, this is because only the curves of domestic consumption are taken and those for industrial and commercial consumption are not included. The analysis proposed in the hypothesis can't be done with this data-set, it would be necessary to have a bigger data set that includes equal amount of data of all strata, and for all zones of the city of Bogota. In the majority of cases, data was only available for some areas of the city, while other areas may have a small amount of data or simply no data as the table D.1 shows. As it was shown in D, some of the localities can not be characterized properly because of the small amount of information available for each case; San Isidro, El Olivo and Rincon Santo were problematic areas because their proper consumption of energy is not available. The correlation matrix for the mean of each locality (Table 8.10) showed how he locations that have a large amount of data tended to consume in a similar way, and this is in turn corroborated by the different boxplots obtained.The residential sector tend to consume in a similar way regardless of the locality in which the family is located. The greatest changes in consumption will occur due to the stratum, where for all cases except stratum 0, consumption will increase as purchasing power increases. Analyzing the data from Appendix B it was possible to discern the time when the average of the population rises to start their journey, however, it was shown as most homes get up at 6 AM (for rural areas it would be 5 AM) regardless of their location , the only places where the population rises later is in Fontibon and Rincon Santo; It is possible that because many offices and industries are located directly in Fontibon and Rincon Santo, people do not find the need to get up so early.
The argument of greater weight showed in Appendix F to not be able to prove the hypothesis, will be the lack of information of the commercial, executive and industrial sectors, nevertheless, with the aid of the geographic analysis it was possible to corroborate the already raised previously. In 8.2 with the use of different maps and charts, it was possible to identify the patterns of consumption for a population during a specific day at a specific hour. As it was mention previously, every population in the city of Bogota tends to consume in a similar manner, routines are generalized in the following way:

1. Every Family starts it's activities around 6 AM. The use of electrical showers, microwaves, electrical ovens, radios and televisions will increase as each family member wakes up.

2. Some members of the family leave the house around 7 AM because they have to work or study, but it's important to include the population that don't leave the house. A great amount of people stay all day at their homes.

3. Some workers tend to come back home during noon to lunch, this might explain why the consumption tend to increase a little bit during 12 PM.

4. Workers and students come back home around 7 PM, this will explain why the load increase during this hours until 9 PM, where the majority of population start going to bed.

As it was mention previously, only residential loads where characterized, the data don't give information of the places where workers and students go after they leave home; the consumption at this specific educational and industrial places will increase around 7 AM, and might decrease around 5 PM, where the majority of people leave to their houses. A complete load characterization of the different zones of the city, will validate the hypothesis stated, and will help **CODENSA** to understand how the load changes depending the hour and the zones of the city.

9. Recommendations for a Correct Demand Management

The data given by Smart Meters should be used to create new models of pricing the consumers, in such way that the system and the user get a benefit. At first, one can easily identify the way the majority of homes consume energy in the day, smart meters can change the way consumers use energy if information of energy prices are provided to the user. One can change the pattern of a population with the use of a pricing system that privileges the use of the system in hours where not many people consume. With the data studied previously, the following states should be consider at the hour of developing the new pricing methodology:

1. The maximum energy consume occurs around 7 PM to 10 PM; consumers should be motivated to consume earlier or later from this period of time. It's important to have a more homogeneous consumption of energy during the day, this will help to reduce the peaks of consumption that were seen in the different boxplots. Another important factor, will be the reduction of the variance of consumption during an specific hour.

2. In order to carry out a complete study, it will be necessary to fully understand the trends of electric power consumption according to type, purpose and use of the most common electrical appliances in the home, covering all the opinions and data of the inhabitants; In addition, the influence of electrical energy on the quality of life of families must be established.
   The development of the study, should apply a questionnaire to heads of household, having as main characteristic the census character of housing. With the help of the smart meters a quantitative phase will be given, which will allow knowing the consumption habits of the families, while with the qualitative the opinion of the inhabitants about their use of the electric power will be known. The variables to be taken into account in the questionnaire will be:

   - Habits and electrical consumption.
   - Amount of inhabitants of each house.
   - Satisfaction with regard to energy consumption.
   - Physical conditions of the housing.
   - Description of the family group and their lifestyle.

3. Many clients act as outliers of the system, smart metering can detect this anomalies and apply economical punishment to this clients. It's necessary for **CODENSA** to classified the different kinds of outliers using a the qualitative analysis mention previously; the company should identify why the families have abnormal patterns. In many cases, technical failures of electrical installations and households will be related to the families that are classified as outliers.

4. The pattern of different zones of the cities should be identified and classified, in this way **CODENSA** will have a frame of comparison for each client depending of the zone it lives. The consumer should need to have this ideal curve of consumption as his optimum energy consumption. For each locality, the characteristics of the electricity demand curve per family should be studied; for this, different physical and economic variables of the populations that inhabit the sector should be taken into account.Identifying the types of housing families (houses, apartments) will be necessary to adapt the daily consumption curves that each specific population will have; In this way, the energy can be charged depending on the needs of each customer in the network.Consumers who do not achieve the goals stipulated by the regulatory body will be punished with much higher bills.

5. Anomalies should be immediately reported to **CONDESA**, in this way, the non-technical losses will be reduced. The smart meter will be able to identify when the client has become an outlier, so it will generate a series of sanctions against the family that uses the power in an improper

way. As smart meters will give information in real time, it is possible to keep a strict control of the activities of the home, reporting during any hour of the day the possible problematic that a house might be facing.

6. Geographical tools should be implemented and used, this will give support for the clients living in a specific zone of the city. Every client will be identify in an interactive map, that can follow his profile as a consumer that has an specific stratum and that lives at an specific zone of the city.The analysis using maps will help to understand the different relationships existing in the city in terms of energy consumption graphically.Maps help consumers to know their position in the consumption chain, allowing consumers to compare each other with their neighbors.The possibility of having maps will open the fact of being able to visualize the different costs of energy in different areas of the city in real time; this will revolutionize the existing pricing system, since each price will be adapted to the restrictions of each community.

7. New methodologies should be proposed that promote electricity savings, such as turning off equipment when it is not used. New habits to encouraging the use of the washing machine with a full load, just like ironing only once a week. The load for lighting can be reduced with the adaptation of LED luminaries; the use of environmental campaigns promoted by governmental entities will help with this purpose.

10. Discussion

In this work, the greatest difficulty in performing the data analysis was due to the quality of the data provided by Codensa. All the months data-set had to be filter to remove the many mistakes made, the majority of them include:

- Duplicate Measures.
- Incomplete Measures.

After filtering, there was no way to validate the quality of the data remaining for starting the analysis. Every data set contains different amount of data, differing greatly with the other months. The following locations didn't contain a significant amount of data:

1. Cogua Veredal.
2. Rincon Santo.
3. San Isidro: Had a strange pattern for some cases studied.
4. El Olivo.

The sample in many cases for these locations was not representative, complicating in this way the validation of the hypothesis. Another great problem presented in the data set, happened for the month of January, where the stratum 5 and Fontibon where not measured,being in-congruent with past measurements.
It is necessary to include that this whole data-set was biased thanks to the amount of data from Stratum 3 group, not giving much information of any other stratum in the way it does with this one. Also, Stratum 5 and Stratum 6 only have data for the locality of Suba, it's necessary to count with data from different groups of localities.
For enhancing the analysis done in this project, more data should be analyzed in order to generate a better conclusion of Bogota's energy consumption. Equal amount of data from each locality should be needed to perform a good analysis that might or not validate the hypothesis.

11. Conclusions

This work has made possible the creation of a unique package for analyzing the smart of meters located in different parts of the city of Bogota; not only using traditional statistical methods, also combining the use of satellite images to form maps that include relevant information of a groups of consumers. Analyzing the data-set was done in 2 different ways:

1. Analyzing the data for each economical stratum.

2. Analyzing the data for each month given.

The first thing to emphasize, obtained from the data is the relationship between energy consumption and the habits of a group. Following this thread, it was possible to identify the consumption patterns of the different homes studied (due to their behavior, it is known that they are homes), evidencing the different routines that each family generally performs throughout the day. Without considering the stratum 0, every family tends to wake really early in the morning, leave the home generating a decrease in the energy, arriving to lunch, and after around 3 PM the energy consumption will start to increase until reaching a maximum around 8PM-9PM. Stratum 6 has the second greatest average amount of consumption, this is related to the amount of income that this families have, where many of them have more households in average that any other population. Stratum 0 has a strange pattern, it represents the group with the greatest average amount of consumption that is related to a energy consumption curve that differs from any other stratum and group; in this case, it can be seen that the routines of this group stars around 4 AM as any other group, but in this case, the consumption increases constantly until 3 PM, where it reaches it's maximum, slowly decreasing until 9 PM. In contrast to the amount of energy that the stratum 0 consume, it comes that they are the group with the lower incomes and greatest unemployment rate, this consumption might be explained if members of this families stay at home during all the day, having an increasing consumption during all day, that could be related to clandestine commercial activities that families do to increase their income.
Using differing set of map strategies, the hypothesis was tried to been validated, the results showed that it was not possible to reach or objective. All localities tend to consume in similar ways during the day, every family has a similar routine, and in this case the majority of data came from Stratum 3, biasing the whole data set for all months. It was not possible to state negative correlations between areas during different periods of time, all locations tend to increase and decrease in consumption in the same way. Using the map correlations lines, we could compare the correlation between different zones, for the majority of cases, the correlation coefficient was near 1, resulting in a thick line that doesn't tell any relevant information referring to the transportation of people. To validate the hypothesis, it would be needed to include data from commercial and industrial zones, this will include offices and business, in this way the population that leaves their homes will go to work to this locations, increasing the energy consumption during the labor hours, while decreasing after 4PM-7PM where the working population come back to their homes.
The boxplots used for understanding the variations of the data, allowed to identify outliers and the pattern of energy consumption of an specific population. The possibility of visualizing the great amount of outliers, in almost any boxplot, is one the benefits of using this technique of data visualization; also, understating the position of the mean in the boxplot, gives an idea of the spreading of each data.

12. Future Work

1. It's necessary to finish the construction of the **R** packet for analyzing data-sets from Smart Meters located anywhere in Colombia. This packet should include different functions for analyzing data using maps that contain every consumer; the map should be able to link each group of consumers with other that have similar consumption characteristics. Maps will be interactive, in such way, that anyone can easily extract information from them.

2. Create functions that could show the effect of real time pricing in consumption; the scenarios of load overlap depending on the price of energy will be simulated. With these functions, the idea is to seek to optimize the resources of the electrical system, in order to perform better demand management.

3. Surveys should be carried out in order to get to know the different families that live in the different localities of Bogota. These surveys will define the role of electric power in people's lives.

4. A greatest data-set should be used in order to validate properly the hypothesis established at the beginning of the document.

5. If it's possible, a comparison between curves of consumption of different cities should be done, in this way, the whole consumption of the country can be measured using the packet done in R.

6. Validate the data given by Codensa using the substation's energy curves that energize the zones of study.

7. Obtain a license for the packet, in this way anyone with educational purposes can use it, limiting the commercial usage.

8. It's necessary to research more about consumption in Fontibon, the great amount of energy used in this locality is not normal for Bogota.

Bibliography

[1] *Communication Networks for Smart Grids - Making Smart Grid | Kenneth C. Budka | Springer.*

[2] *Studies in Classification, Data Analysis, and Knowledge Organization.*

[3] CODENSA. Carga Máxima para el Sector Residencial, May 2010.

[4] Congreso de Colombia. Ley 505 de 1999 Nivel Nacional, June.

[5] CONSORCIO CORPOEMA-CUSA. Caracterización Energética del Sector Residencial Urbano y Rural en Colombia, Mar. 2012.

[6] Grupo Técnico Proyecto BID. Smart Grids Colombia Visión 2030 Parte I, Apr. 2016.

[7] Jenny Paola Lozano Celis and William Camilo Guzmán Espitia. *Evaluación de Demanda de Energía Eléctrica según Hábtios de Consumo Actuales en la Ciudad de Bogotá.* Grupo de Ingestigación GESETIC, Universidad Distrital José de Caldas, Bogotá D.C, 2016.

[8] Juan Felipe Valencia Uribe. *Aplicaciones de Grandes Datos en los Medidores Inteligentes del Consumidor de Electricidad Residencial Colombiano.* Trabajo final de maestría presentado como requisito parcial para optar al título de: Magister en Ingeniería de Sistemas Energéticos, Universidad Nacional de Colombia, Medellín, Colombia, 2014.

