A critical study of power consumption patterns in Indian Apartments

Nagasuri Venkata Apurupa

IIIT-D-MTech-CS-GEN-18-MT16036

July, 2018

Indraprastha Institute of Information Technology
New Delhi

Thesis Committee
Dr. Pushpendra Singh (Advisor)
Dr. Sambudhho Chakravarthy (Internal Examiner)
Dr. Arun Balaji Buduru (Internal Examiner)

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science

Certificate

This is to certify that the thesis titled **"A critical study of power consumption patterns in Indian Apartments"** submitted by **Nagasuri Venkata Apurupa** for the partial fulfillment of the requirements for the degree of *Master of Technology* in *Computer Science & Engineering* is a record of the bonafide work carried out by her under our guidance and supervision at Indraprastha Institute of Information Technology, Delhi.In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.The results contained in this thesis have not been submitted in part or full to any other university or Institute for the award of any degree/diploma.

Dr. Pushpendra Singh

Department of Computer Science

Indraprastha Institute of Information Technology, New Delhi

Abstract

Energy conservation plays an important role in the economic development of a country. The total share of building 's energy consumption in the country 's energy consumption is increasing every year in India. Thus, it is essential to study the various features that impact the energy consumption in the domestic house holds which can be used for effective policy making for energy conservation and also advise individuals on how to monitor and control their energy consumption. The purpose of our study is to explore the various factors that are affecting the total power consumption in domestic households in India. The factors include the floor level of the apartment, family size, orientation of the building, no of adults and children. The study was conducted on the faculty apartments in the IIIT-Delhi campus, India from the year 2014. The significant features were identified using regression analysis on the data. Interaction effects between the features that impact the energy consumption was also studied. So from our analysis, we have found out that floor level, orientation, family size, the presence of a working couple are having a significant impact on the energy consumption patterns of the household. We have also developed an android app that will help household dwellers to monitor their usage in real time and also to check their historical usage and compare the trends. They can get an estimate on the number of units consumed and average power consumed for the selected duration of time.

Acknowledgments

I would like to express my sincere gratitude to my advisor Dr. Pushpendra Singh for providing me an opportunity to work with him. I would be grateful for his constant guidance and encouragement.

I would also like to thank my thesis defense committee members, Dr. Sambudhho Chakravarthy and Dr. Arun Balaji Buduru for their valuable feedback during the thesis defense.

I would like to thank Haroon Rashid (PhD Scholar at IIIT-Delhi) for mentoring me during the thesis, providing me constant feedback during my work and report writing.

I also would like to thank Garvita Bajaj(PhD Scholar at IIIT-Delhi) for helping me deploy my app in the server and providing the necessary devices during the development and testing and timely support.

I also thank IIIT-Delhi faculty members and our Director Dr. Pankaj Jalote for allowing us to perform experiments on the faculty residence apartments.

Lastly, I am very much indebted to my parents for their emotional and moral support throughout my masters without which my journey would not have been smooth.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Energy conservation plays a vital role in the economic development of a country. The energy consumed by either domestic or commercial buildings contribute to a greater extent in the country 's total energy production. In India, energy consumption in Indian households has tripled since 2000 [1]. Building energy consumption accounted for a share of 24% in total energy consumption in India in 2016 (See Figure 1.1) and is predicted to increase by an average of 2.7% every year by 2040. According to U.S Energy Information Administration(EIA), International Energy Outlook 2017, India 's building energy consumption would be fastest growing among other countries by 2040.

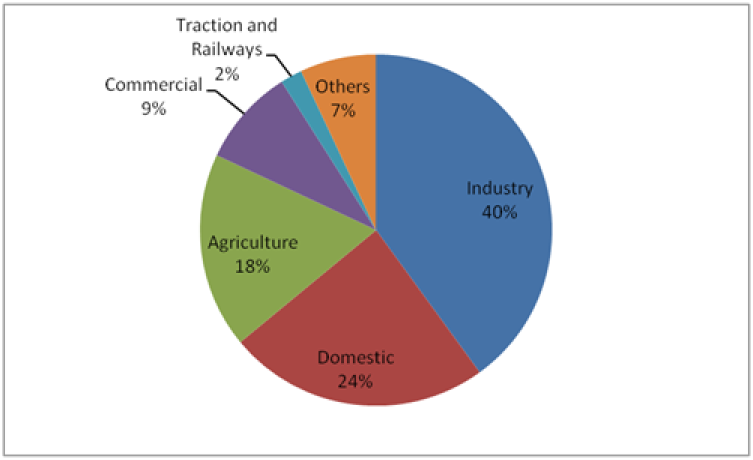

Figure 1.1: Sector Wise electricity Consumption in India in 2016-17

Some of the factors affecting the increase in building energy consumption are development in technology resulting in new appliances at lower prices, increase in household incomes, quick electrification measures in India [1]. With the growth in technology, we can build appliances which are not only energy efficient(green appliances or energy aware appliances) but also affordable, thus increasing their usage and subsequently increases the energy consumption. So, it is the hour of need to optimize the consumption of electricity in buildings. This can be achieved by understanding the various factors that affect energy consumption in domestic households.

1.2 Aim

Buildings encompass a different set of activities which affect the energy consumption differently. Building's energy consumption comes from various factors like TV's , refrigerators and different heating and cooling appliances. Building design, dwelling factors, and other socio-economic factors also have a significant effect on the energy consumption. Our work mainly focuses on the study of how socio-economic factors like family size and dwelling factors like floor level and orientation of the building affect the energy consumption in a domestic building.

Orientation of the building is also one of the factors that decides the amount of sunlight that is entering the building thus influencing the energy consumption accordingly. The amount of sunlight entering the building also depends upon the temperature of the region where the building is located. Building orientation is the way the building and it 's windows are positioned. A good orientation will minimize the energy consumption and also provides natural ventilation.

1.3 Deliverables

In our paper, we study the impact of different socio-economic and dwelling factors on the energy consumption of a typical residential building. We have also developed a mobile application which enables the household dwellers to track their energy consumption every day so that they can take measures to control their energy consumption at their fingertips.

1.4 Thesis Outline

Chapter 2 discusses related work and literature survey. Chapter 3 discusses about data collection and data pre-processing. Chapter 4 discusses the exploratory data analysis which includes Peak analysis of data, Graphical analysis of our data and clustering analysis.Chapter 5 explains the Regression Analysis and results. Chapter 6 explains the development of EMonitor app and it 's features and Chapter 7 discusses Conclusions and Future Work.

Chapter 2

Related Work

This chapter details the research work done on understanding how the socio-economic and dwelling factors affect the building energy consumption. Socio-economic features refers to characteristics of occupants in the house such as number of occupants in the house, presence of children , annual household income etc,. while dwelling features are the characteristics of the building like building type, building age, no of floors and rooms.

2.1 Socio-economic features

Huebner et al. [2] explains by how much percentage does the different features like building factors, socio-demographics and behaviors affect the total energy consumption and also summarizes the effect of each feature alone on the total consumption. It also addresses the multi-collinearity problem between the predictors. Their findings showed that the most significant predictor of energy consumption was the dwelling's size and type in building variables, household size in socio-demographic predictors and length of the heating season in heating behaviors.

The effects of socio-economic and dwelling factors including the presence of teenagers and electric space heating were studied by Lomas et al. [3] . Odd ratio analysis was done to study the effect of factors on total energy consumption in UK domestic buildings. Results showed that the presence of teenagers, having electric space heating as the primary form of heating and portable electric heating was having a high impact on electricity consumption while the employment status and education level of the Household Representative Person, the number of floors and fixed electric heating has no impact on high electricity consumption.

Another study on the role of occupants and socio-economic features by Steemers et al. [4] using regression analysis show that besides climate, socioeconomic factors and occupant behaviors are significant.

