PRECON: Pakistan Residential Electricity Consumption Dataset

Ahmad Nadeem, Naveed Arshad
Lahore University of Management Sciences
Lahore, Pakistan
[16060026,naveedarshad]@lums.edu.pk

ABSTRACT

Buildings consume on average of over 40% of energy throughout the world[1]. Therefore, it is crucial to fully understand the consumption behaviour of building occupants for energy efficiency, efficient load balancing and better demand-side management. To this end, small number of datasets are available from developing countries, particularly South Asia, that can model consumption behaviours of a wide range of residential electricity users. In this paper, we present **PRECON** dataset, collected over a period of one year, of electricity consumption patterns for 42 residential properties having varied demographics. Data is collected for the whole house consumption and from high powered devices as well as from major areas of the building. This dataset can play a pivotal role for distribution companies and policymakers to use data-driven optimization of generation, perform better demand-side management and improve energy efficiency.

KEYWORDS

PRECON, Electricity, Consumption, Dataset

ACM Reference Format:

Ahmad Nadeem, Naveed Arshad. 2019. PRECON: Pakistan Residential Electricity Consumption Dataset. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems (e-Energy '19), June 25–28, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3307772.3328317

1 INTRODUCTION

Several countries across the globe are striving to shift their entire electricity generation to renewable resources [6, 12, 15]. However, the electricity supply and demand from such resources is often variable and intermittent. To run the electrical power systems on these resources, we need a concrete understanding of both generation and demand.

Renewable share in the power sector is on the rise, and with more solar PV and wind energy installations, the variability and intermittency will only increase. In such a scenario, the share of guaranteed dispatchable energy will decrease. Thus, it is important to control the energy demand according to the availability of energy from renewable resources using better pricing or innovative demand-side management. To this end, insight into consumer behaviour is important. Traditional smart meters only provide data at a granularity of 15 minutes or more. A lot of information related to energy consumption behaviour is not captured at this granularity. Therefore, data is needed at a finer scale where better insights can be derived on consumer behaviour.

The widespread availability of data at a finer scale is not possible. Therefore, a representative sample is typically employed. In this paper, we present a dataset where we have collected data from 42 residential buildings at one-minute interval. Not only the data of the whole building is captured but also the energy utilization of high powered devices as well as the energy consumption of different areas of the properties is captured in the dataset.

The aim of this data collection exercise is to understand the residential electricity consumption profiles of households in developing countries where the energy market is flourishing. Noor et al. [18] argues that extensive power system planning is required due to the unidirectional relation between GDP and electricity consumption of South Asian countries. As per capita electricity usage in these areas is increasing, a lot of research is required to keep up.

Several datasets have been collected previously, but few of these are available for public use. To the best of our knowledge, there is no dataset available for developing countries which has such a vast and demographically varied sample size as PRECON. The aim of this data collection and processing exercise is to understand the electricity consumption patterns of users in the developing world. A sound realization of consumption patterns can help in the development of intelligent smart grids and better demand-side management tools.

This paper provides details of electricity consumption data and meta-data of houses having varied demographics such as monthly income of the household, number of people in the house, area of the house and the built year of the house. Also, we have provided details on high electricity consumption devices, various rooms and the load profile of the whole house. During the data collection process, several problems common to the developing world were encountered. These problems included lack of standardized wiring in the households, multiple conventional electricity meters in each house provided by the utility and wiring issues due to alternate energy sources such as generators and Uninterruptable Power Supplies (UPSs).

Section 2 of this paper provides details on other datasets collected on electricity consumption behaviours and describes how PRECON is different. Then, section 3 introduces the process of data collection. It explains how the data is collected and processed. Section 4 describes various details of the data collected for this paper. The paper culminates by providing a conclusion to the work described in this paper.

e-Energy '19, June 25–28, 2019, Phoenix, AZ, USA

ACM ISBN 978-1-4503-6671-7/19/06.
https://doi.org/10.1145/3307772.3328317

2 RELATED WORK

In recent years, a lot of research has focused on understanding behaviours of electricity consumers. Such research has become more significant after the adoption of wind and solar PV as generation resources which are inherently intermittent. A comprehensive energy consumption dataset is required to better model and forecast electricity consumption behaviours. UK-DALE [10] is one such dataset. In this dataset, energy consumption of five households in the UK is recorded at 16 kHz for varying durations. The duration of data collection for these households varies from one month to two years. Additionally, only one household in the respective dataset has four occupants while all the remaining households have two occupants.