[9] T. Kodinariya and P. Dan Makwana. Review on Determining of Cluster in K-means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1:90–95, Jan. 2013.

[10] Maria Angélica Arroyo Bernal. *Análisis de la información y caracterización del consumo de energía eléctrica de los estratos 1 y 2 caso EMCALI, como base para la implantación de programas de respuesta de la demanda.* Proyecto Fin de Carrera Ingeniería Eléctrica, Universidad de los Andes, Bogota, Colombia, Dec. 2016.

[11] VICTORIA EUGENIA PÉREZ VÉLEZ. *ESTUDIO PRELIMINAR SOBRE LA VIABILIDAD DE LA IMPLEMENTACIÓN DE MEDIDORES INTELIGENTES DE ENERGÍA EN LOS ESTRATOS 1, 2 Y 3 DE CALI.* Trabajo de Grado, Universidad del Valle, Santiago de Cali, Colombia, 2013.

[12] William H. Kersting. *Distribution System Modelling.* The ELECTRIC POWER ENGINEERING Series. CRC Press, United States of America, 2002.

Appendices

A. Variance of Load Factors and Energy

Table A.1: Business Days: Variance of Total Energy Mean.

	0	1	2	3	4	5	6
Puente Aranda	3174894	No Data	No Data	3212457	No Data	No Data	No Data
Suba	4939145	No Data	3838889	3022352	No Data	3394435	2450527
Engativa	1428780	No Data	No Data	3665142	No Data	No Data	No Data
Fontibon	3671346	No Data	No Data	2087075	3235037	No Data	No Data
Cogua	7918139	2250331	2934192	2555134	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	2475253	No Data	No Data
Cogua Veredal	No Data	111163.4	2336761	2315138	No Data	No Data	No Data
Zipaquira	3299049	2320373	2144597	2803244	7483920	No Data	No Data
Rincon Santo	No Data	No Data	2062190	No Data	No Data	No Data	No Data
San Isidro	No Data	1367615	1870053	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	2168377	No Data	No Data	No Data	No Data

Table A.2: Business Days: Variance of Load Factor Mean.

	0	1	2	3	4	5	6
Puente Aranda	0.00951	No Data	No Data	0.0161	No Data	No Data	No Data
Suba	0.00396	No Data	0.0154	0.01961	No Data	0.00895	0.00563
Engativa	0.00556	No Data	No Data	0.01196	No Data	No Data	No Data
Fontibon	0.001372	No Data	No Data	0.01701	0.01782	No Data	No Data
Cogua	0.02947	0.038601	0.015447	0.01722	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	0.02973	No Data	No Data
Cogua Veredal	No Data	0.007976	0.014233	0.04329	No Data	No Data	No Data
Zipaquira	0.026297	0.012771	0.015226	0.018551	0.008621	No Data	No Data
Rincon Santo	No Data	No Data	0.010836	No Data	No Data	No Data	No Data
San Isidro	No Data	0.023093	0.011152	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	0.010602	No Data	No Data	No Data	No Data

Table A.3: Holidays: Variance of Total Energy Mean.

	0	1	2	3	4	5	6
Puente Aranda	3462866	No Data	No Data	3419479	No Data	No Data	No Data
Suba	4261459	No Data	3547594	3170530	No Data	1767200	3684048
Engativa	2442272	No Data	No Data	3561537	No Data	No Data	No Data
Fontibon	3392589	No Data	No Data	2197835	3007978	No Data	No Data
Cogua	8889408	1956428	3002232	2610560	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	2874645	No Data	No Data
Cogua Veredal	No Data	16380.5	1979386	1340301	No Data	No Data	No Data
Zipaquira	1972341	2705331	2356827	2832433	434312	No Data	No Data
Rincon Santo	No Data	No Data	2213015	No Data	No Data	No Data	No Data
San Isidro	No Data	No Data	2921609	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	1926899	No Data	No Data	No Data	No Data

Table A.4: Holidaus: Variance of Load Factor Mean.

Locality	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Puente Aranda	0.01027	No Data	No Data	0.01735	No Data	No Data	No Data
Suba	0.008587	No Data	0.014736	0.021583	No Data	0.00207	0.008287
Engativa	0.035621	No Data	No Data	0.014269	No Data	No Data	No Data
Fontibon	0.00709	No Data	No Data	0.017562	0.017896	No Data	No Data
Cogua	0.030312	0.012173	0.019449	0.027618	No Data	No Data	No Data
Kennedy	No Data	No Data	No Data	No Data	0.038907	No Data	No Data
Cogua Veredal	No Data	0.010527	0.019833	0.038214	No Data	No Data	No Data
Zipaquira	0.031447	0.02004	0.022178	0.021786	0.001358	No Data	No Data
Rincon Santo	No Data	No Data	0.022981	No Data	No Data	No Data	No Data
San Isidro	No Data	No Data	0.01182	No Data	No Data	No Data	No Data
El Olivo	No Data	No Data	0.007886	No Data	No Data	No Data	No Data

B. Rates of Change of Curves of Mean Consumption

Table B.1: Rates of Changes of Mean Consumption for Localities for the period 12AM-6AM

Locality	12AM-1AM	1AM-2AM	2 AM-3AM	3 AM-4 AM	4 AM-5 AM	5 AM-6 AM
Puente Aranda	-24.906	-15.389	-8.878	-2.180	9.635	29.719
Suba	-21.154	-11.374	-5.930	-0.462	17.287	24.523
Engativa	-22.533	-14.113	-6.916	-1.836	12.559	35.674
Fontibon	-20.818	-12.688	-8.215	-2.186	0.711	14.757
Cogua	-17.518	-8.687	-5.563	5.405	27.378	10.698
Kennedy	-23.843	-13.088	-6.422	1.345	1.156	11.856
Cogua Veredal	-13.300	-8.350	-4.861	12.038	26.154	5.767
Zipaquira	-23.608	-15.032	-6.817	-1.485	9.032	17.830
Rincon Santo	-7.097	-13.329	-8.778	8.289	12.025	26.459
San Isidro	-20.670	-8.428	-10.078	17.348	30.739	15.384
El Olivo	-22.723	-31.951	6.305	10.459	17.306	62.328

Table B.2: Rates of Changes of Mean Consumption for Localities for the period 6AM-12PM

Locality	6AM-7AM	7AM-8AM	8AM-9AM	9AM-10AM	10AM-11AM	11AM-12PM
Puente Aranda	19.453	0.789	1.323	1.704	2.076	1.892
Suba	7.402	-2.986	0.986	3.063	1.874	1.778
Engativa	14.904	-2.378	-0.240	-1.228	0.241	1.558
Fontibon	16.854	5.146	2.336	2.346	3.799	2.490
Cogua	4.739	7.605	7.304	2.560	-1.831	1.347
Kennedy	10.830	9.485	0.956	1.597	6.607	-1.415
Cogua Veredal	7.059	7.649	-2.814	8.571	-5.885	-1.105
Zipaquira	17.658	7.681	1.639	10.142	6.339	2.696
Rincon Santo	33.638	-11.195	8.134	-1.114	-18.096	-1.923
San Isidro	-7.294	20.244	-0.964	-2.086	5.571	-11.475
El Olivo	-31.294	-25.854	-7.310	-15.189	7.289	16.249

Table B.3: Rates of Changes of Mean Consumption for Strata for the period 12AM-6AM

Locality	12AM-1AM	1AM-2AM	2 AM-3AM	3 AM-4 AM	4 AM-5 AM	5 AM-6 AM
Stratum 0	-6.147357	-2.502763	-0.940435	-0.327567	-0.0799813	7.101921
Stratum 1	-20.048545	-14.3024	-7.446032	2.1391822	36.882603	10.800933
Stratum 2	-21.669301	-13.39999	-6.120893	3.2487284	21.2052982	23.67206
Stratum 3	-21.966161	-12.5659	-7.273376	-1.591432	12.566087	28.233985
Stratum 4	-19.597972	-13.57059	-7.571371	-1.485651	0.63707889	18.378585
Stratum 5	-26.378056	-22.58115	-23.15556	-15.36276	-7.2445461	10.260179
Stratum 6	-12.234644	-7.533977	-8.115154	0.4713931	-4.2368555	20.918881

Table B.4: Rates of Changes of Mean Consumption for Strata for the period 6AM-12PM

Locality	6AM-7AM	7AM-8AM	8AM-9AM	9AM-10AM	10AM-11AM	11AM-12PM
Stratum 0	14.185923	9.135951	9.4017353	5.4041514	5.4455432	2.8189484
Stratum 1	9.183989	8.952551	14.958555	0.8466994	-5.3928519	1.5074764
Stratum 2	4.783839	-1.038695	-0.279839	2.881477	1.6731525	1.6525276
Stratum 3	14.542406	-2.300203	-0.913016	-0.775636	0.2031479	1.4980265
Stratum 4	15.181459	8.917822	-0.443878	0.7926124	2.8186883	0.5031088
Stratum 5	79.789401	-12.90964	-22.23168	12.601354	-14.247053	-5.3495609
Stratum 6	14.461532	-15.94465	4.9288181	-4.336086	-1.5926521	5.3426506

C. Strata Analysis

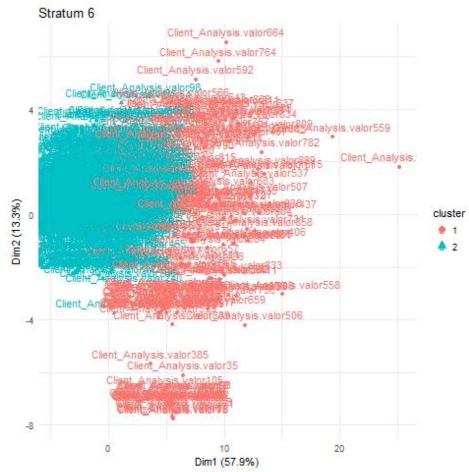

Figure C.1: Cluster Stratum 6

Using the outliers removal function, the intention will be to form a group of analysis for each case with a minor quantity of outliers. This function will be applied for each stratum data-set, this will be done, in order to just have a unique mean of consumption for a 24 hours interval. Using the boxplots, the variation of the data can be easily recognize.

C.1 Stratum 0

Being the stratum 0 the one with minor economical incomes per family, one might expect that their energy consumption was the minor from the whole set; this hypothesis was invalidated with the results obtained:

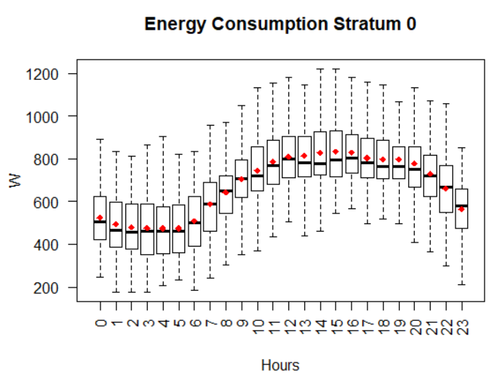

Figure C.2: Mean Consumption for Stratum 0 in a 24 Hour Lapsus

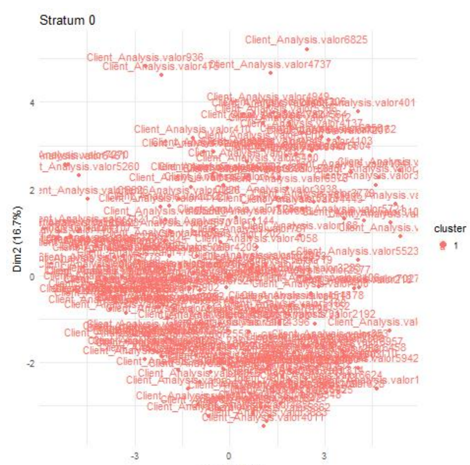

Figure C.3: Dispersion of Data

In C.2 it can be easily seen how the mean is usually equal or near to the median. The variation for this data-set is low, usually all the group of clients tend to consume in a similar way for all hours, this except for some high amount of outliers that are seen around 7 PM to 10 PM.

Table C.1: Variation Coefficient Stratum 0

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.275	10	0.199	20	0.234
1	0.286	11	0.189	21	0.234
2	0.284	12	0.165	22	0.239
3	0.298	13	0.179	23	0.270
4	0.316	14	0.185		
5	0.284	15	0.178		
6	0.289	16	0.175		
7	0.288	17	0.174		
8	0.247	18	0.184		
9	0.231	19	0.213		

This stratum is the one will less variation for each data. Also the mean represents a unique pattern of energy use, where it tends to increase until the afternoon, decreasing until 11 PM. It's necessary to understand the routines of this population, in order to find the reason of this rare curve that is similar for all the group. A possibility might be related with the amount of people that don't leave the house during the morning.