Mutual information criteria and random forests model were used in [5] to find out the importance of socio-economic features and the results revealed that the number of occupants, the age of occupants, no of children and adults were among the top 10 predictors in energy forecasting.

According to studies made in [6], [7] , household size is also shown to have to positive effect on the total energy usage.Labanderia et al. [8] found that place of residence, household size, age, education or labour force participation are the significant explanatory variables on the energy consumption based on the study conducted on the residential households in Spain.

2.2 Dwelling features

The effect of building orientation on the energy consumption was studied in [9]. The authors have conducted simulations on office buildings in different temperature zones and studied the orientation of buildings and the effect of glazing on windows on optimum energy consumption. A study made on Portugal residential buildings show that both household and dwelling characteristics have a significant impact on electricity consumption [10]. Another study [11] on 1628 households show that floor area, location and weather are the most import factors contributing to the electricity consumption. Brian

Norton et al. [12] studied the patterns in residential dwellings and found a correlation between the annual electricity consumption and floor area. [13] studied the role of geometry parameters like building shape, window to wall ratio and orientation on the energy consumption in the school buildings in China and found out that by variation of geometric parameters, they got a maximum of 13.6% of energy savings.

Chapter 3

Data Set

3.1 Data Collection

We have collected time series dataset of 27 apartments from the faculty residence apartment building of Indraprastha Institute of Information Technology Delhi(Figure 3.1). Each apartment is installed with a smart meter for measuring the usage, and the data is stored in a cloud server. The faculty residence consists of three series in each floor. Apartments in series one are oriented towards south-east direction while Apartments in series two towards the south and series three towards the south-west. We have worked on data in floors from 2 to 9 from the year 2014 to till May 2018. Also, we also collected family size, number of adults and children.

Figure 3.1: IIIT-Delhi Faculty Building

3.2 Data Pre-processing

The original data set is a time series of power consumption values(in Watts) with a sampling rate of 30 seconds. As the original data set is computationally intensive to handle with that sampling rate, we have downsampled the data set.

We have experimented on data at different sampling rates like hour wise, day wise, week wise and month wise. For resampling hour wise, we have used mean resampling technique where the data points are replaced by the mean of the data points within the given hour. We have also conducted experiments by excluding weekends and daytimes that gave us better insights.

Chapter 4

Data Analysis: An exploratory approach

In this chapter, we have studied different methods for finding peak consumption in a typical household and developed a system that summarizes the duration of peak consumption patterns of the apartments. Also, we have explored if there are any patterns among the consumption styles of the dwellers using graphical analysis and clustering.

4.1 Peak Analysis

Peak identification in a time-series data is useful in many applications.A peak indicates a sudden increase in the power or the energy consumption.Peak detection will help us identify patterns in the energy consumption and will be useful in detecting high consumption rates.

We have studied the data to detect and summarize the peak power consumption behaviors of the consumer. We have compared different approaches of peak detection on our data and developed a system that would summarize the time and duration of peak consumption patterns during a week. For peak detection, we have compared the following algorithms.

4.1.1 Z-score approach

This approach uses the concept of moving mean. A moving mean is used to analyze the data points by calculating the series of means of different subsets of the original data set. A moving mean is mainly used to smooth the time series data.

In this approach, if a given data point is 'x' number of standard deviations away from the moving mean, then the algorithm considers the data point as a peak. The algorithm takes lag of the moving window and threshold as inputs from the user. The lag of the window is the size of the window and threshold is the value of 'x' mentioned earlier. Here the standard deviation is the standard deviation of the moving window.

When experimented with various values of lag and threshold, the algorithm gave more number of false positives. We are finding whether a point is a peak by comparing if the point is greater than the standard deviation of the moving mean. As the moving mean changes for every point, it calculates peaks locally thus giving rise to more number of false positives.

4.1.2 Smoothing Approach

Vlachos et al [15] describes an algorithm where the time series data is first smoothed using a moving average filter and the peaks are identified. A data point is considered as a peak if it is larger than x times the standard deviation of the smoothed time series data set.

When experimented on our data set, compared to Z-score algorithm we observed less number of false positives. With this approach, there were number of false negatives

4.1.3 Adaptive thresholding

Beukelman et al. [16] proposed a technique to compute the threshold automatically to adapt to the noise-levels in time series data. In this approach after filtering the data using the moving average, we calculate the threshold using the following formula

$$threshold = (max + abs\_avg)/2 + K * abs\_dev$$

Here max is the maximum value in the smoothed time-series data, abs_avg is the average of the absolute values in the time-series, abs_dev is the mean absolute deviation, and K is a user-specified constant. Thus a data point is considered as a peak if its value is greater than the threshold.

4.1.4 S1 and S5 approaches

These approaches are proposed by Palshinkar G [17] which computes a score for each data point which denotes the *spikiness* of the data point. In S1 approach, the score for a point is calculated as the average of 1. maximum of signed distances between the point and its k left neighbors 2. maximum of signed distances between the point and its k right neighbors. A data point is considered as a peak if its score value(s) satisfies the following equation

$$s - mean >= h * std\_dev$$

Here mean and std_dev are mean, and standard deviation of the time-series data respectively and h is a user-defined constant.When experimented with our data, we have used a window size of 4(K value) and the value of h to be one which gave us good results.

In S5 scoring approach, for each data point, if the point is greater than the mean of the moving window and difference of point and mean of the window is greater than threshold times the standard deviation, it is considered as a peak point.

Out of the above two approaches, we got better performance in S1 compared to that of S5.

4.1.5 Sorting Approach

Another approach which we tried was to sort the time series data first and consider those data which is greater than the standard deviation of the data by the mean value of the data. We can stop traversing the data if the required number of peaks are found.

The performance of the above algorithms when tested on our time-series data in the months of January, July and October was shown in Figure 4.1 and Figure 4.2