Smart [2] is another such dataset which provides details of electricity consumption for seven residential properties located in Western Massachusetts, USA. In this dataset, instead of installing smart meters at several homes, the focus is on obtaining detailed data for a limited number of households. The dataset contains electricity consumption data at 1/900 Hz sampling rate, along with several other details such as ambient temperature, humidity and several binary events such as opening and closing of doors and occupancy of rooms monitored through motion sensors.

Dataport database [19] is another source of a similar dataset containing electricity consumption data of 700 households for more than four years. However, this dataset is not publically available, and the monitored households are located in Texas, USA which reduces its utility for understanding load profiles of residential electricity consumers in developing regions like South Asia.

One dataset that is of particular interest is iAWE[3], collected in 2013. It contains data from a single household in New Delhi, India for 73 days. However, this dataset does not captures the energy usage patterns of households over the whole year. Several other datasets have been published in the last decade such as BLUED[9], for a single household in Pittsburgh, USA; AMPds[13], for a single household in Canada; GREEND [16] which recorded 9 households for a year; RAE [14], DRED[22], REFIT[17] and ECO[4] all of which recorded electricity consumption in developed countries like Canada, Netherlands, UK and Switzerland respectively.

PRECON is different from all these datasets in three aspects; location, duration and number of households monitored. Figure 1 illustrates the difference between PRECON and other publically available energy consumption datasets. PRECON stands out from the rest of the datasets because it monitors a greater number of households than any other similar dataset. It is also important to notice that only two datasets are available that are moinitoring residential electrciity consumption and are related to developing countries.

Figure 2 represents the spatial distribution of households for which the data is collected in PRECON. The selected residential properties are scattered all over the city, which enables us to cover various types of households, varying in financial status, daily activities and usage pattern of home appliances.

3 DATA COLLECTION PROCESS

The first task in data collection process is to select locations for installation of smart meters. Lahore is selected as the location for all smart meter installations. It is the second largest city and is also the cultural hub of Pakistan with a population of more than 10 million.

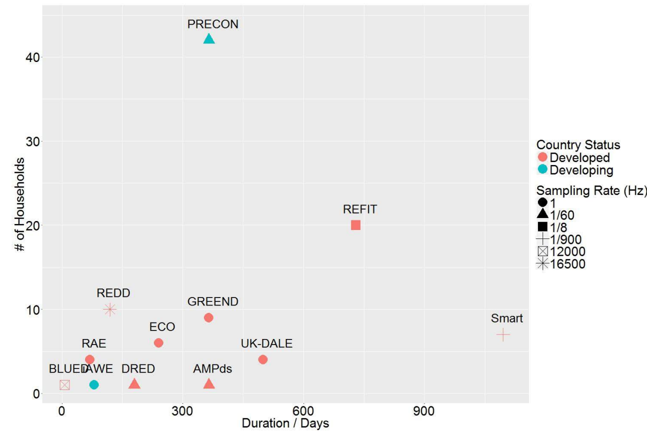

Figure 1: Attributes of PRECON and other similar datasets.

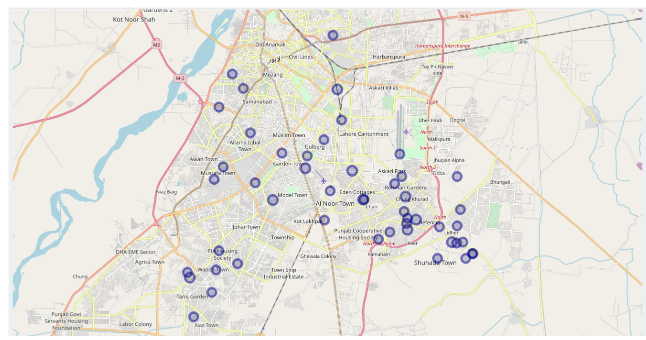

Figure 2: Spatial distribution of residential properties for which the data is collected.

The next step in the process is to find a representative subset of the electricity consumers. This can only be achieved if the subset is diverse.