C.2 Stratum 1

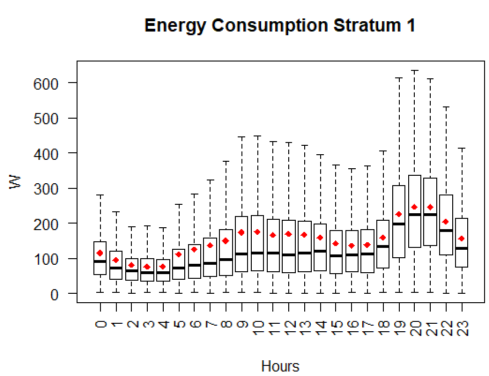

Figure C.4: Mean Consumption for Stratum 1 in a 24 Hour period

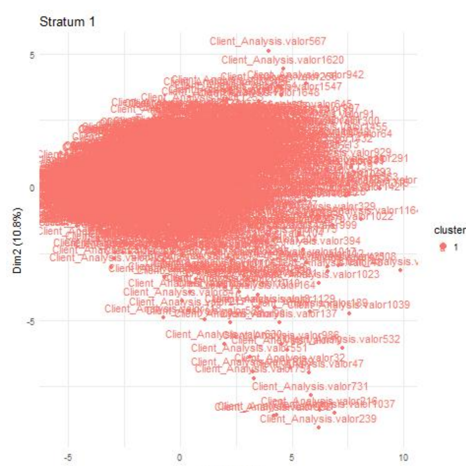

Figure C.5: Dispersion of Data

Table C.2: Variation Coefficient Stratum 1

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.798	10	1.013	20	0.638
1	0.868	11	0.937	21	0.620
2	0.904	12	0.995	22	0.649
3	0.864	13	0.993	23	0.702
4	0.888	14	0.935		
5	1.092	15	0.943		
6	1.103	16	0.786		
7	1.097	17	0.799		
8	1.037	18	0.770		
9	0.986	19	0.721		

The majority of energy curves will be similar to this one, where the habits of a normal worker family will be visualized in the curve. Usually an augment of energy will be seen around 5 AM, having a peak around 7 AM, after this, a reduction of energy will be seen until the first hours of the afternoon, where the amount of energy will slowly increase until night.

C.3 Stratum 2

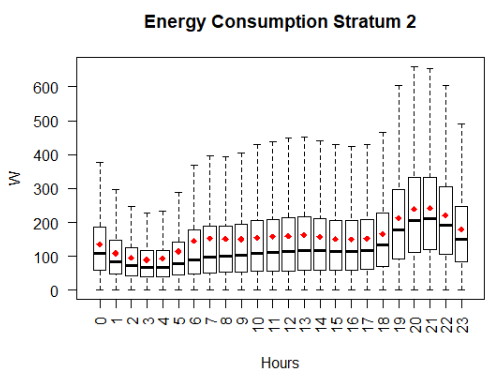

Figure C.6: Mean Consumption for Stratum 0 in a 24 Hour period

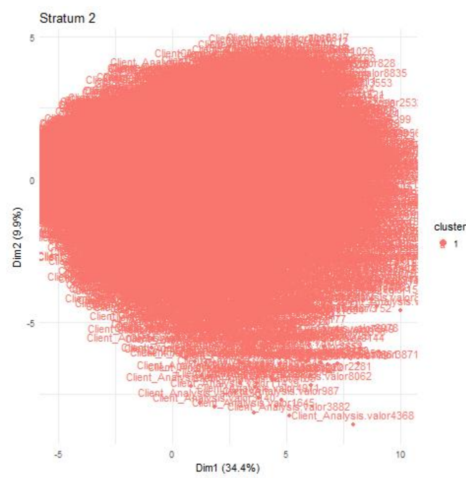

Figure C.7: Dispersion of Data

The amount of outliers found for the stratum 2 case is enormous, with a variation coefficient of 1.077 during 6 AM of the morning.The large variation in hours of tomorrow could give CODENSA a clue about houses that are badly stratified.The electrical pattern of the group tends to be normal during the whole day. One important aspect that the Table C.3 shows is the constant decrease of the variation coefficient after 8 AM, this might be probably related, with the hours of waking up of families. This strata tends to consume in a similar manner after 6 AM to 5 PM.

Table C.3: Variation Coefficient Stratum 2

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.738	10	0.935	20	0.698
1	0.768	11	0.927	21	0.669
2	0.802	12	0.920	22	0.669
3	0.819	13	0.922	23	0.701
4	0.887	14	0.895		
5	0.993	15	0.854		
6	1.077	16	0.835		
7	1.017	17	0.828		
8	1.002	18	0.793		
9	0.956	19	0.735		

C.4 Stratum 3

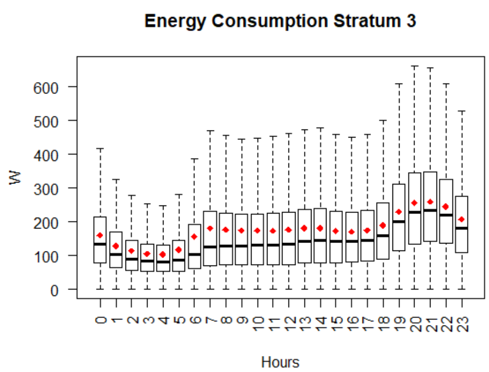

Figure C.8: Mean Consumption for Stratum 3 in a 24 Hour period

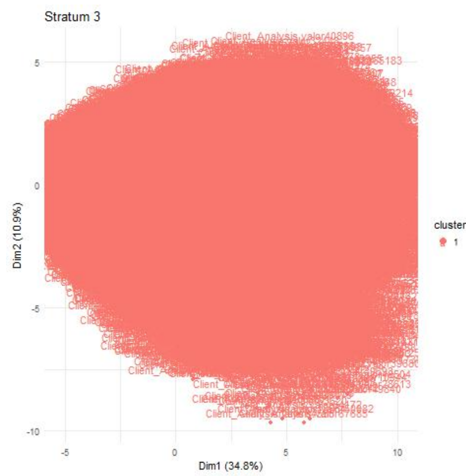

Figure C.9: Dispersion of Data

The amount of data providing from this group, might explain the great amount of outliers that exists during all the day. The variation coefficient tends to decrease constantly after 12 PM, with an small increase at 11 PM.

Table C.4: Variation Coefficient Stratum 3

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.673	10	0.835	20	0.621
1	0.704	11	0.831	21	0.603
2	0.735	12	0.835	22	0.598
3	0.746	13	0.802	23	0.628
4	0.761	14	0.770		
5	0.880	15	0.741		
6	0.965	16	0.720		
7	0.911	17	0.711		
8	0.866	18	0.688		
9	0.855	19	0.652		

C.5 Stratum 4

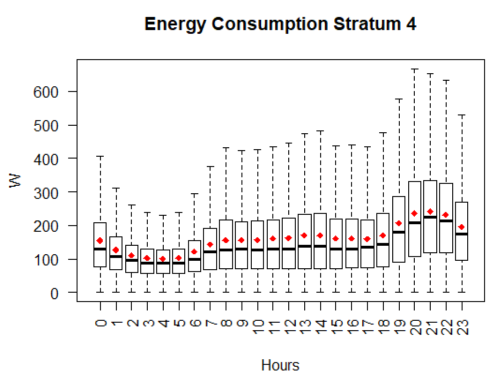

Figure C.10: Mean Consumption for Stratum 4 in a 24 Hour period

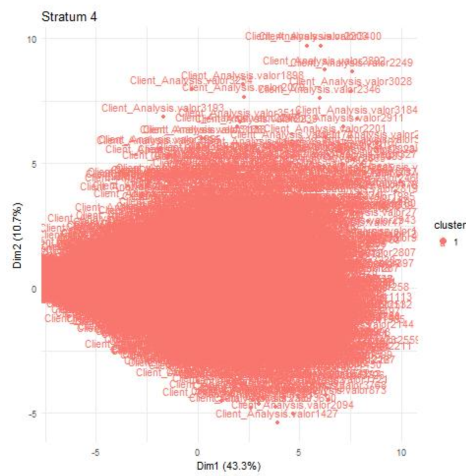

Figure C.11: Dispersion of Data

The energy consumption for this stratum presents similar characteristics when compared with stratum 1, 2 and 3, visually seen, a lower average consumption compared to stratum 3. Unlike the previous cases, in this there is no evidence of a constant reduction of the coefficient of variation in the day, there is evidence of a tendency to remain constant.

Table C.5: Variation Coefficient Stratum 4

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.686	10	0.761	20	0.697
1	0.656	11	0.785	21	0.652
2	0.642	12	0.785	22	0.636
3	0.636	13	0.794	23	0.673
4	0.621	14	0.781		
5	0.640	15	0.776		
6	0.727	16	0.772		
7	0.716	17	0.740		
8	0.748	18	0.742		
9	0.738	19	0.713		

C.6 Stratum 5

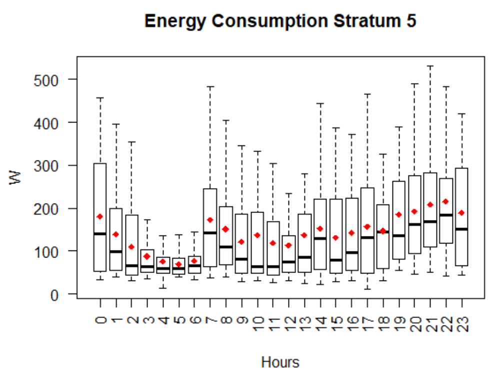

Figure C.12: Mean Consumption for Stratum 5 in a 24 Hour period

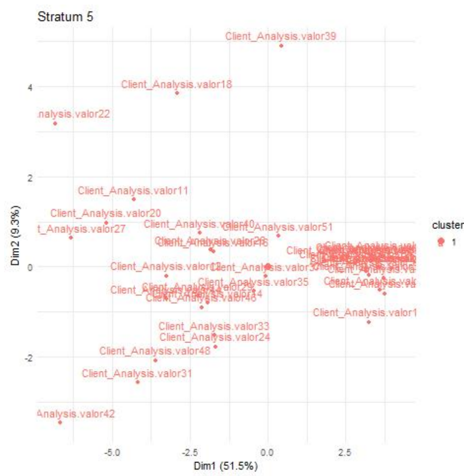

Figure C.13: Dispersion of Data

Table C.6: Variation Coefficient Stratum 5

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.738	10	0.946	20	0.619
1	0.745	11	0.969	21	0.601
2	0.767	12	0.873	22	0.610
3	0.684	13	1.036	23	0.676
4	0.622	14	0.740		
5	0.462	15	0.785		
6	0.558	16	0.721		
7	0.742	17	0.774		
8	0.726	18	0.627		
9	0.729	19	0.759		

Due to the small amount of data available for stratum 5, this curve should not be considered as the average characteristic of this group. A consumption similar to that of the other strata is visualized, verifying the same daily cycles in each case. As in the previous case, the variation coefficient tends to remain constant during the day, existing a great variation for 1 PM of 1.036.