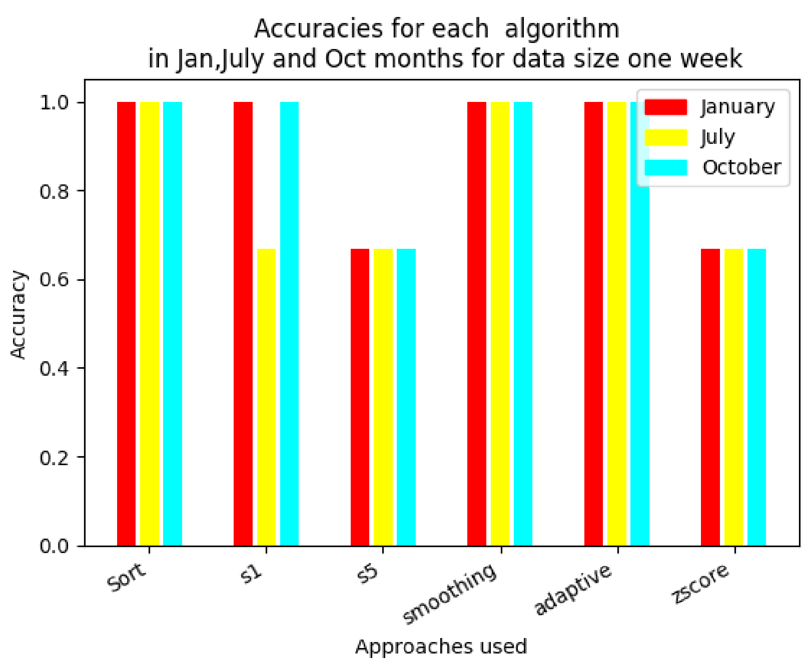

Figure 4.1: Accuracy values for a typical week data

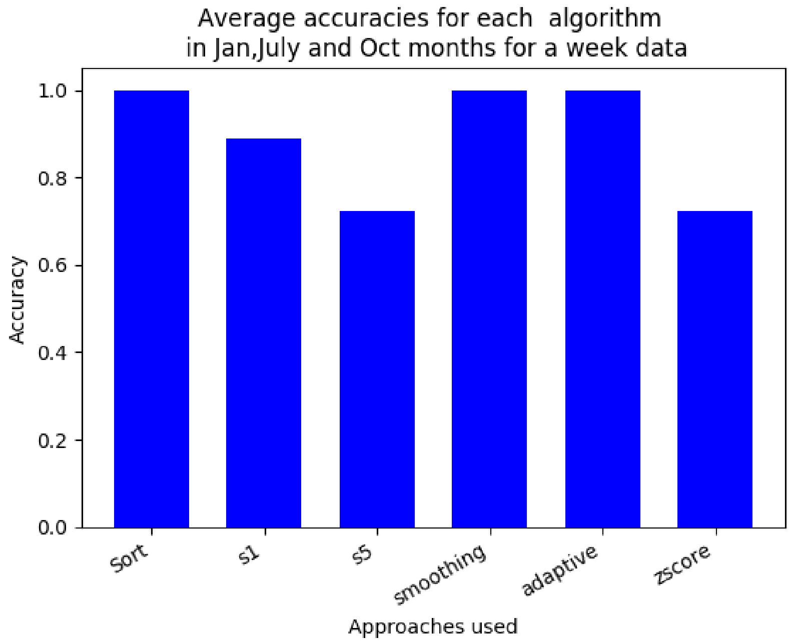

Figure 4.2: Average accuracy of each algorithm

We can observe that from Figures 4.1 and 4.2, Smoothing, Sorting and Adaptive thresholding approaches performed well with 100% accuracy while the least performing approaches were Z-Score and S5 with 66% accuracy.

When studied with data set having two weeks data in the months of January, July and October the performance of the approaches are shown in Figure 4.3 and Figure 4.4

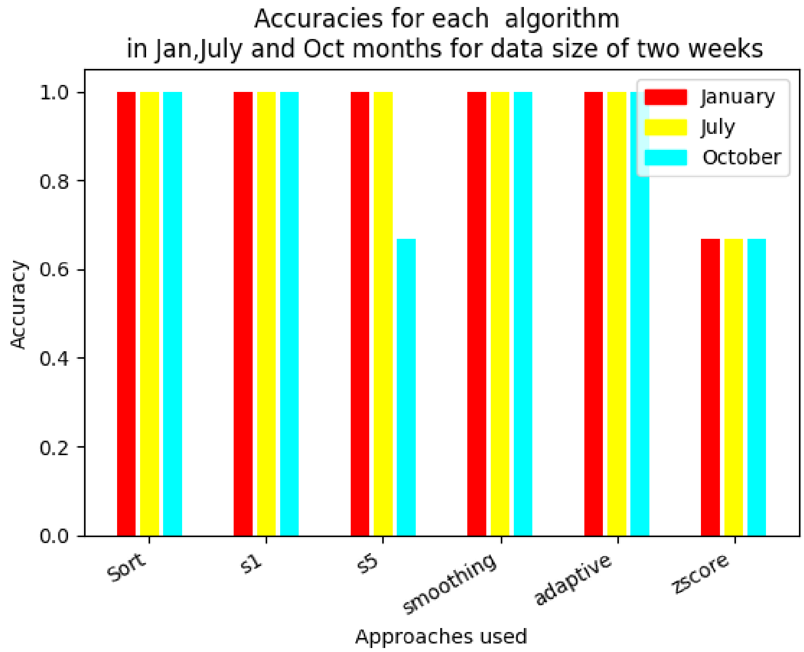

Figure 4.3: Accuracy values for two weeks data

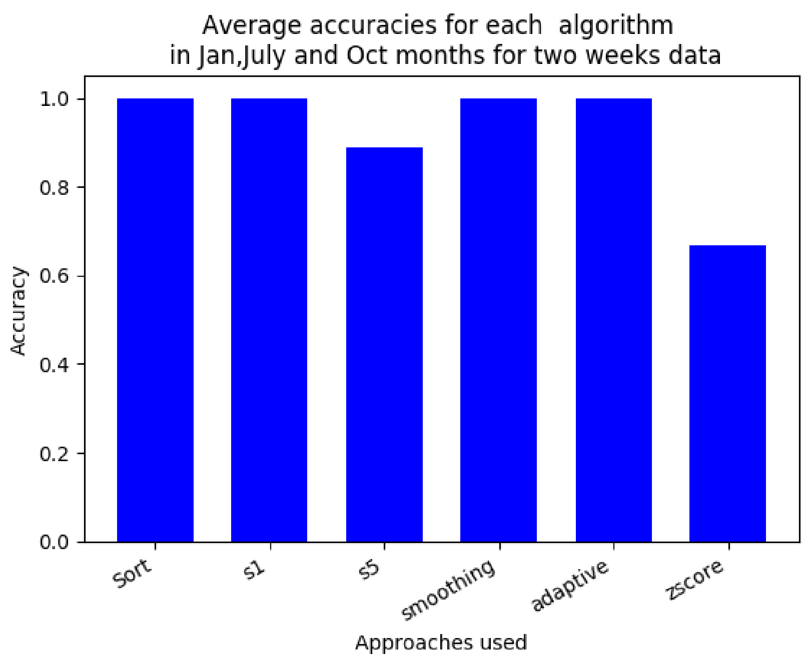

Figure 4.4: Average accuracy of each algorithm for two weeks data

We can observe that from Figures 4.3 and 4.4, Smoothing, S1 scoring, Sorting and Adaptive thresholding approaches performed well with 100%

accuracy while the least performing approaches were Z-Score and S5.

When the same experiments were repeated with a data set having power consumption values of four weeks, the results are shown in Figure 4.5 and Figure 4.6