For this task, a large number of households were required that were willing to share their electricity consumption profiles. For a household to be selected, several considerations were made. We selected households which had at least two persons living in it to get a better demand profile for multiple residents. It was also ensured that the property must be owned by its occupants so that there is a high chance of the property being occupied during the data collection period. Another condition was that the house should have only one meter provided by the utility. Some households register multiple meters with the utility company for each floor of the house which creates complications while installing our smart meters. Utility companies in Pakistan provide a three-phase or a single-phase connection to its customers. We have tried to include both types of these connections in our data collection exercise.

After an initial selection procedure, a detailed form was filled by a resident of all selected households, providing various details about the household which are listed in appendix A below. The detailed form was designed to cover all the significant loads that can greatly impact our understanding about their usage. Air conditioners were particularly focused on because of their large share in the electricity consumption in Pakistan. Other than this, demographics of the households were also recorded such as the total number of adults, children and senior citizens.

The selection of smart meter to be used was decided after a trial of several brands. In the end, eGauge[8] was finalized to perform the task of data collection. This energy meter has three voltage sensors and 12 current channels. One of the advantages of these meters is their built-in solid-state memory that can retain data of up to a year at 1-minute granularity. So even if there is no internet connection, data is saved in the device. The device also has a built-in web server and is fully configurable over the web.

The smart meter installation phase started in February 2018 and ended in May 2018. The selected households were visited by our installation team consisting of three people, including at least one technician. The smart meters were installed in the distribution panel of the household, with CT connections to the main circuit breaker and selected sub-circuit breakers, which connect high powered device or individual rooms to the mains. The wiring structure in households was unorganized and had several complicated nodes and meshes. Figures 3 and 4 illustrate a typical household wiring and our installed smart meter. Wiring standards were not followed and no labelling was provided for circuit breaker identification. Several households had UPS connections, which provide few hours of backup for mostly lighting and fans in case of power outages, which are quite frequent in Pakistan. Power lines of UPS connected appliances run through a power inverter/converter.

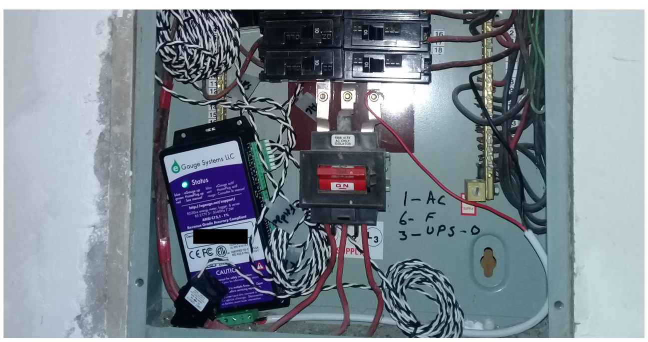

Figure 3: Smart meter installed in one of the selected households

Due to chaotic wiring in the circuit breaker it was difficult to identify the appliances and their corresponding sub-circuit breakers in the distribution panel. To overcome this issue we manually toggled each sub-circuit breaker which allowed us to match each appliance with its respective sub-circuit breaker. This allowed us to create a circuit diagram for the whole household. After the detection process, CTs were installed at desired circuit breaker outputs. To pass the complexities of a UPS, CTs were only installed on the output of the sub-circuit breaker that feeds power to UPS when it is charging. This has only been done for few households. As the smart meter is powered from the main supply provided by the utility, it is powered off in case of a power outage. However, the household might be using some power from the UPS. This usage of back-up energy is not directly recorded by our installed smart meter. However, the smart meter does record the charging of back-up batteries once the power supply from the utility is restored.

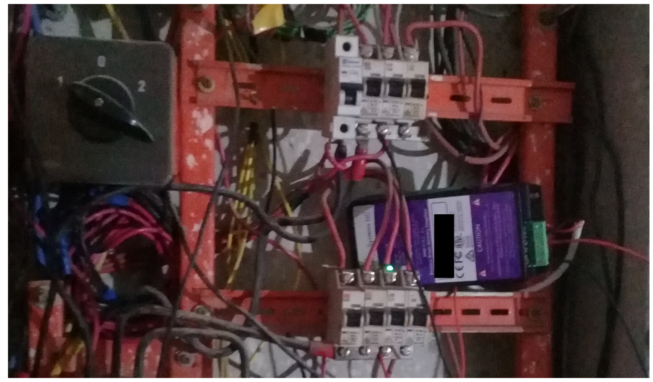

Figure 4: Smart meter installed in one of the selected households. Note the disorganization in wiring

Once the smart meter is installed, it is important to connect it to the internet since regular visits for check-up and data retrieval from the smart meter are inconvenient and time-consuming for the data retrieval team. Three ways have been adopted to connect the smart meter to the internet. One way is by directly connecting the smart meter to the home internet router using LAN cable. This is the most direct way, as it does not require any additional equipment.