C.7 Stratum 6

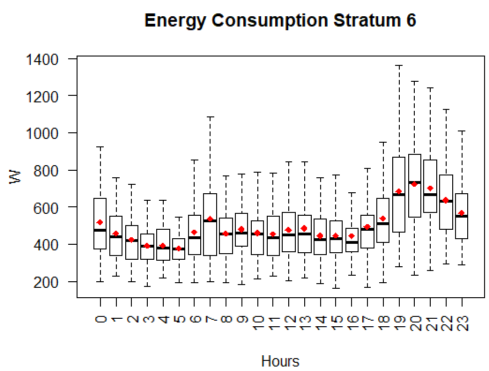

Figure C.14: Mean Consumption for Stratum 6 in a 24 Hour period

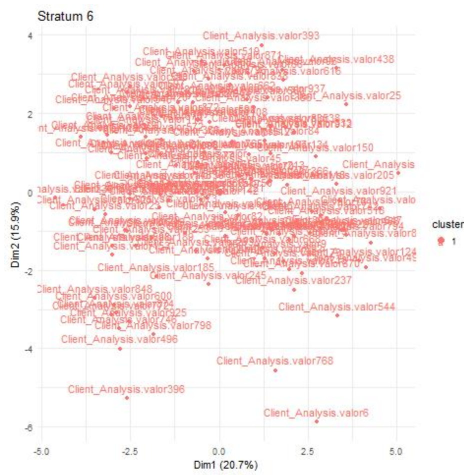

Figure C.15: Dispersion of Data

Table C.7: Variation Coefficient Stratum 6

Hour	sd/mean	Hour	sd/mean	Hour	sd/mean
0	0.326	10	0.319	20	0.323
1	0.309	11	0.331	21	0.293
2	0.298	12	0.327	22	0.311
3	0.250	13	0.341	23	0.303
4	0.259	14	0.282		
5	0.223	15	0.285		
6	0.341	16	0.300		
7	0.419	17	0.329		
8	0.316	18	0.325		
9	0.301	19	0.359		

For this group, the individuals tend to consume in a similar way, just like in the stratum 0 case; in this case, it can be seen the reduction in consumption after the 6 AM peak like in other strata.For this stratum, it is possible to justify its high consumption due to the purchasing power of the residential units cataloged as stratum 6, however, the behavior of the curve is greater in comparison with other strata; In order to perform a rigorous analysis, it will be necessary to know the most common appliances for each stratum, in order to know the reasons for their curves. The variation Coefficient remains low and constant for the day, while the boxplot reveals an increase in the consumption for all hour of the days in comparison to other strata.

C.7.1 Strata per Localities

The problem with validating the hypothesis, refers in the amount of measurements done in the different areas of the city, some of them have a lot of information, while others have few relevant data for the study.

C.7.1.1 Stratum 0

Engativa is the locality with the majority of data;the localities of Bogota have much more data in comparison with those that are in their periphery.

Table C.8: Measurements per Localities for Stratum 0

Locality	# Measurements
Puente Aranda	20664
Suba	32064
Engativa	58488
Fontibon	41688
Cogua	4056
Zipaquira	18528
Cogua Veredal	552
Rincón Santo	144

C.7.1.2 Stratum 1

Stratum 1 will have a great amount of data from Cogua, this might explain why the augment of energy starts around 4 AM a little faster in comparasion with other stratums.

Table C.9: Measurements per Localities for Stratum 1

Locality	# Measurements
Cogua	21096
Zipaquira	17256
Cogua Veredal	552
San Isidro	840
Rincon Santo	72

C.7.1.3 Stratum 2

This a stratum with a great amount of data from different locations of the city and it's surroundings. The greatest amount of data comes from Suba.

Table C.10: Measurements per Localities for Stratum 2

Locality	# Measurements
Suba	206088
Cogua	68040
Cogua Veredal	39960
Rincon Santo	11304
Zipaquira	50712
San Isidro	5832
El Olivo	2856

C.7.1.4 Stratum 3

Stratum 3 represent the biggest data-set of study, it's size is so vast, that will probably bias the whole data-set. There is a great homogeneity between the number of measurements and the number of locations, being Engativa the locality with the greatest amount of measurements.

Table C.11: Measurements per Localities for Stratum 3

Locality	# Measurements
Puente Aranda	363912
Engativa	607560
Fontibon	236136
Suba	489648
Cogua	28560
Cogua Veredal	1344
Zipaquira	20448

C.7.1.5 Stratum 4

As it can be seen in the Table C.12, Fontibon is the locality that groups the majority of meditions done for this stratum.

Table C.12: Measurements per Localities for Stratum 4

Locality	# Measurements
Fontibon	75624
Kennedy	18192
Engativa	2952
Zipaquira	792

C.7.1.6 Stratum 5

Data is only available from Suba, representing that the sample is not representative to perform a proper study.

Table C.13: Measurements per Localities for Stratum 5

Locality	# Measurements
Suba	1248

C.7.1.7 Stratum 6

Suba is the only locality with stratum 6 data; doing a comparison between the data of this stratum, with the data of stratum 5, the amount of measurements is almost 20 times more.

Table C.14: Measurements per Localities for Stratum 6

Locality	# Measurements
Suba	22536

D. Localities Analysis

As it can be seen in the Table D.1, the majority of data corresponds to the localities of Suba, Engativa, Fontibon and Puente Aranda. In order to compare the consumption between the different localities of the city, it will be necessary to identify the consumption curves of their different populations, to make a proper contrast between localities. It is important to mention, how are regions of the city with a low number of measurements, which in many cases, will not be sufficient to perform a proper load characterization; this list will include Kennedy, Rincon Santo, El Olivo and San Isidro. In general, every locality tends to consume energy in a similar manner, the correlation analysis will reveal more valuable information, that might help in the hypothesis validation.

Table D.1: Amount of Data per Locality

Locality	**Amount of Data**
Puente Aranda	384576
Suba	751884
Engativa	669000
Fontibon	353448
Cogua	121752
Kennedy	18192
Cogua Veredal	42408
Zipaquira	107736
Rincon Santo	11520
San Isidro	6672
El Olivo	2856

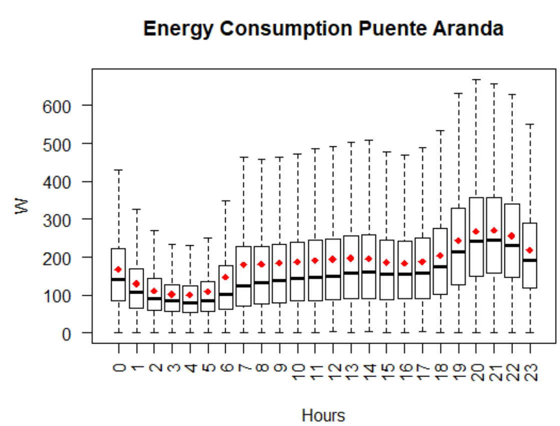

Figure D.1: Puente Aranda Energy Consumption

As in any locality, the consumption of energy tends to adapt to the routine of each family. Early in the morning, and late at night, the median of the population tends to be equal to the mean. For this population, the consumption tends to be flat for a period that goes from 7 AM to 5 PM, with an increase in the consumption until 9 PM; after 9 PM, the consumption tends to decrease until 5 AM.

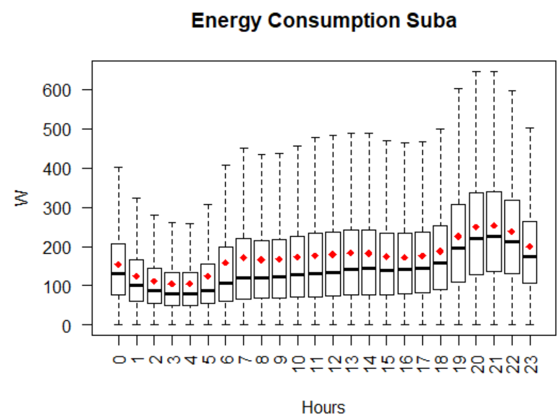

Figure D.2: Suba Energy Consumption

For the Suba locality, the variance will be an important factor, since as boxplots show, the range in which users consume is great. The population of this locality in general, rises around 6 AM, having an even consumption until the end of the afternoon. The consumption peak for this locality is between 8 PM and 9 PM.

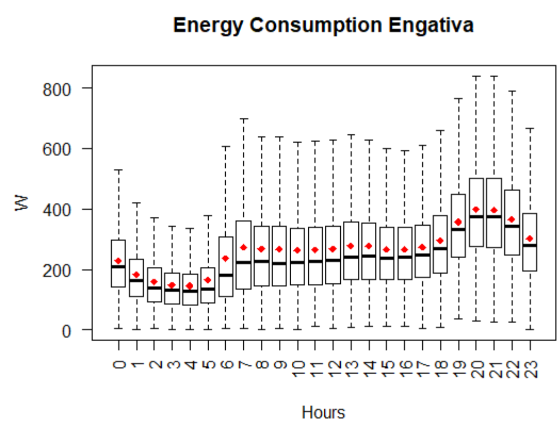

Figure D.3: Engativa Energy Consumption

Engativa consumption is presented in the figure D.3, this locality tends to have a greater consumption during the early morning, compared to other localities. As in the majority of cases, the consumption tends to be flat during the period that goes from 8 AM to 5 PM. The peaks of consumption for this locality will happen around 8 PM.

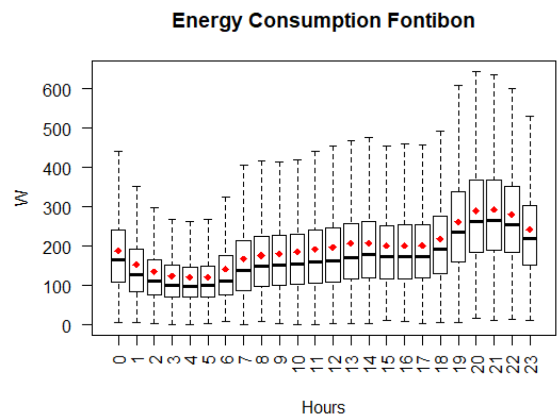

Figure D.4: Fontibon Energy Consumption

For Fontibon, the group consumption tends to be less in comparison to Engativa and Puente Aranda. For the early morning, in this locality, the mean tends to be equal to the median. For this specific locality, the boxplots reveal a more homogeneous consumption (small variance).

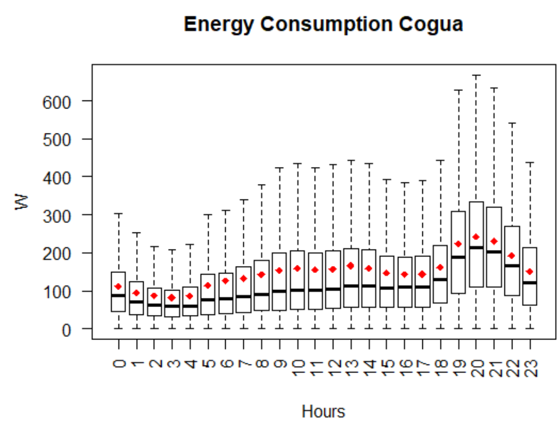

Figure D.5: Cogua Energy Consumption

Cogua represents a town near Bogota, for this reason, many of the residents of the town will have similar manners of consumption compared to any locality of the city. The peak of consumption for this town, will be around 8 PM, representing also, the greatest variance.

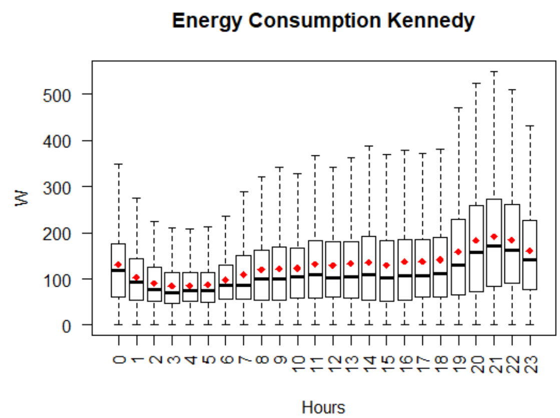

Figure D.6: Kennedy Energy Consumption

This locality located in the south of Bogota, has a small database; however, the data taken will allow to characterize in some way the consumption in it. In this case, it is verified as the population tend to get up at 5 AM in the morning, having a similar consumption throughout the day, until reaching a maximum around 9 PM. For this specific locality, the mean tends to be equal in the majority of hours, to the median.

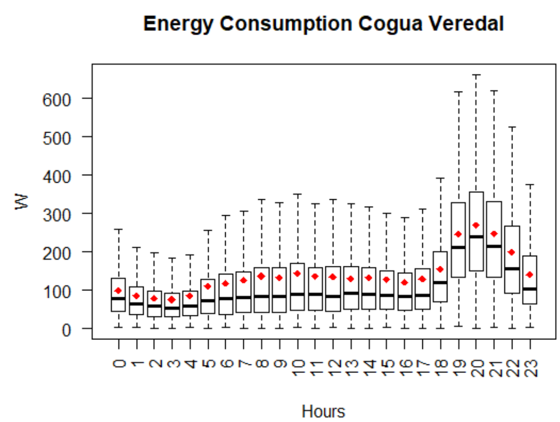

Figure D.7: Cogua Veredal Energy Consumption

Cogua Veredal is a rural population, where the majority of people work and live in farms. Working in farms might explain how the consumption increase around 5 AM, one hour earlier than in Bogota or urban Cogua. The consumption in this place tends to be lower than in the city as the boxplots show.