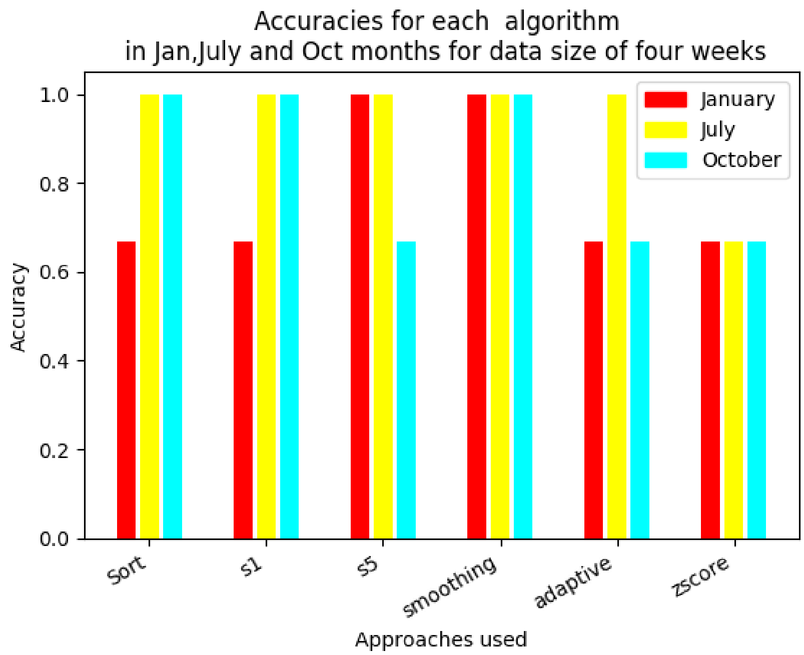

Figure 4.5: Accuracy values for four weeks data

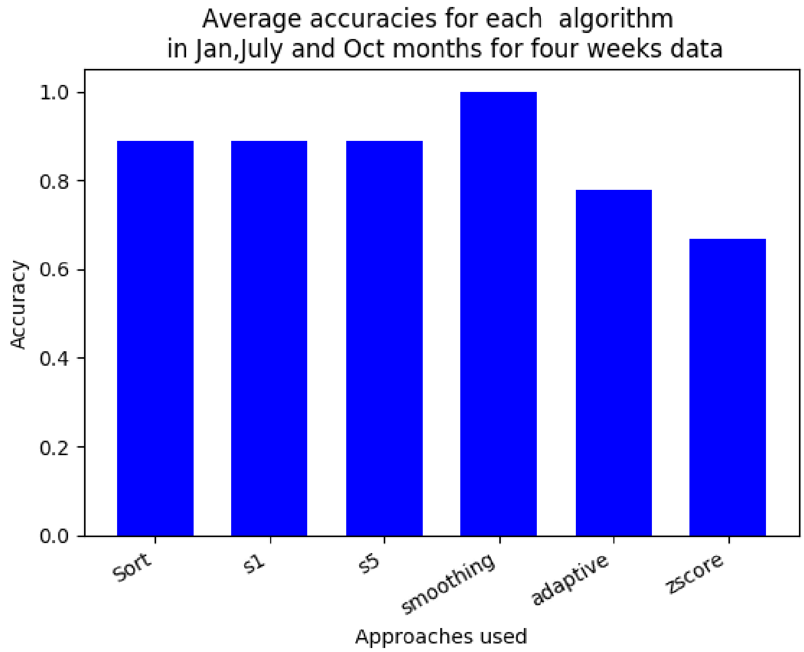

Figure 4.6: Average accuracy of each algorithm for four weeks data

We can observe that from Figure 4.5 and Figure 4.6, while in the month of January we get S5 and Smoothing approaches to give 100% accuracy, in July we see that algorithms gave same performance except for Z-score and in October, we see that Sorting, S1, and Smoothing gave the highest accuracy. So, on an average, the best performing approach was the smoothing approach.

If the above experiment is repeated with a data set with weekends removed, the performance of each of the algorithm are shown in the Figure 4.7 and Figure 4.8

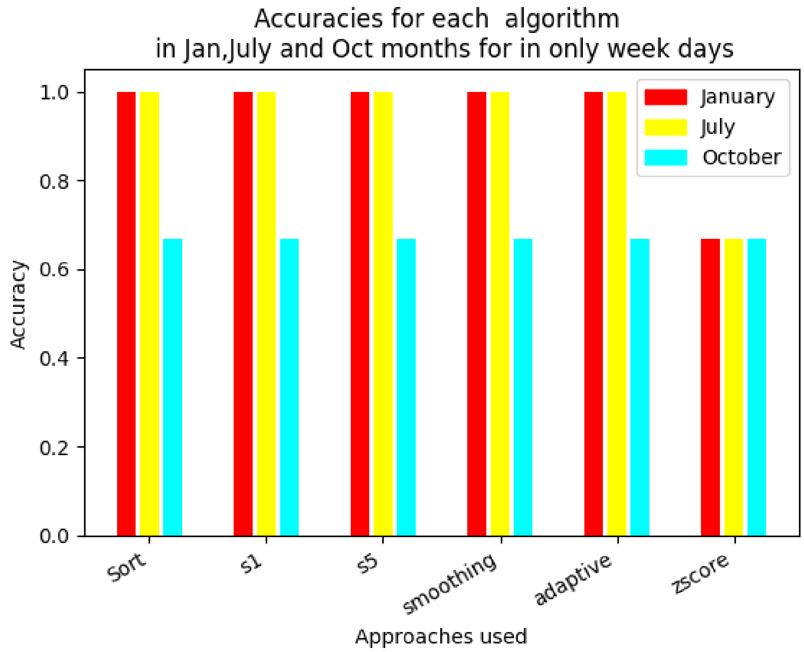

Figure 4.7: Accuracy values for weekdays data

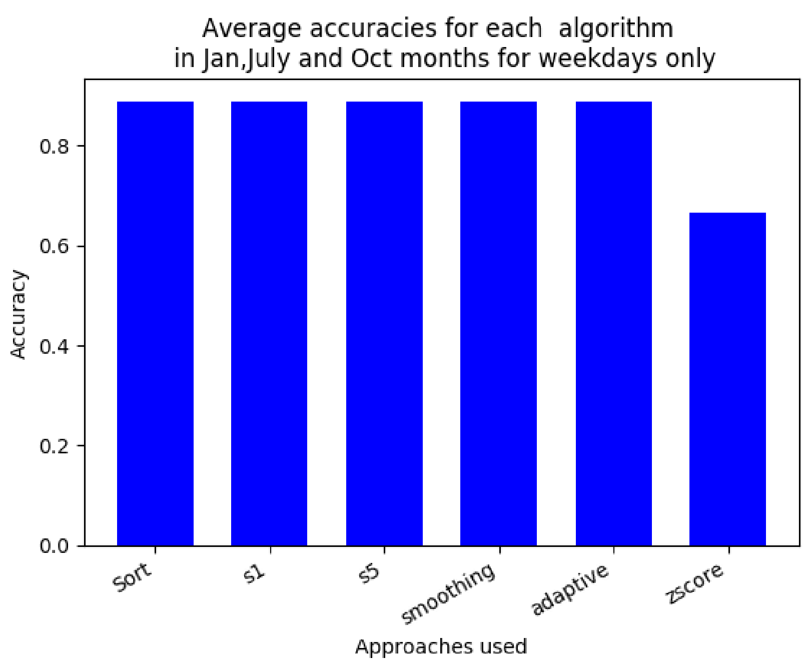

Figure 4.8: Average accuracy for weekdays data

We can observe that sorting and smoothing approaches are similar except for smoothing the time series data. Also, the Z-score algorithm gave the least performance because of moving standard deviation. S5 sometimes works similar to that of S1 when constant values were tweaked. For weekdays only data, smoothing, sorting, s1 and s5 gave similar performance. For weekends only data, sorting and s1 approaches worked better than smoothing approaches.

Once we found out the best performing algorithm, we have used it to find peaks and duration during a week by dividing a day into different windows and determining which window the maximum number of peaks fall into. A sample output for the summarizing the peaks is shown in Figure 4.9.Here we can see that a day is divided into different windows and patterns are found.

```
Using sort
Peaks occur between  20 hours and 24 hours with avg peak value 1658.52099626
[8, 0, 0, 4, 4, 12]
 Accuracy  0.666666666667

Using s1
Peaks occur between  20 hours and 24 hours with avg peak value 1595.94999427
[7, 0, 0, 4, 5, 12]
 Accuracy  0.666666666667

Using s5
Peaks occur between  20 hours and 24 hours with avg peak value 1670.24959237
[5, 1, 0, 5, 4, 13]
 Accuracy  0.666666666667

Using smoothing
Peaks occur between  20 hours and 24 hours with avg peak value 1496.76231917
[8, 0, 0, 3, 5, 12]
 Accuracy  0.666666666667

Using adaptive
Peaks occur between  20 hours and 24 hours with avg peak value 1815.88465612
[8, 0, 0, 4, 2, 14]
 Accuracy  0.666666666667

Using zscore
Peaks occur between  4 hours and 8 hours with avg peak value 897.974798638
[27, 28, 28, 28, 28, 28]
 Accuracy  0.666666666667
```

Figure 4.9: A sample output for summarizing peaks of different algorithms

4.2 Graphical Analysis

To visually explore the consumption patterns of the apartments, we have compared the trends apartments-wise, series-wise and months-wise.

4.2.1 Apartment Wise Plots

Here we have taken each apartment and plotted their average power consumption in a day for each month in all different years. Thus, we could get possible insights on how the consumption is varying monthly and what are the months that are getting higher consumption patterns.

Figures 4.10 to 4.12 shows plots in the year 2017 for the apartments in series of the building in the eighth floor.