The second way of internet connectivity is by using a home-plug [5] that uses power line communication (PLC). A home-plug is connected to the households internet router and plugged into the single phase wall socket. Now the power lines for that particular phase can also be a medium for communication to another such home-plug plugged into the same phase. Our smart meters also have a built-in Home-plug for this purpose. However, the PLC system that we use has a limited range. The signal can only travel 100 ft in power lines before it is attenuated to the extent that it is not detected by the other Home-plug.

The third way is by using 4G enabled Wi-Fi dongles sold by the telecommunication companies. This requires another device called the Wi-Fi-LAN converter. As the name suggests, it transforms Wi-Fi signal from 4G devices to LAN, which is fed to the smart meter's LAN port. This method is only adopted in households where either there is no internet connection provided by the household or the internet router is more than 100 ft away from the distribution box.

Figure 5 shows the data flow diagram, The voltage and current sensors (Clamp-On current transformers) send voltage and current readings respectively, from the distribution box to the smart meter, which are processed by the smart meter. The smart meter calculates power consumed using the formulas configured while installing the smart meter. If the smart meter is online, the data is available for download through API provided by the smart meter[7]. A Python script regularly calls the API to get the required data. The data is stored in the raw form and then processed. In the processing phase, the data is row bind in a single file for every month, as the python script outputs a CSV file for each day. Both monthly and daily data is stored in the system and a back up is also created for safe keeping.

The monthly files will be publically available for download for each household.

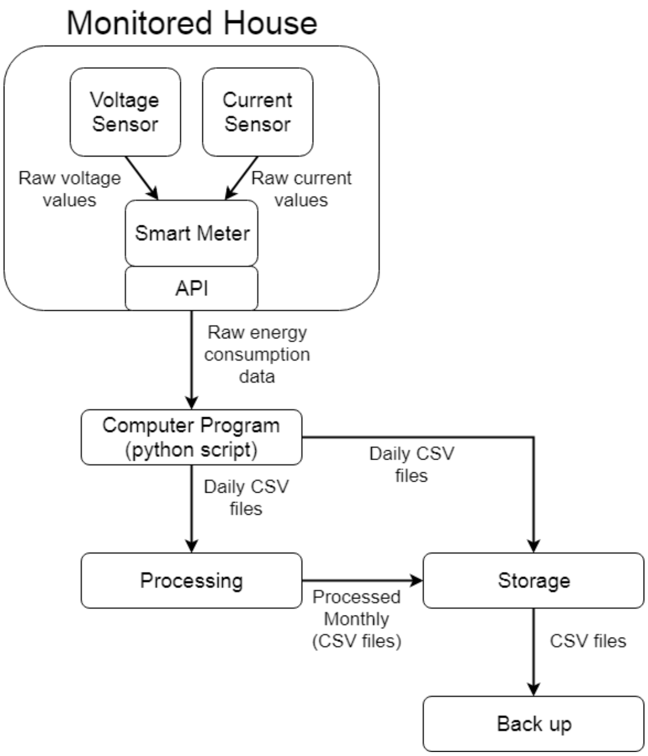

Figure 5: Data flow diagram for data collection architecture

4 DATA DESCRIPTION

This dataset presents electricity consumption data of 42 households, recorded at a minute interval. At the time of submission of this paper, eight months of data has been recorded, starting from June 1, 2018. For each day, data is stored in an individual CSV file for each household. Each file contains 1440 rows corresponding to each minute of the day and a varying number of columns. The number of columns varies because for each different household a different number of appliances are selected for monitoring. However, the first two columns are always for the timestamp and the total usage of the household in kW. These columns are named Date_Time and Usage_kW respectively. The rest of the columns are abbreviated, for example, the column BR_kW shows the energy consumption in the bedroom of that particular household. As said before, for monitoring the appliances air conditioners are preferred, so a column named ACBR_kW shows the energy consumption of an air conditioner installed in the bedroom of that particular household. As each household has a varying number of appliances being monitored, the directory of each household contains a description of each column in the electricity consumption files.