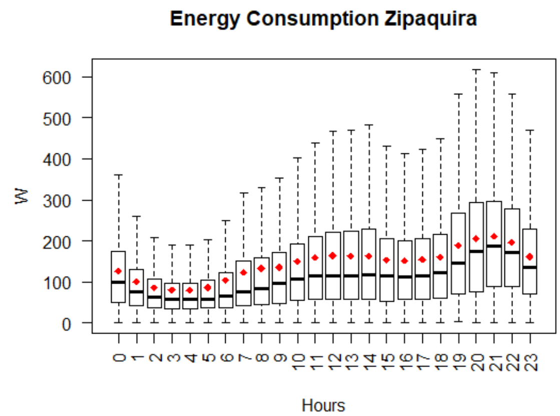

Figure D.8: Zipaquira Energy Consumption

Zipaquira consumption is different from any other locality with a similar amount of data. The consumption tends to increase around noon, to decrease until 6 PM, where it will increase until 9 PM.

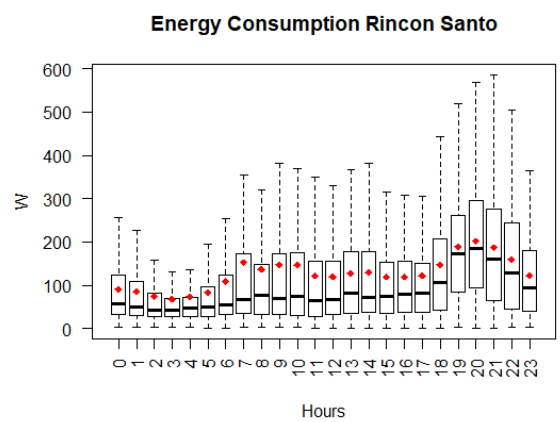

Figure D.9: Rincon Santo Energy Consumption

The variance for Rincon Santo tends to increase greatly around 7 PM. In this specific locality, during the afternoon, the mean tends to be extremely different in comparison to the median.

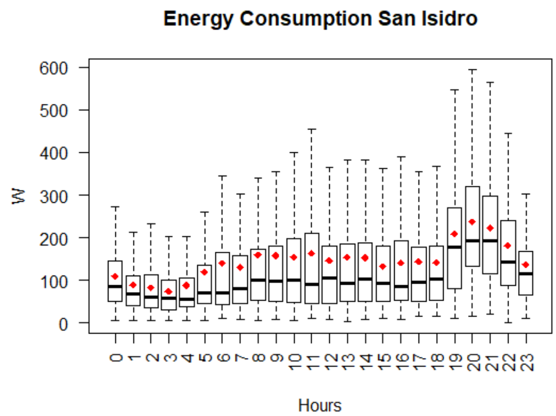

Figure D.10: San Isidro Energy Consumption

San Isidro is a locality with a few amount of available data, from the data given, it can be proven how the consumption at this locality tends to be similar to any other of the city. The variance tends to be great around 7 PM.

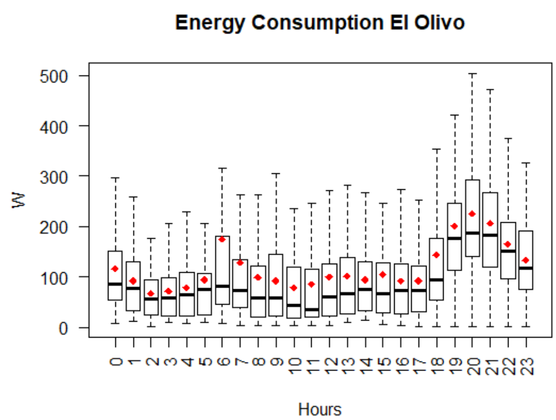

Figure D.11: El Olivo Energy Consumption

El Olivo is the locality with the smallest amount of information. This boxplot D.11 will not represent how the population actually use energy in their daily routines.

E. Weekdays Analysis Including Box-plots of Consumption.

The pattern of the consumers tends to be equal for every stratum and location of the city. In order to visualize the mean obtained after doing a k-means clustering procedure to obtain the different outliers that will removed in order to have only one group. After removing the many outliers that exists for each case, the mean of each data set is obtained. It can be seen, that for all cases, the mean is similar to the one obtained from the stratum 3 data-set, this will probably show how all the data-set is bias thanks to the criterion used for developing the measurements. The average 24 hours mean demand will be compared between each month and weekday.

E.1 November 2016

For the month of November/2016 the consumption tends to be similar between the labor days of the week (Monday to Saturday), with a small variance in the early morning, and a great variance that remains mostly constant since 7 AM. For Tuesdays, Wednesdays and Fridays the 7 AM consume is bigger in comparison to any other days of the week.
The boxplots of the days of the weeks may us identify how are the routines of the whole population of the city, for example, it's easy to see how during weekends, people tend to wake up later (8 AM for Saturdays, and 9 AM for Sundays).

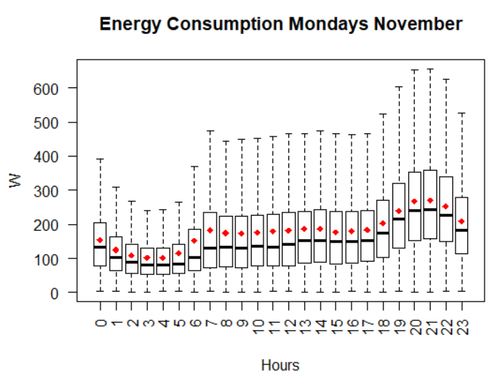

Figure E.1: Energy Consumption for Mondays/November 2016

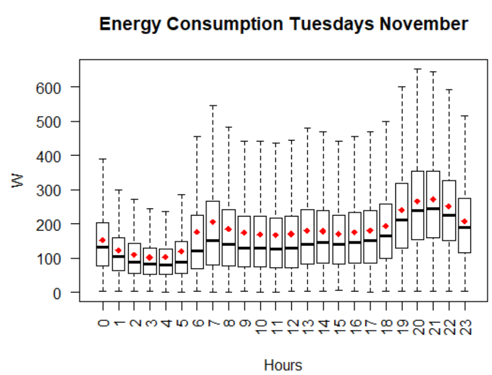

Figure E.2: Energy Consumption for Tuesdays/November 2016

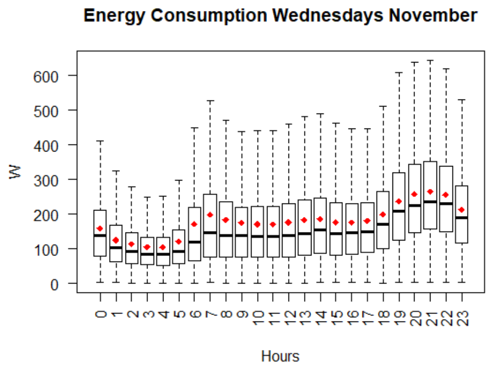

Figure E.3: Energy Consumption for Wednesdays/November 2016

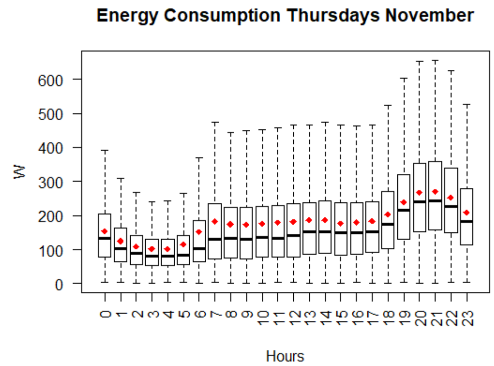

Figure E.4: Energy Consumption for Thursdays/November 2016

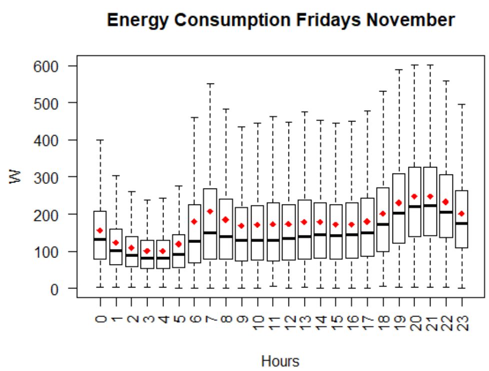

Figure E.5: Energy Consumption for Fridays/November 2016

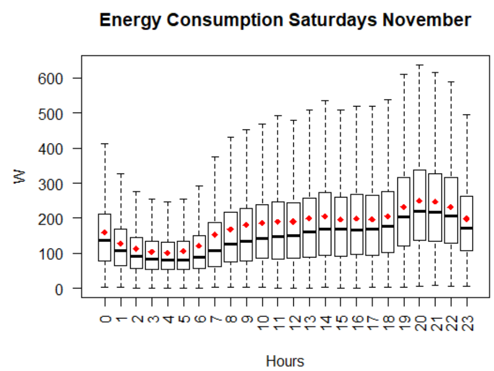

Figure E.6: Energy Consumption for Saturdays/November 2016

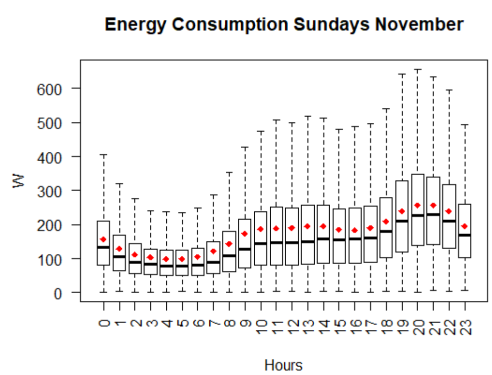

Figure E.7: Energy Consumption for Sundays/November 2016

Table E.1: Correlation Between Weekdays for November

Weekdays	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Monday	1.000	0.982	0.989	0.979	0.974	0.948	0.929
Tuesday	0.982	1.000	0.997	0.998	0.991	0.882	0.846
Wednesday	0.989	0.997	1.000	0.995	0.990	0.904	0.871
Thursday	0.979	0.998	0.995	1.000	0.994	0.881	0.841
Friday	0.974	0.991	0.990	0.994	1.000	0.885	0.838
Saturday	0.948	0.882	0.904	0.881	0.885	1.000	0.982
Sunday	0.929	0.846	0.871	0.841	0.838	0.982	1.000

The correlation coefficient between days tend to be great for all days, decreasing for any pair between any day an Sunday. As it can be seen in E.1, the weekend tends to have a different pattern in comparison with any other day, this can be verified in the boxplots for the weekdays of November. For Saturdays and Sundays, the increase in energy consumption stars around 8 AM for Saturdays and at 9 AM for Sundays, while during normal labor days, the increase in consumption stars around 7 AM.

Table E.2: Mean Total Energy Consumption Characteristics for November by User

Characteristic	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Average Demand W	178.209	178.546	179.858	179.860	175.703	177.909	173.203
Maximum Demand W	269.954	272.382	265.326	269.227	247.503	250.460	256.164
Hour of Maximum Demand	21:00	21:00	21:00	21:00	20:00	20:00	20:00
Load Factor	0.660	0.655	0.678	0.668	0.710	0.710	0.676

The average demand and load factor, tends to be similar for all cases, increasing on the days of the middle of the week. As it can be seen at Table E.2, the lowest amount of average demand is for Sundays, while Thursdays tends to have the higher amount. An interesting result will be the decrease in the hour in which the peak occurs, going from 9 at night to 8 at night for weekend days including Friday; this behavior will be due to workers arriving at their homes earlier than on other days.

E.2 December 2016

For the month of December/2016 the consumption didn't variate much in comparison to the previous month. As in the previous case, the variance tends to increase after 6 AM as the boxplots show.For this month, people tend to consume a little more at night due to the festivities, and this will be reflected in the data,in this specific case, we see a maximum of consumption for Sunday at 9 PM, one hour higher compared to last month.