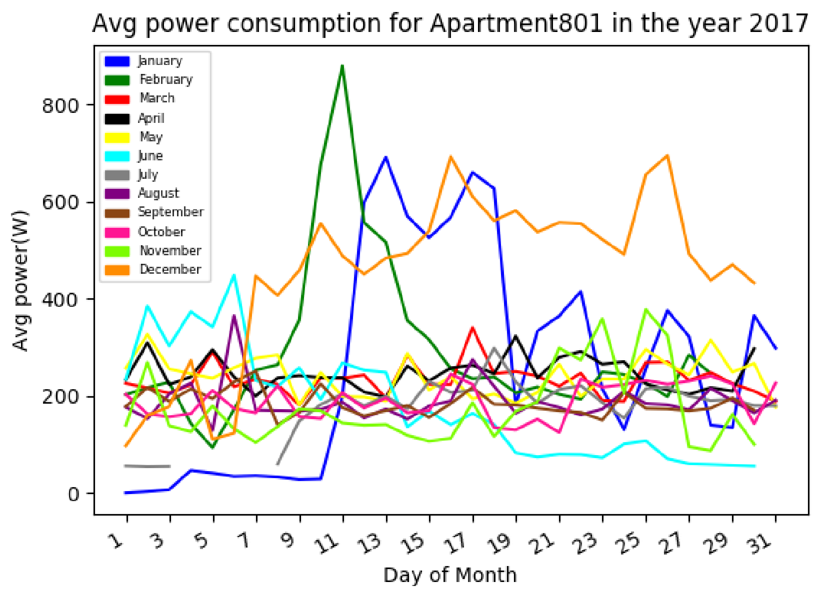

Figure 4.10: Average power consumption for Apartment 801 in 2017

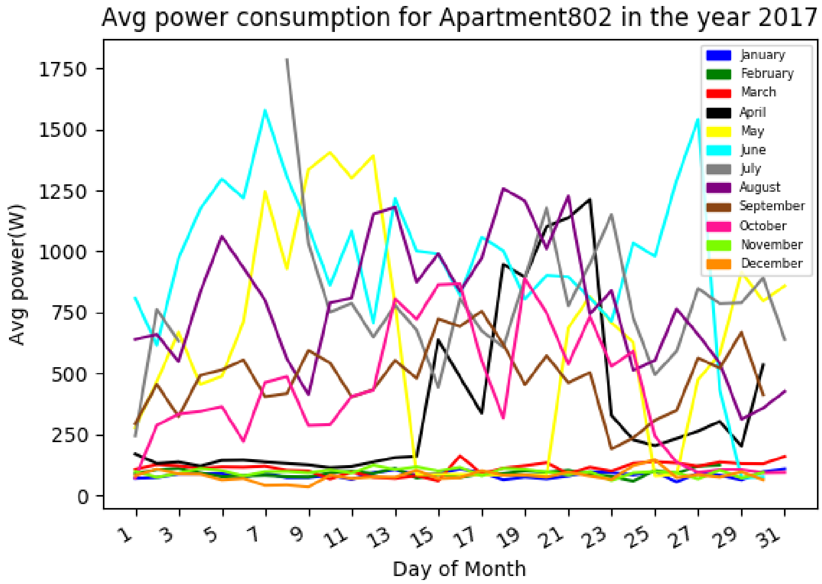

Figure 4.11: Average power consumption for Apartment 802 in 2017

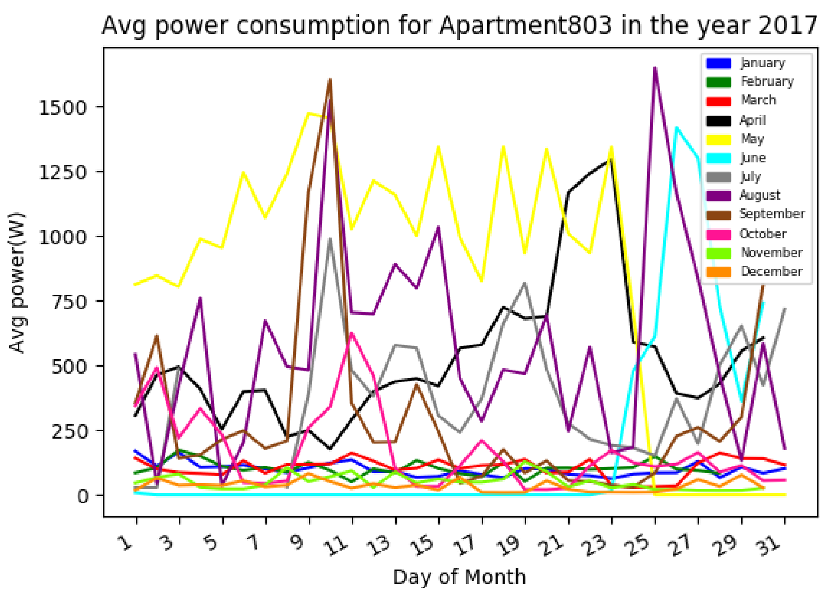

Figure 4.12: Average power consumption for Apartment 803 in 2017

So by manually analyzing the graphs, we got the following insights. Most of the apartments had relatively higher consumption patterns in the months of May, June, July compared to other months. This raise in power consumption can be attributed to peak summers in the region and consequent use of air conditioning equipment. However, there are some apartments which had higher consumption patterns observed in January and December. This can be attributed to relatives joining the tenant or one of the couple taking leave and staying in the apartment for possibly for maternal reasons.

4.2.2 Series Wise Plots

We have tried to analyze the patterns comparing all the apartments in each series(each wing of the building). As there are three wings in the building, we have three different plots comparing all the apartments in that wing for a given month and year.

Figures 4.13 to 4.15 shows the consumption for Apartments in the month of August.

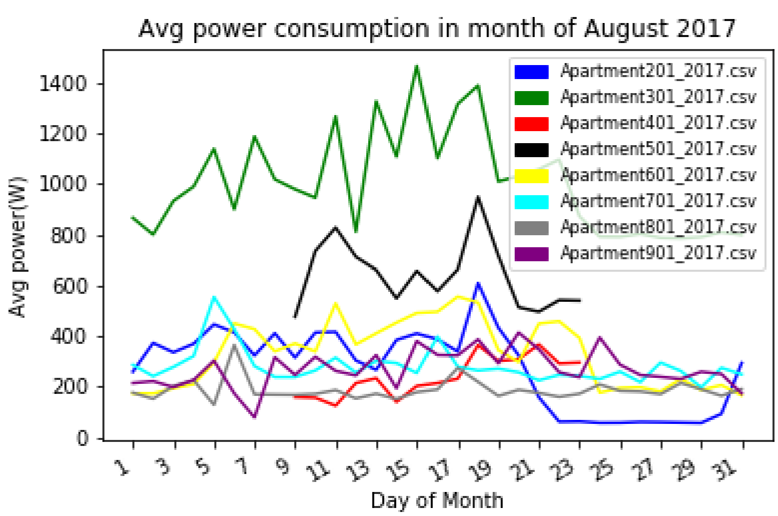

Figure 4.13: Average power consumption for south-east oriented apartments in August 2017

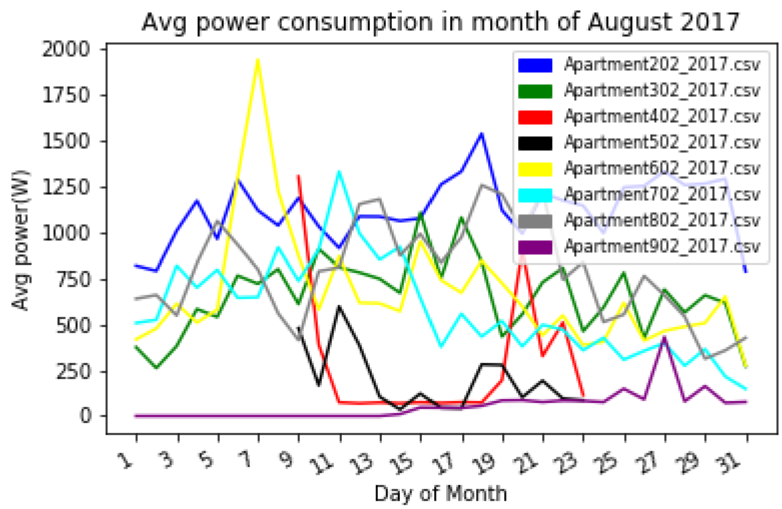

Figure 4.14: Average power consumption for south oriented apartments in August 2017