Other than electricity consumption data, several attributes related to the households are also recorded. These attributes are summarized in the table 1. The demographics of the household include the total number of people in the households, and it is further divided into the number of children (age < 14), adults (14 <= age < 60) and senior citizens (age > 60). We have also recorded the number of permanent and temporary residents because in some households occupants leave for several days for jobs, studies and other similar activities and are only at home for some part of the year. The table 2 in the appendix shows the summary of some of the demographical attributes of the houses where our smart meters are installed. Build year refers to the year the house was built. It is believed that the architecture of a house affects the energy consumption of its occupants and the architecture can be inferred from the build year of the house [11]. Property area is another attribute that is considered to have a positive correlation with the energy consumption in that household. Other recorded attributes include number of floors and number of rooms. The meta-data file also contains information about the ceiling height and the number of each type of rooms in a house, i.e. bedrooms, kitchen, living rooms and drawing rooms. It was also recorded whether the house has any heat insulation installed. One thing of particular interest is that the average occupancy of hoouseholds in Pakistan is 6.45 according to the lattest census in 2017 [21] which is quite close to our sample's average household occupancy of 6.19. In table 3, summary of some of the electrical loads of the households is summarised. Every houshold in our dataset has an air conditioner installed, which shows the impact of temperature on our demand load. We are expecting to see a positive correlation between temperature and electricity cosnumption of any household.

Connection type describes whether the electrical connection provided by the utility to the mains is single-phase, double-phase or three-phase. Most households have either single or three-phase connections. Other than this, all the significant electrical loads of the household are also recorded which include the number of LED lights, tube-lights, fans, refrigerators, washing machines, water dispensers, water pumps, electric iron, electronic devices and electric heaters. The meta-data also has information about the UPSs installed at each household. We recorded the wattage of the inverter/converter in watts and the capacity of lead-acid batteries in Ah.

5 CONCLUSION

To our knowledge, PRECON dataset is the first attempt to collect extensive residential energy consumption information from South Asia in particular Pakistan. Buildings consume more than 50% [20]of electricity in Pakistan. Therefore, any intervention for anything related to energy utilization in buildings requires insights of energy consumption by building occupants. Moreover, distributed solar and other captive methods of energy generation require an in-depth understanding of consumer behaviour for better sizing and optimization of captive units. Since Lahore is located just a few kilometres from India, this dataset can also be used for assessing consumer behaviour with respect to climatic and other natural events for Northern India as well.

REFERENCES

[1] CA Balaras, K Droutsa, AA Argiriou, and DN Asimakopoulos. 2000. Potential for energy conservation in apartment buildings. *Energy and buildings* 31, 2 (2000), 143–154.

[2] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. 2012. Smart*: An open data set and tools for enabling research in sustainable homes. *SustKDD, August* 111, 112 (2012), 108.