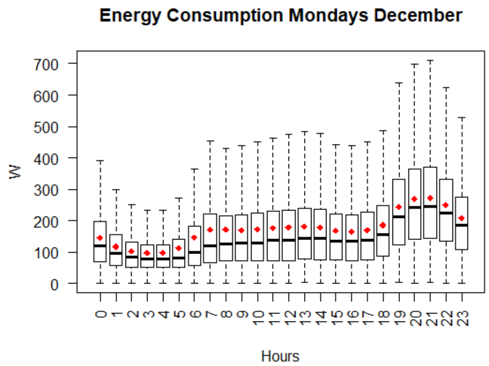

Figure E.8: Energy Consumption for Mondays/December 2016

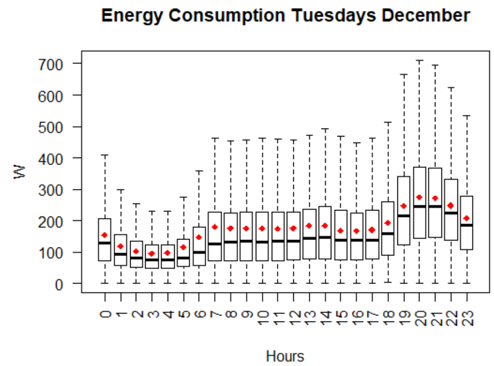

Figure E.9: Energy Consumption for Tuesdays/December 2016

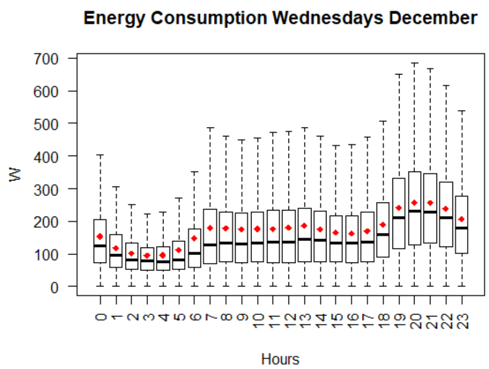

Figure E.10: Energy Consumption for Wednesdays/December 2016

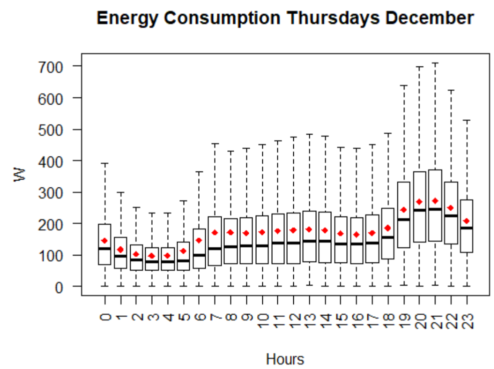

Figure E.11: Energy Consumption for Thursdays/December 2016

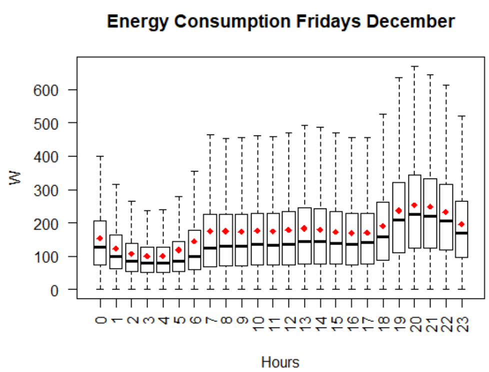

Figure E.12: Energy Consumption for Fridays/December 2016

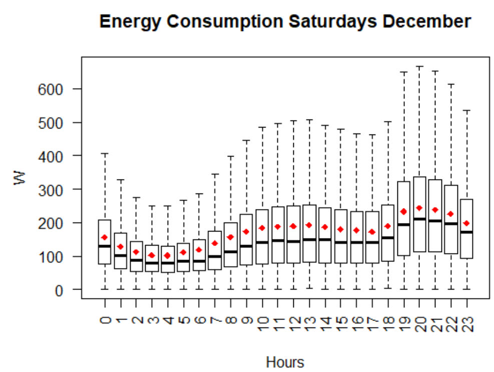

Figure E.13: Energy Consumption for Saturdays/December 2016

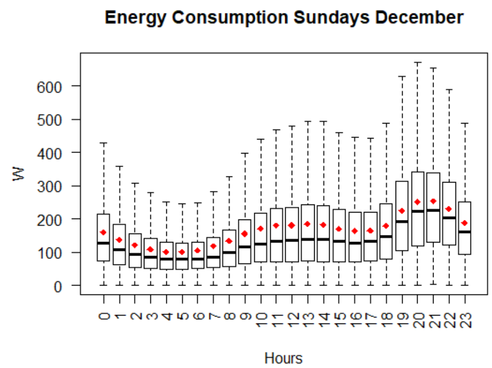

Figure E.14: Energy Consumption for Sundays/December 2016

Table E.3: Correlation Between Weekdays for December

Weekdays	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Monday	1.000	0.998	0.994	0.996	0.993	0.951	0.928
Tuesday	0.998	1.000	0.996	0.995	0.996	0.947	0.918
Wednesday	0.994	0.996	1.000	0.995	0.997	0.952	0.910
Thursday	0.996	0.995	0.995	1.000	0.995	0.969	0.942
Friday	0.993	0.996	0.997	0.995	1.000	0.961	0.922
Saturday	0.951	0.947	0.952	0.969	0.961	1.000	0.973
Sunday	0.928	0.918	0.910	0.942	0.922	0.973	1.000

During December, all days have a great correlation between each other, this will be with no exceptions.Great correlation between every day, something that remains invariant since December.This could be because December is a holiday season, where children are usually no longer studying, spending more time at home consuming energy every day of the week.

Table E.4: Mean Total Energy Consumption Characteristics for December by User

Characteristic	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Average Demand W	173.286	175.247	172.514	173.663	172.265	170.747	165.799
Maximum Demand W	272.224	274.490	257.115	263.230	253.224	242.897	253.677
Hour of Maximum Demand	21:00	20:00	20:00	20:00	20:00	20:00	21:00
Load Factor	0.637	0.638	0.671	0.660	0.680	0.703	0.654

There is an interesting small decrease in the maximum demand and average demand for the days of December, this might be thanks to the amount of families that decide to travel to other places around this time. The load factor during Saturdays represents the maximum for all cases. The Hour of Maximum Demand tends to be 8 PM for all week, excepting Sundays and Mondays where it is 9 PM, this again, might be thanks to the holidays.

E.3 January 2017

For the month of January/2017, there was a decrease in the consumption of energy during the day in comparison to the two previous month studied, significantly reducing the load in the afternoons of the days Mondays, Tuesdays, Wednesdays and Thursdays.The load could be reduced due to the large number of people who are not in the city, or because many homes are not completely inhabited.Another advantage of the smart meter, is that with the help of this, you can inform how many people were living in the home on a specific day.

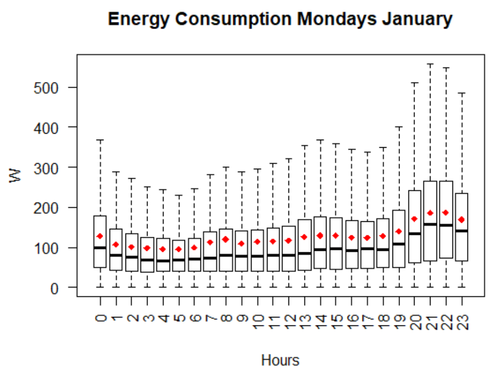

Figure E.15: Energy Consumption for Mondays/January 2017

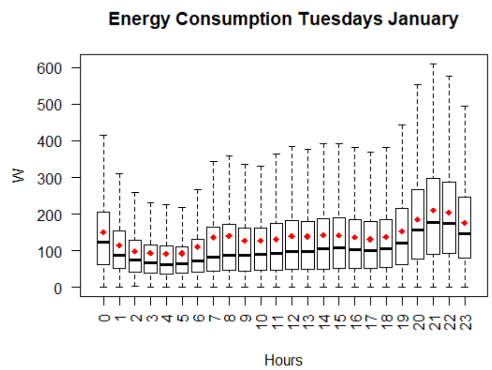

Figure E.16: Energy Consumption for Tuesdays/January 2017

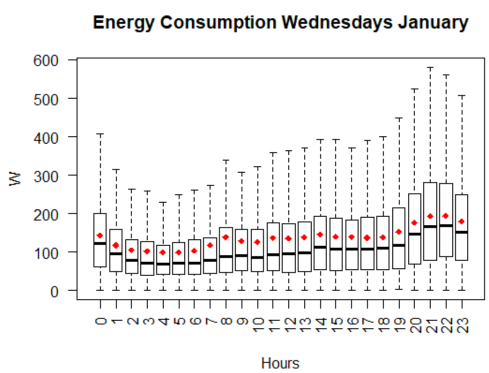

Figure E.17: Energy Consumption for Wednesdays/January 2017

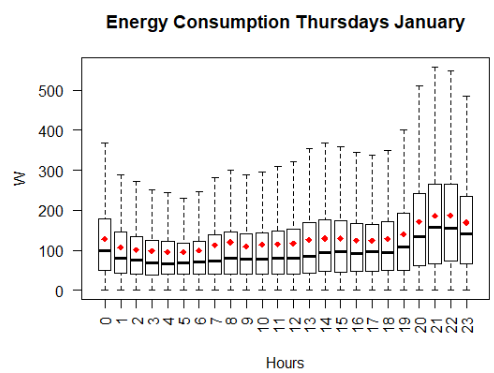

Figure E.18: Energy Consumption for Thursdays/January 2017

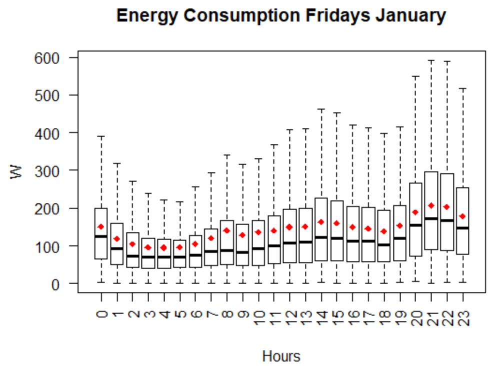

Figure E.19: Energy Consumption for Fridays/January 2017

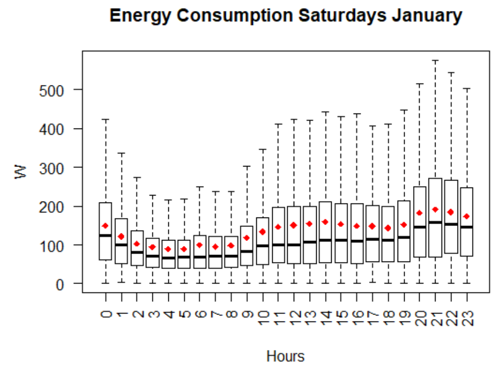

Figure E.20: Energy Consumption for Saturdays/January 2017

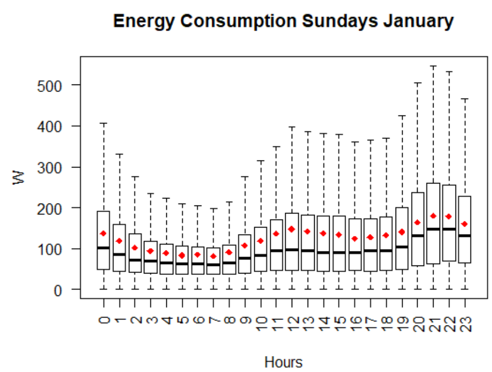

Figure E.21: Energy Consumption for Sundays/January 2017

Table E.5: Correlation Between Weekdays for January

Weekdays	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Monday	1.000	0.970	0.979	0.977	0.950	0.874	0.882
Tuesday	0.970	1.000	0.980	0.978	0.969	0.869	0.865
Wednesday	0.979	0.980	1.000	0.977	0.981	0.913	0.911
Thursday	0.977	0.978	0.977	1.000	0.974	0.916	0.905
Friday	0.950	0.969	0.981	0.974	1.000	0.942	0.922
Saturday	0.874	0.869	0.913	0.916	0.942	1.000	0.978
Sunday	0.882	0.865	0.911	0.905	0.922	0.978	1.000

The Table E.5 shows how the correlation between weekdays tends to be great for the majority of cases, showing a great contrast between the consumption of energy in the first two days of the week, compared to the last days of it.

Table E.6: Mean Total Energy Consumption Characteristics for January by User

Characteristic	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Average Demand W	126.344	138.230	137.414	135.145	142.705	136.803	125.835
Maximum Demand W	187.684	210.208	193.907	196.327	206.782	190.992	180.840
Hour of Maximum Demand	22:00	22:00	22:00	22:00	22:00	22:00	22:00
Load Factor	0.673	0.658	0.709	0.688	0.690	0.716	0.696

The average 24 hours demand is the smallest in all the data set, this could be linked to the reduction of high consumption that happens around the end of every year. Another factor that might be happened in the decrease of the average demand, might be related with the mentality of saving money after wasting it during December. Suddenly, for this month the hour of maximum demand was at 10 PM for all days of the week, showing a pattern that wasn't seen in the previous month.