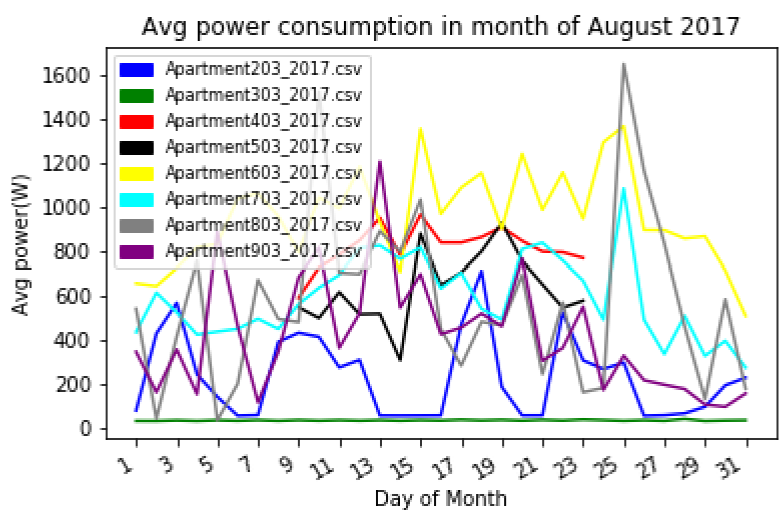

Figure 4.15: Average power consumption for south-west oriented apartments in August 2017

With series wise plots, we could observe the influence of floor level on the power consumption of the apartments oriented in the same direction. By analyzing such graphs, we found out that although most of the households in the higher floors have shown relatively higher consumption patterns, there are some apartments from lower levels which shows higher consumption patterns.

4.3 Clustering

We have clustered data to see how the apartments are grouped with respect to family size, floor level etc. We have used K-Means Clustering algorithm on the data to find the clusters.

4.3.1 k-Means Clustering

k-Means clustering is an unsupervised learning technique to find groups in the given data.The number of groups is given by the K value.It iteratively assigns the data points to the K clusters such that within cluster sum of squares(Variance) is minimum. [18]

4.3.2 K-Means Algorithm

Given a set of observations X and no of clusters K , the algorithm proceeds as follows [19]

- Randomly select K centres.
- **Assignment step:** For each data point in X
    1. Calculate the distance between the data point to each of the cluster centres
    2. Assign the data point to the cluster having the minimum distance.
- **Update step:** Recalculate the new clusters which is the mean of all the data points in the cluster.

The algorithm converges when the assignment does not change from the previous iteration.

4.3.3 Applying k-Means on Apartments data

To find the apartments which fall under same cluster based on their total monthly power consumption , we have used K means on the transformed data which has the total power consumption for each apartment. We have also experimented clustering with weekly total power consumption,total power consumption for 10 days and 15 days

4.4 Observations

Figures 4.16 4.17 and 4.18 shows how apartments are clustered and their trends in different months in the year 2016.

Figure 4.16: Clusters in the month of April 2016

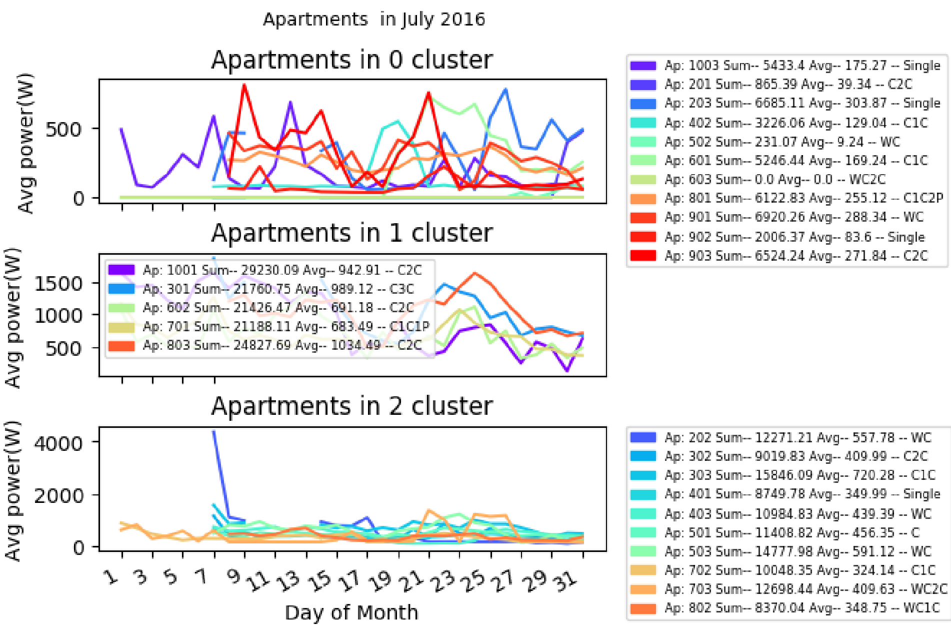

Figure 4.17: Clusters in the month of July 2016

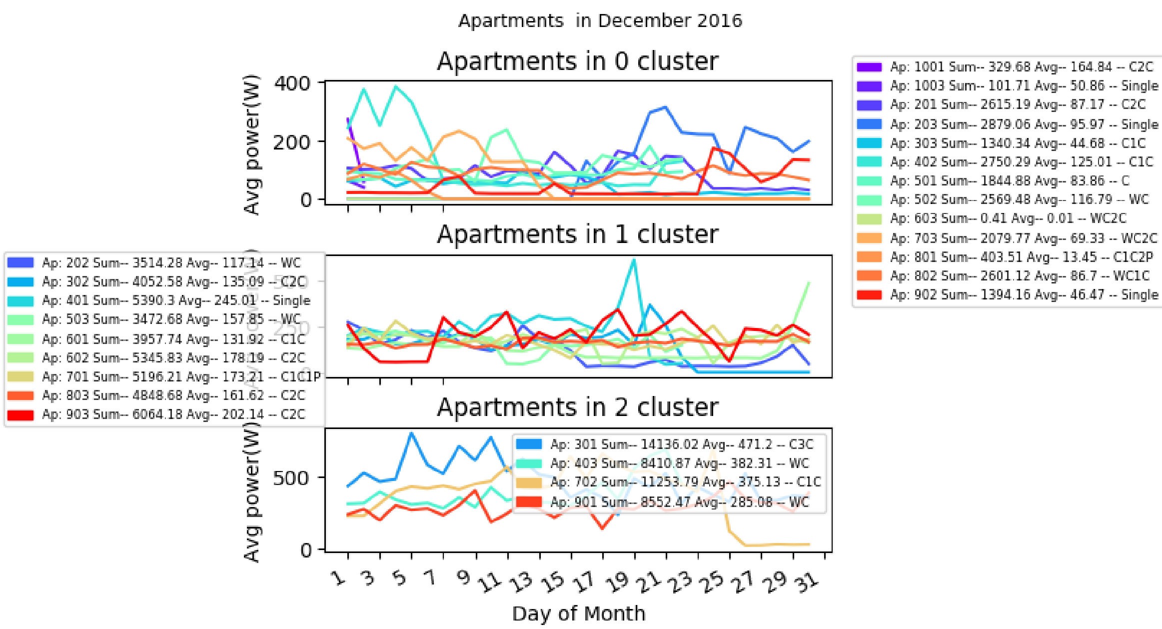

Figure 4.18: Clusters in the month of December 2016

4.4.1 Interpretation of Clustering Results

After interpreting the clustered plots of the apartments we observed the following points

- In most of the cases we got the apartments belonging to the same category(Single Person/Couple/Couple with Children/Working Couple etc.) are in the same cluster.
- The clusters in which the apartments are having families belonging to either Single Person category or Working couple category are having the total consumption of the apartments lower compared to other clusters.
- Sometimes we are observing deviations from the above mentioned point and in such cases the apartments are either belonging to higher floor level or belongs to either 1 series or 3 series.
- We also observed that not always the apartments belonging to the lower floor level are in the cluster with lesser total power consumption.
- We can also observe that family size also has an impact along with floor level.