[3] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B Srivastava. 2013. It's Different: Insights into home energy consumption in India. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 1–8.
[4] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. 2014. The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms. In *Proceedings of the 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys 2014). Memphis, TN, USA*. ACM, 80–89.
[5] TP-Link Technologies Co. 2019. Powerline Adapters. https://www.tp-link.com/uk/products/list-18.html
[6] David Connolly, Henrik Lund, Brian Vad Mathiesen, and Martin Leahy. 2011. The first step towards a 100% renewable energy-system for Ireland. *Applied Energy* 88, 2 (2011), 502–507.
[7] eGauge Systems LLC. 2019. eGauge XML API. https://www.egauge.net/media/support/docs/egauge-xml-api.pdf
[8] eGauge Systems LLC. 2019. Monitor Electricity on Every Circuit with Precision and Accuracy. https://www.egauge.net/
[9] Adrian Filip. 2011. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*. 2012.
[10] Jack Kelly and William Knottenbelt. 2014. UK-DALE: A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand. *ArXiv e-prints* 59 (2014).
[11] Rihab Khalid and Minna Sunikka-Blank. 2018. Evolving houses, demanding practices: A case of rising electricity consumption of the middle class in Pakistan. *Building and Environment* 143 (2018), 293–305.
[12] Henrik Lund and Brian Vad Mathiesen. 2009. Energy system analysis of 100% renewable energy systemsâĂŤThe case of Denmark in years 2030 and 2050. *Energy* 34, 5 (2009), 524–531.
[13] Stephen Makonin, Fred Popowich, Lyn Bartram, Bob Gill, and Ivan V Bajic. 2013. AMPds: A public dataset for load disaggregation and eco-feedback research. In *Electrical Power & Energy Conference (EPEC), 2013 IEEE*. IEEE, 1–6.
[14] Stephen Makonin, Z Jane Wang, and Chris Tumpach. 2018. RAE: The rainforest automation energy dataset for smart grid meter data analysis. *Data* 3, 1 (2018), 8.
[15] Ian George Mason, SC Page, and AG Williamson. 2010. A 100% renewable electricity generation system for New Zealand utilising hydro, wind, geothermal and biomass resources. *Energy Policy* 38, 8 (2010), 3973–3984.
[16] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D'Alessandro, and Andrea M Tonello. 2014. GREEND: An energy consumption dataset of households in Italy and Austria. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*. IEEE, 511–516.
[17] David Murray, Jing Liao, Lina Stankovic, Vladimir Stankovic, Richard Hauxwell-Baldwin, Charlie Wilson, Michael Coleman, Tom Kane, and Steven Firth. 2015. A data management platform for personalised real-time energy feedback. In *Procedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting*.
[18] S Noor and MW Siddiqi. 2010. Energy consumption and economic growth in South Asian countries: a co-integrated panel analysis. *International Journal of Human and Social Sciences* 5, 14 (2010), 921–926.
[19] Oliver Parson, Grant Fisher, April Hersey, Nipun Batra, Jack Kelly, Amarjeet Singh, William Knottenbelt, and Alex Rogers. 2015. Dataport and nilmtk: A building data set designed for non-intrusive load monitoring. (2015).
[20] M Sohail and MUD Qureshi. 2011. Energy-efficient buildings in pakistan. *Science Vision* 16 (2011), 27–38.
[21] Tribune. 2017. Census 2017: Family size shrinks. https://tribune.com.pk/story/1491353/census-2017-family-size-shrinks/
[22] Akshay SN Uttama Nambi, Antonio Reyes Lua, and Venkatesha R Prasad. 2015. Loced: Location-aware energy disaggregation framework. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 45–54.

Appendices

A META-DATA: ATTRIBUTES STORED FOR EACH HOUSEHOLD

Table 1: Attribute of the household & their description

Attribute	Description
Ownership of the property	Is the property owned by the household?
Property Area	Area of the property in sq. ft.
Floors	Number of floors of the household.
Build Year	Year in which construction of the property was completed.
Electrical Connection Type	Is the connection type single or three-phase.
Ceiling Height	Height of the ceiling in feet.
Ceiling Insulation	Wheter heat insulation is used or not?
Rooms	Total number of rooms, such as bedrooms, kitchen, living room and drawing room, and their description.
Residents	Number of residents and their distribution i.e. children (age<14), adults (14 < age < 60), senior citizens (age > 60) and temporary & permanent residents.
Air Conditioners	Number of air conditioners, thier brand, installation year and tonnage.
Other appliances	Number of other appliances in the household including refrigerator, washing machines, iron, water dispenser, lighting loads, fans and electronic devices.

[1]Electronic Devices include phones, tablets and laptops and other such devices

Table 2: Summary of Demographics of the Households

	# of People	# of Children	# of Adults	# of Senior Citizens	Property Area [Sq. ft]	Build Year
Min	3	0	2	0	681	1976
First Quartile	5	1	3	0	2450	1998
Median	6	1.5	4	1	2723	2005
Mean	6.19	1.71	4.1	0.88	5215	2003
Third Quartile	7	3	4.7	2	5445	2012
Max	11	5	8	2	32670	2015

Table 3: Summary of Electrical load of the Households.

	# of Air Conditioners	# of Washing Machines	# of Water Pumps	# of Electronic devices [1]
Min	1	0	0	0
First Quartile	5	1	0	5
Median	6.5	1	1	8
Mean	7.38	1	0.7	11.3
Third Quartile	10	1	1	11.5
Max	14	2	3	15