E.4 February 2017

For the month of February/2017, the average consumption increased a lot in comparison to the previous month; in many cases, it even duplicated. The variance also increased greatly during the morning, with a huge increase around 7 AM.There is evidence of an increase in cargo throughout the day, including early morning hours. It is possible that due to the increase in temperatures, the refrigerators of the residences have consumed a greater amount of energy to operate.The highest energy consumption is evident for the middle of the week, otherwise, the weekend will have the lowest consumption of energy.

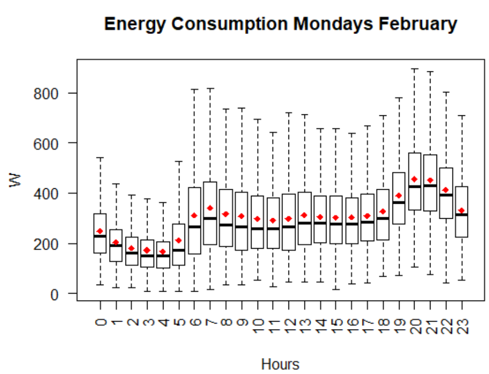

Figure E.22: Energy Consumption for Mondays/February 2017

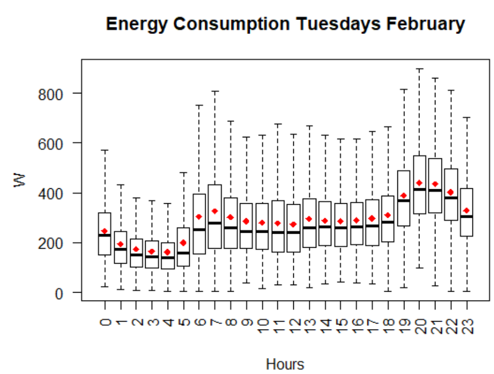

Figure E.23: Energy Consumption for Tuesdays/February 2017

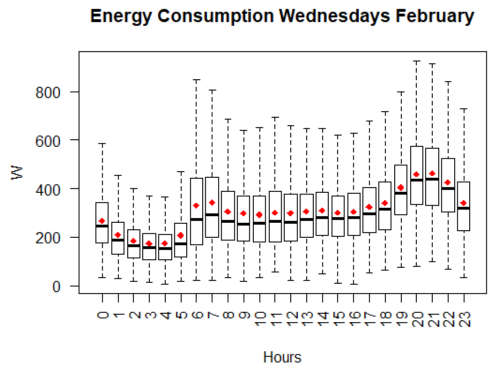

Figure E.24: Energy Consumption for Wednesdays/February 2017

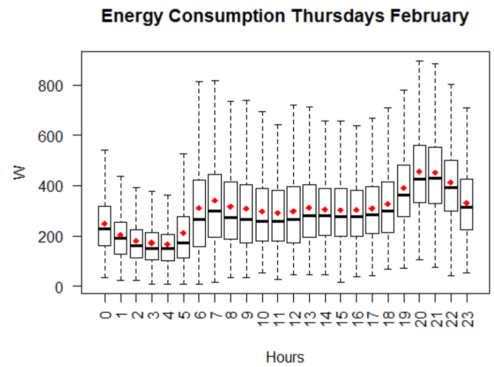

Figure E.25: Energy Consumption for Thursdays/February 2017

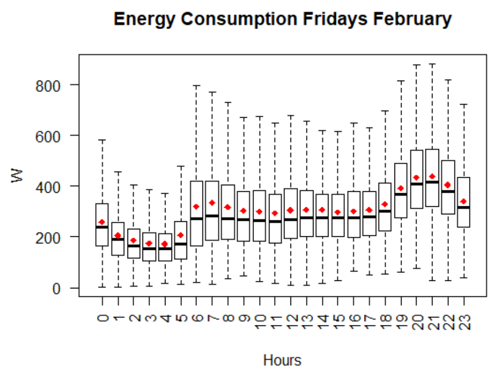

Figure E.26: Energy Consumption for Fridays/February 2017

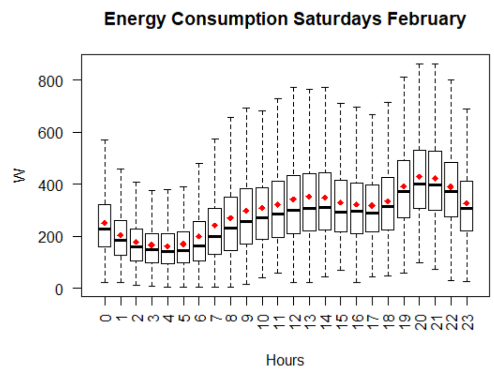

Figure E.27: Energy Consumption for Saturdays/February 2017

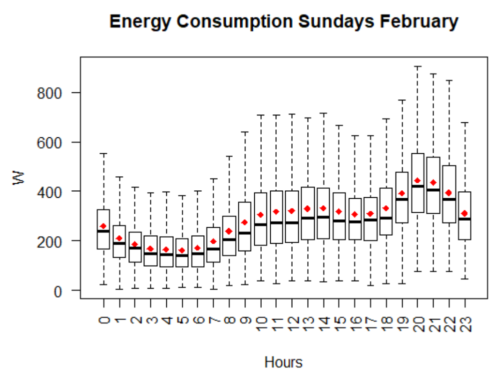

Figure E.28: Energy Consumption for Sundays/February 2017

February is the month with the lest amount of outliers for each day of the week compared with the pass months. The boxplots shows how is an increase in the consumption of energy for this month for the last hours of the day.

Table E.7: Correlation Between Weekdays for February

Weekday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Monday	1.000	0.997	0.994	0.997	0.997	0.880	0.845
Tuesday	0.997	1.000	0.997	0.998	0.996	0.870	0.840
Wednesday	0.994	0.997	1.000	0.998	0.995	0.872	0.845
Thursday	0.997	0.998	0.998	1.000	0.997	0.877	0.849
Friday	0.997	0.996	0.995	0.997	1.000	0.881	0.843
Saturday	0.880	0.870	0.872	0.877	0.881	1.000	0.984
Sunday	0.845	0.840	0.845	0.849	0.843	0.984	1.000

During this month, there is a strong correlation between labor days, while for the weekend, the correlation between Sunday or Saturday with any labor day is weak.

Table E.8: Mean Total Energy Consumption Characteristics for February by User

Characteristic	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Average Demand W	301.677	289.578	306.756	303.893	301.425	294.923	287.080
Maximum Demand W	455.423	439.755	463.393	453.811	439.681	429.246	444.768
Hour of Maximum Demand	20:00	20:00	21:00	21:00	21:00	20:00	20:00
Load Factor	0.662	0.658	0.662	0.670	0.686	0.687	0.645

Surprisingly the average demand and maximum demand increased in an abrupt way for the month of February, in some cases it duplicated. This can be linked with the increase in the temperature that the city lived during this month. The maximum demand as all cases, was around night.

F. Geographical Analysis Using Leaflet and ggmap.

F.1 November

For this month, the mean power consumed at a specif hour at different localities was studied. As it can be seen, in Figures F.1,F.2,F.3, the different localities tend to pattern in similar ways across the day. The consume tends to be similar between localities and strata.

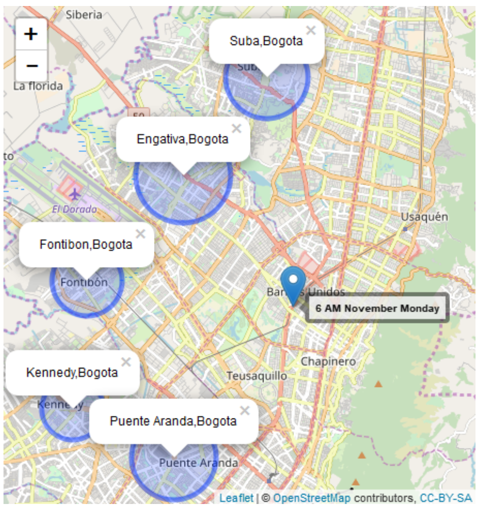

Figure F.1: Mean Consumption for Mondays Nov/16 6 AM

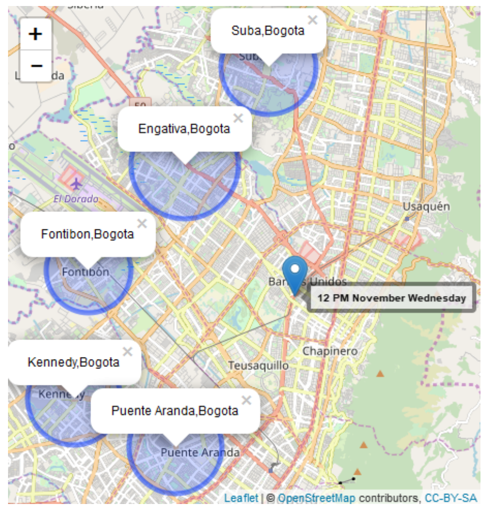

Figure F.2: Mean Consumption for Wednesdays Nov/16 12 PM

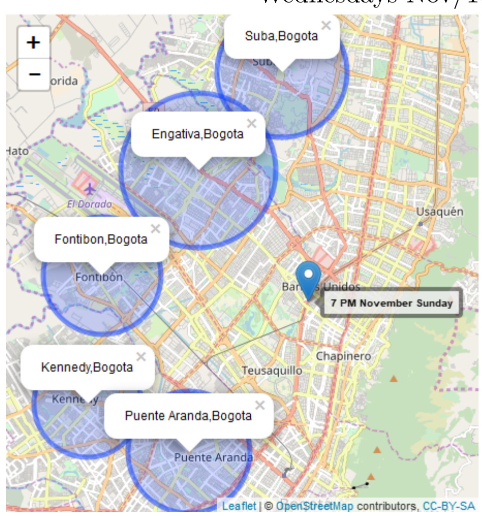

Figure F.3: Mean Consumption for Sundays Nov/16 7 PM

Engativa for all cases, was the locality with the greatest mean of W, this can be seen in the maps created.

Table F.2: Mean Consumption in Watts for the Weekdays and Hours of Study November 2016

Day	Saturday			Sunday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Engativa	140.79	217.34	268.88	121.34	224.97	285.16
Suba	128.97	192.18	233.24	100.39	189.76	240.00
Kennedy	92.06	166.20	180.45	87.40	148.26	199.62
Fontibon	108.61	168.02	211.50	104.38	157.33	216.35
Puente Aranda	116.06	190.83	226.59	99.45	200.45	228.05
Cogua Veredal	No Data	No Data	No Data	114.50	125.00	156.50
Cogua	62.92	126.92	225.15	92.07	194.64	119.14
Zipaquira	83.00	63.00	456.00	121.00	42.00	543.00

Table F.1: Mean Consumption in Watts for the Weekdays and Hours of Study November 2016

Day	Monday			Wednesday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Engativa	175.70	202.46	284.60	197.73	197.09	280.22
Suba	152.28	170.91	232.45	178.82	176.67	228.06
Kennedy	114.05	156.13	168.97	130.22	170.96	197.72
Fontibon	129.87	165.11	215.68	144.68	155.19	212.21
Puente Aranda	156.56	191.41	237.37	175.02	174.36	226.34
Cogua Veredal	115.00	381.00	110.00	No Data	No Data	No Data
Cogua	117.31	219.77	248.46	94.79	175.95	215.42
Zipaquira	66.00	56.00	176.00	No Data	No Data	No Data

Table F.3: Coefficient of Variation November

Hour	06:00 a. m.	12:00 p. m.	07:00 p. m.
Monday	0.263	0.470	0.262
Wednesday	0.245	0.076	0.125
Saturday	0.257	0.396	0.355
Sunday	0.121	0.359	0.520

The coefficient variation between mean data for localities tends to be small for all cases.

F.2 December

As in the previous case, all loclities tend to consume in similar ways depending the hour of the day. Rincon Santo and El Olivo tend to be the ones with the smallest mean for the majority of hours and days studied. Engativa as in the previous case, is the locality with the biggest amount of consume in Watts.