4.5 Conclusions

Most of the apartments had relatively higher consumption patterns in May, June, July compared to other months. By analyzing series wise plots, we found out that although more upper floors to be having relatively higher consumption patterns, there are also apartments in lower floors showing high patterns.So, this shows that apart from floor level there are also other factors like family size, the family type that impacts power consumption.From the clustering analysis, we found that the apartments with working couple to be having lower power consumption and also the family size to be instrumental in power consumption.

Chapter 5

Regression Analysis

Regression Analysis is a statistical procedure for finding the relationship among variables. It mainly focuses on finding the relationship between the dependent variable and one or more independent variable(s). A dependent variable is the target variable which we are trying to analyze, and the independent variables are those variables that we assume to have an impact on the dependent variable. Thus, by using regression analysis we can find out [20] the

- The significance of the relationship between the dependent and independent variables.
- The strength of the influence of one or more independent variable(s) on the dependent variable.

In this study, we would like to explore how energy consumption of a typical residential buildings tenant is influenced by socio-economic and dwelling factors such as floor number, floor orientation, etc.

5.1 Linear Regression

In this regression, the predictor variable is continuous while the independent variables can either continuous or discrete, and the regression line should be linear. Linear regression is represented by the following equation

$$Y = a + b * X + e$$

where a is the intercept, b is the slope and e represents the error term.

In the case of multiple linear regression we have multiple independent variables in contrast to the simple linear regression where we have one independent variable. Multiple linear regression can be represented by

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_n X_n + e$$

where $\mathrm{b_0}$ is the intercept , $\mathrm{b_1, b_2.....,b_n}$ are the coefficients of the independent variables $X_1, X_2...., X_n$ and e is the error.

Least square method is used for fitting the regression line by calculating the parameters of the best fit line that minimizes the sum of the squared prediction errors.

5.1.1 Interactions

An interaction occurs when the effect of variable on the dependent variable depends on the value of another variable.It means the influence of two variables on the dependent variable is not additive.Interactions occur mostly in regression analysis, ANOVA (Analysis of Variance).The regression equation that includes two way interaction effect can be represented by

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 + e$$

Here the term $\mathrm{X_1} X_2$ is the interaction term which is the product of the two dependent variables.Thus the coefficient term $\mathrm{b_3}$ can be interpreted as the amount of change in the slope of Y on X_1 when X_2 changes by a unit. [21]

5.1.2 Use of lm() function of R language

We have used linear regression to find out the significance of each of the features on the total power consumption. In R, we have a function with the name lm() which is used for fitting a simple regression model. The most important parameters [22] for this function are the

- formula- which is used to describe the parameters for the model.The format for specifying formula is

$$Y \sim X$$

  where Y is our dependent variable and X is the independent variable.

- data - the data set which we are working on

To summarize the model, we can use summarize() function which gives most of the statistical information about the model like t-test, F-test and significance values. [22]

5.1.3 Experiments and Results

We have tried to analyze the influence of some of the socio-economic and dwelling factors on the power consumption.

5.2 Understanding the influence of Floor level on power consumption

To find the influence of floor level on the power consumption of apartments we have used regression analysis taking the floor level as the feature.

Upon regression analysis on the data set of the year 2015, which contains all south-east facing apartments, we got the following output.

Multiple R-squared: 0.4956, Adjusted R-squared: 0.4487
F-statistic: 10.56 on 8 and 86 DF, p-value: 3.29e-10

So based on the p-value obtained,we can conclude that the floor level has a significant positive impact on the power consumption for the given data set containing south-east facing apartments and explains around 45 % of the variance in the output.

When the same experiment is repeated on year 2016 data set we got
F-statistic: 6.187 on 7 and 80 DF, p-value: 8.454e-06.

Based on the p-value here, the floor level is having a significant impact on the power consumption for the given data and explains 30 % of the variance in the output.

For the year 2017 and apartments in the south east direction we got
Multiple R-squared: 0.4421, Adjusted R-squared: 0.3907
F-statistic: 8.603 on 7 and 76 DF, p-value: 9.887e-08

Based on p-value here floor level is a significant feature for the given data and explains 39 % of the variance in the output.

Now experimenting on the apartments of south-facing direction, in the year 2014 we found the floor level to have a significant impact on the power consumption explaining 11% of variance in the power consumption and in 2015 floor level is explaining 34 % of variance in power consumption, in 2016 it is 25 % and in 2017 it is 33 %. When the data is tested with all years consisting of apartments of south-facing direction, floor level is an important feature explaining 17 % of the variance in power consumption.

For the apartments in the south-west facing direction, in the year 2014 we found the floor level to have a significant impact on the power consumption explaining 15% of variance in the power consumption and in 2015 floor level is explaining 29 % of variance in power consumption, in 2016 it is 19 % and in 2017 it is 17 %. When the data is tested with all years consisting of apartments of south-west facing direction, floor level is a important feature explaining 11 % of the variance in power consumption.

When all the apartments are considered over all years floor level is a significant feature explaining 15 % of variance in power consumption.

When family size and a group variable indicating if there is a working couple or not are also included, for the year 2015 (with weekdays and evening data), we found all the features are significant features.

5.3 Understanding the influence of orientation and interactions between floor level, orientation of the building, family size and family type on power consumption

Now considering the effect of orientation also. As already mentioned, the apartments in series one, two and three are oriented towards south-east, south and south-west directions respectively.

We have experimented with weekdays evening data of all apartments for the year 2014 with orientation also added as a feature. After the outlier removal, we found out that orientation is a significant feature in 2014 dataset (4 % variance in total output is explained by orientation). And the order of consumption is South<South-East<South-West($2<1<3$). Although Floor level has a positive effect on the power consumption(as the floor level increases the power consumption increases), feature floor alone is having p-value $0.08>0.05$ making it not significant at 5 % significance level.The feature indicating the presence of a working couple is found to be significant feature.

The combined model with interactions that include interactions between floor level and orientation of the building, floor level and family size, floor with group , group with orientation showed significant positive effect of floor and family size(i.e with increase in these features, we can see an increase in power consumption) and significant effect of group and orientation where South East facing apartments are at a lower level followed by south facing and south west oriented apartments. The total R-squared value for this model is 0.15

While working on data set in 2015, we found similar results as that of 2014 and even the floor level also determined and a significant feature. Upon adding interactions, we got a model with R-square value of 0.24.

For the year 2016 and 2017, we found out that orientation, floor level as a significant feature.The feature indicating the presence of working couple is found to be significant. When interactions between the features are also included, we got a model with an R-square value of 0.17.

When the model is trained with all the years at once, we got a combined model with interactions included to be having an R-squared value of 0.12.

When the same experiments are repeated with a data set of weekdays evening data that are added month wise, we found an increase in the R-squared value of 0.20. Thus the features are explaining 20% of the variance in total consumption.

We have also found that in all the years, in a family with a working couple/a single person, the power consumed is decreasing with an increase in floor level compared to a normal family where it shows a positive effect.This explains the plots where even lower level floors are shown to have high trends. An apartment from the south-west direction is shown to have higher consumption than from the apartment in the south orientation.Also, with increase in family size , floor level is also shown to have a positive effect on the power consumption.