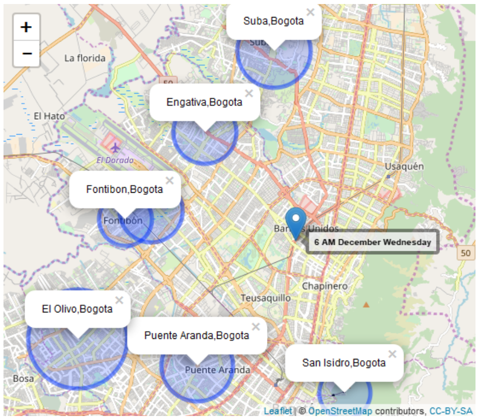

Figure F.4: Mean Consumption for Wednesdays Dec/16 6 AM

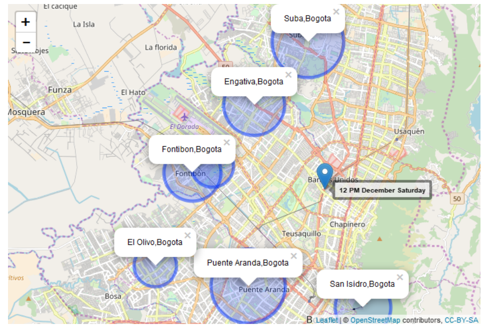

Figure F.5: Mean Consumption for Saturdays Dec/16 12 PM

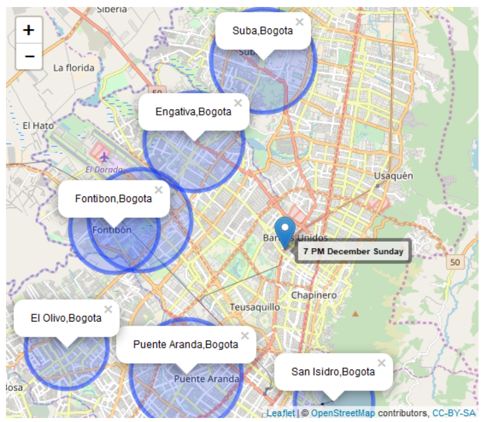

Figure F.6: Mean Consumption for Sundays Dec/16 7 PM

Table F.4: Mean Consumption in W for the Weekdays and Hours of Study December 2016

Day	Monday			Wednesday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Fontibon	121.04	145.32	208.18	119.54	146.33	195.88
Puente Aranda	161.44	200.61	262.52	162.37	203.71	264.79
Suba	157.24	202.41	249.18	164.01	199.26	245.27
Cogua	153.70	175.14	272.11	151.51	173.46	277.42
Rincón Santo	194.56	123.92	240.36	143.29	110.29	249.29
Zipaquira	184.66	210.45	253.83	157.10	236.17	338.17
Engativa	151.62	187.09	277.95	141.48	195.67	268.37
El Olivo	127.10	131.30	206.00	220.10	61.90	203.60
San Isidro	126.78	154.72	214.06	115.29	192.82	275.41
Cogua Veredal	124.28	114.55	246.12	130.12	154.43	256.01

Table F.5: Mean Consumption in W for the Weekdays and Hours of Study December 2016

Day	Saturday			Sunday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Fontibon	99.84	161.81	195.27	94.26	153.29	197.93
Puente Aranda	126.17	216.74	262.11	113.76	204.74	249.54
Suba	128.49	205.71	239.86	109.08	197.44	232.64
Cogua	129.37	189.90	268.46	99.33	176.54	248.48
Rincón Santo	93.90	123.00	216.97	88.88	140.59	229.78
Zipaquira	126.85	240.67	259.62	105.43	195.30	222.51
Engativa	133.27	176.68	227.04	143.25	177.95	221.03
El Olivo	180.55	122.09	170.00	75.36	79.93	182.93
San Isidro	92.78	160.48	193.35	74.89	150.89	178.11
Cogua Veredal	113.76	139.96	247.01	98.16	157.07	204.27

Table F.6: Coefficient of Variation December

Hour	06:00 a. m.	12:00 p. m.	07:00 p. m.
Monday	0.171	0.214	0.106
Wednesday	0.197	0.305	0.154
Saturday	0.208	0.229	0.146
Sunday	0.198	0.224	0.116

The variation coefficient for the data is small, reveling how in general all localities tend to consume in a similar manner. There is evidence of a growth in the average consumption in El Olivo and San Isidro sectors.

F.3 January

For the month of January the average consumption tended to differ too much between localities.This strange behavior may be due to the fact that it is the month with the least amount of data, and in addition, by including San Isidro as a locality to be evaluated, the way in which the data are distributed will be affected. San Isidro is a town with few data, so the risk of bias is high, causing what will be seen below:

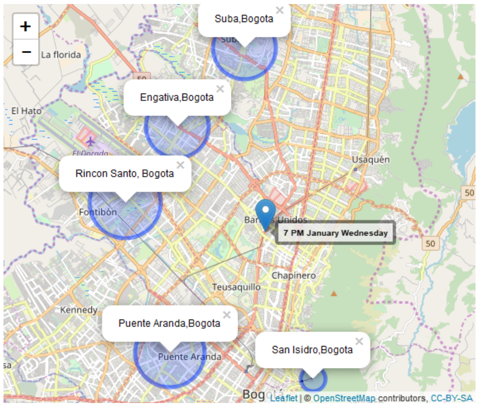

Figure F.7: Mean Consumption for Wednesdays Jan/17 7 PM

Figure F.8: Mean Consumption for Saturdays Jan/17 7 PM

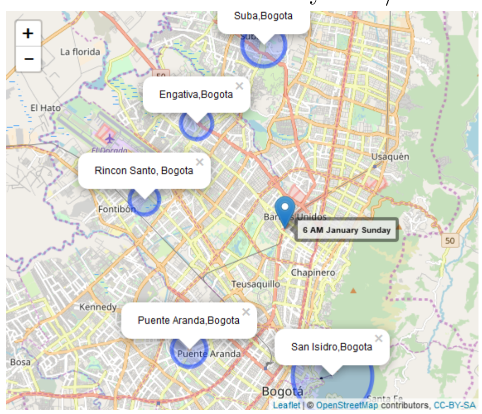

Figure F.9: Mean Consumption for Sundays Jan/17 6 AM

Table F.7: Mean Consumption in W for the Weekdays and Hours of Study January 2017

Day	Monday			Wednesday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Rincón Santo	182.00	132.86	151.29	117.25	50.50	164.75
Suba	117.04	105.15	143.73	122.59	112.79	147.26
Puente Aranda	83.00	134.43	136.05	89.88	155.08	160.12
Zipaquira	89.02	114.71	131.84	93.44	129.46	151.60
Cogua	122.44	133.55	146.66	126.72	172.83	155.79
Engativa	70.59	121.69	148.13	77.38	167.97	145.99
Cogua Veredal	156.57	156.00	168.43	94.33	72.50	166.00
San Isidro	462.00	43.00	588.00	445.00	54.00	58.67

San Isidro for some specifics days of the week, and at specific hours, have an increase in it mean, this is representing an anomaly that requires further study from **CODENSA**.

Table F.8: Mean Consumption in W for the Weekdays and Hours of Study January 2017

Day	Saturday			Sunday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Rincón Santo	60.77	97.77	146.77	71.59	155.95	142.50
Suba	108.12	142.72	158.49	99.40	135.24	145.24
Puente Aranda	83.86	154.40	138.82	83.11	153.58	138.35
Zipaquira	93.46	166.44	128.25	75.81	140.77	137.31
Cogua	126.63	149.81	165.65	69.09	208.88	136.84
Engativa	78.33	165.28	169.81	74.26	156.47	136.04
Cogua Veredal	105.25	156.31	190.31	71.13	124.44	188.06
San Isidro	314.67	48.00	75.00	185.50	388.00	45.50
El Olivo	142.00	230.50	236.00	No Data	No Data	No Data

Table F.9: Coefficient of Variation January

Hour	06:00 a. m.	12:00 p. m.	07:00 p. m.
Monday	0.795	0.287	0.775
Wednesday	0.837	0.439	0.244
Saturday	0.612	0.343	0.281
Sunday	0.431	0.473	0.296

Due to the strange pattern of some localities, the coefficient of variation between localities tend to be greater in comparsion with any other month. This might be thanks to weird mean of consumption of the San Isidro zone.

F.4 February

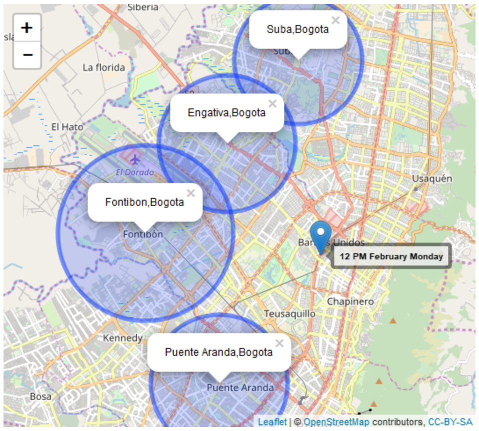

Figure F.10: Mean Consumption for Mondays Feb/17 12 PM

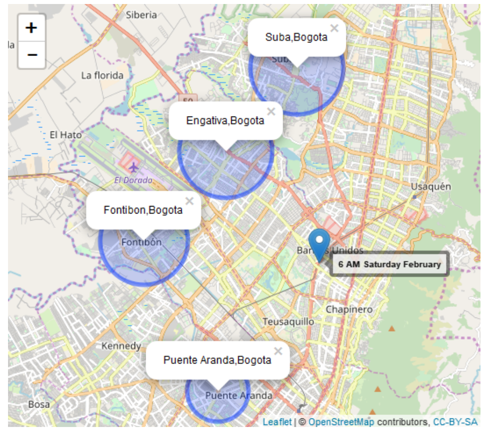

Figure F.11: Mean Consumption for Saturdays Feb/17 6 AM

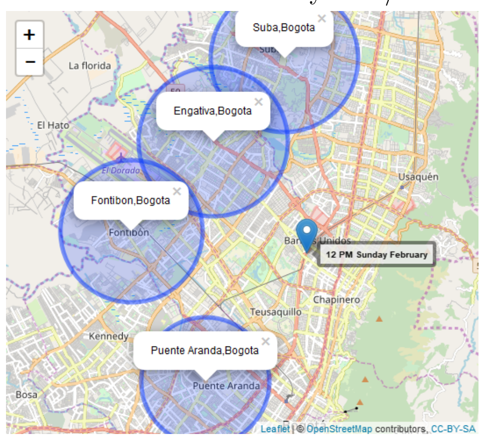

Figure F.12: Mean Consumption for Sundays Feb/17 12 PM

Table F.10: Mean Consumption in W for the Weekdays and Hours of Study February 2017

Day	Monday			Wednesday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Zipaquira	217.87	302.83	419.83	171.69	325.36	445.83
Engativa	324.39	298.10	398.91	175.45	298.60	409.57
Suba	294.22	280.84	356.14	171.92	286.02	382.23
Fontibon	224.39	381.25	385.36	176.36	355.25	394.66
Cogua	270.75	323.25	377.50	119.00	520.00	568.00
Puente Aranda	311.73	298.18	324.91	149.77	244.68	340.41

Table F.11: Mean Consumption in W for the Weekdays and Hours of Study February 2017

Day	Saturday			Sunday		
Locality	06:00 a. m.	12:00 p. m.	07:00 p. m.	06:00 a. m.	12:00 p. m.	07:00 p. m.
Zipaquira	156.13	338.97	436.50	153.21	312.41	330.74
Engativa	204.00	348.34	397.05	179.35	323.92	401.32
Suba	201.79	320.30	377.31	163.47	320.24	378.20
Fontibon	189.00	364.92	370.62	193.23	307.40	372.90
Cogua	175.00	267.50	285.00	91.67	391.67	460.67
Puente Aranda	135.31	291.15	345.42	105.62	280.10	332.62

As it was mentioned before, the consumption in the month of February skyrocketed, possibly due to the high temperatures registered during that period of time. All locations compared to past data, tended to consume more.In this specific month, it is evident that it is important to take into account variables such as temperature in the load characterization of a system. The hot months will represent a greater consumption of electrical energy, since the refrigeration processes;they will require a greater amount of power; In the cases of houses that have refrigeration systems, an increase in the load will be seen due to this.

Table F.12: Coefficient of Variation February

Hour	06:00 a. m.	12:00 p. m.	07:00 p. m.
Monday	0.163	0.113	0.088
Wednesday	0.141	0.285	0.185
Saturday	0.153	0.114	0.138
Sunday	0.275	0.115	0.127

The greatest increase in consumption was registered in the areas surrounding Bogota, such as Cogua and Zipaquira. The coefficient of variation remained low, due to the fact that there were no cases such as San Isidro that occurred in the previous month.