5.4 Conclusions

Our experimentation corroborated that a residential apartments dwellers power consumption very much depends on their apartments orientation and floor level. We have tried to see if the number of children or presence of children in the family does have a significant impact on the power usage of the usage of the household, but the results have showed that they do not have a significant impact on the households.

Chapter 6

Design and Development of EMonitor: An Android app for the perusal of tenants

As a part of the thesis work , we have developed a mobile app called **EMonitor** that would enable the residents of the apartments to monitor their energy consumption levels.

6.1 Features

The features which are included with this app are:

- The user is able to view the current day's power consumption of his/her Apartment with 10-minutes sampling interval.
- The user can compare his consumption with the average, minimum and maximum levels of the building and also that of other apartments(Privacy settings are included here)
- User can also view the historical statistics of their power consumption in the following granularity levels.

1. **Day-Wise**: User can select a particular date and view his consumption.
2. **Month-Wise**: User can view patterns month wise
3. **Year-wise**: User can see historical patterns in based on year.

- Even in the statistics section, user can compare his data to others at all levels of granularity.
- We also included a feature that would give average power consumption for the time range selected and the estimated bill amount. So by this the user can find out if the power consumption is high, medium or low and can take necessary measures if required.
- We also implemented a privacy feature in this app whereby , the faculty members(users) with registered official email-id can have access to install this app.In case if app is installed with any other email-id, then a ***push notification*** is sent to the official user of this app who can either accept or reject the request. This would be useful if other family members of the same apartments wants to use the app.

6.2 Technologies Used

6.2.1 Front-End

As it is an android app, the front end is developed using XML and Android programming.

6.2.2 Back-End

We have used **Firebase** as back-end database for our app. App interacts with back-end through **REST API** calls.For getting the dynamic real time data which at a private cloud server, we have used R code to extract that dynamic and then uploaded to firebase database using a python script which has the facility to connect with firebase database.

This python script is scheduled to run every 10 minutes that will extract data from portal and upload it into firebase database.

6.3 Screenshots

Figure 6.1 is the login screen of EMonitor app.User is authenticated through google authentication. If a new user tries to log in through unofficial email id, first a push notification is sent to the official user and only after the official user accepts the request, the new user can use the app.

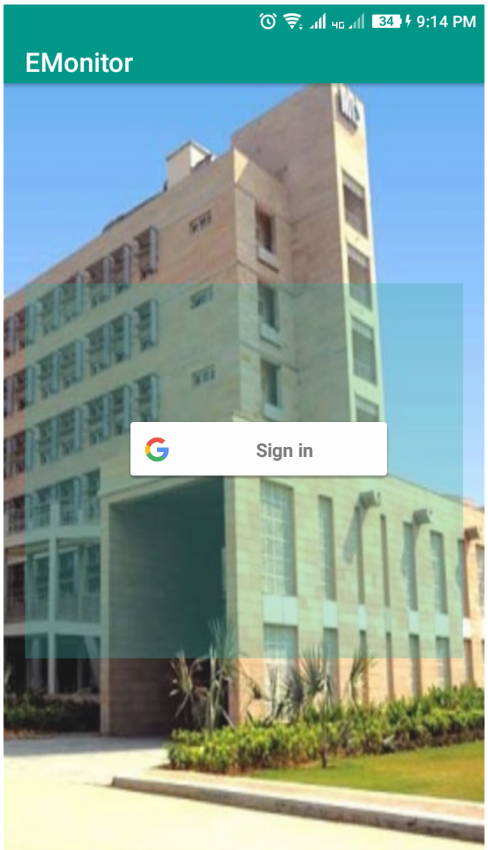

Figure 6.1: Login Screen

After the user is authenticated, for the first time app installation we get a registration page as shown in Figure 6.2. Here, the user can enter the basis details of the apartment.

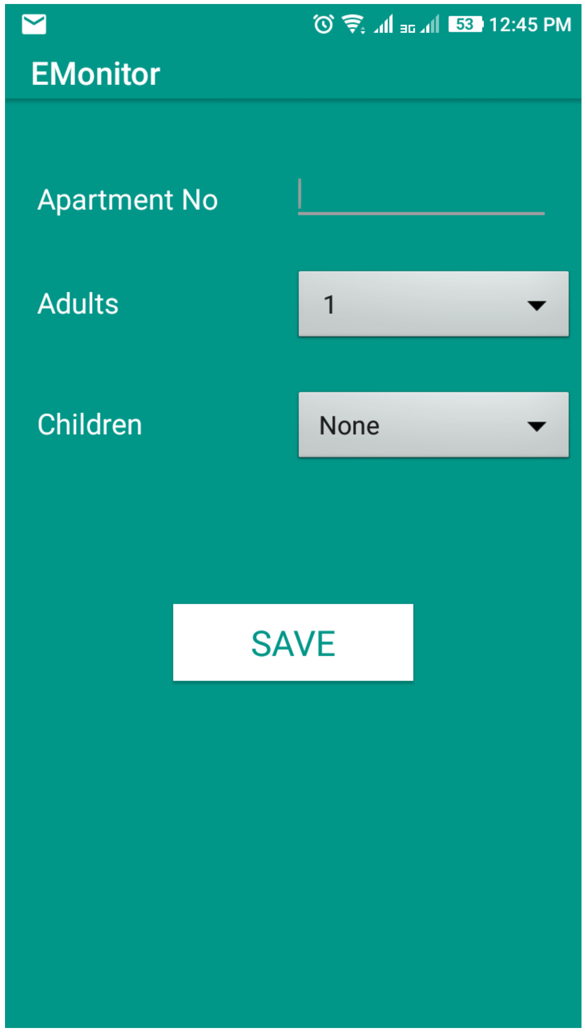

Figure 6.2: Registration page

After successful registration/login, the user is directed to the screen in Figure 6.3, where he can view the real time power trends on that day, average power consumption and number of units consumed.

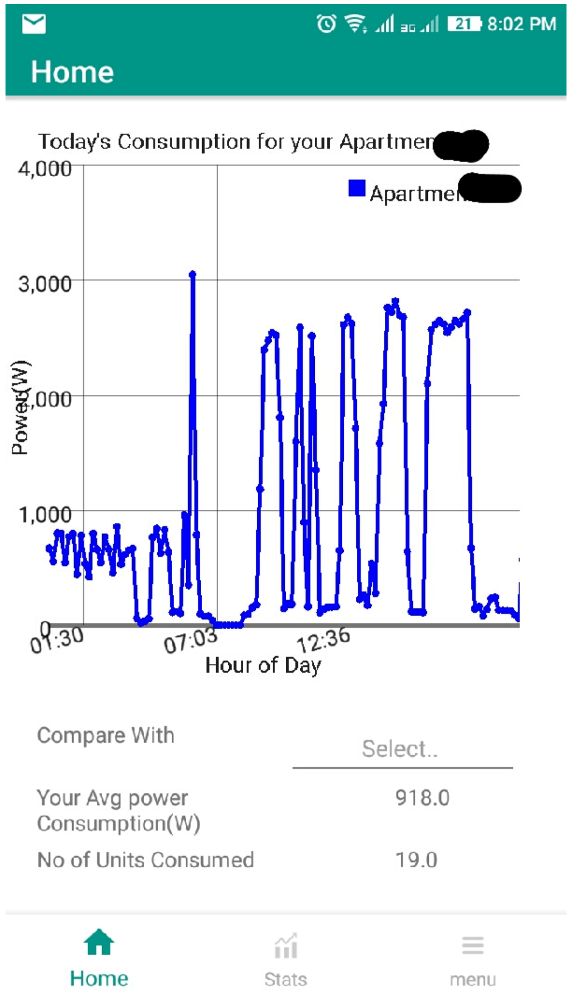

Figure 6.3: Home Screen-Real time Consumption

If a user selects the option to compare with other apartments, the user gets a dialog box as in Figure 6.4, where he/she can choose at most three options to compare the trends with. The comparative trends will be shown as in Figure 6.5

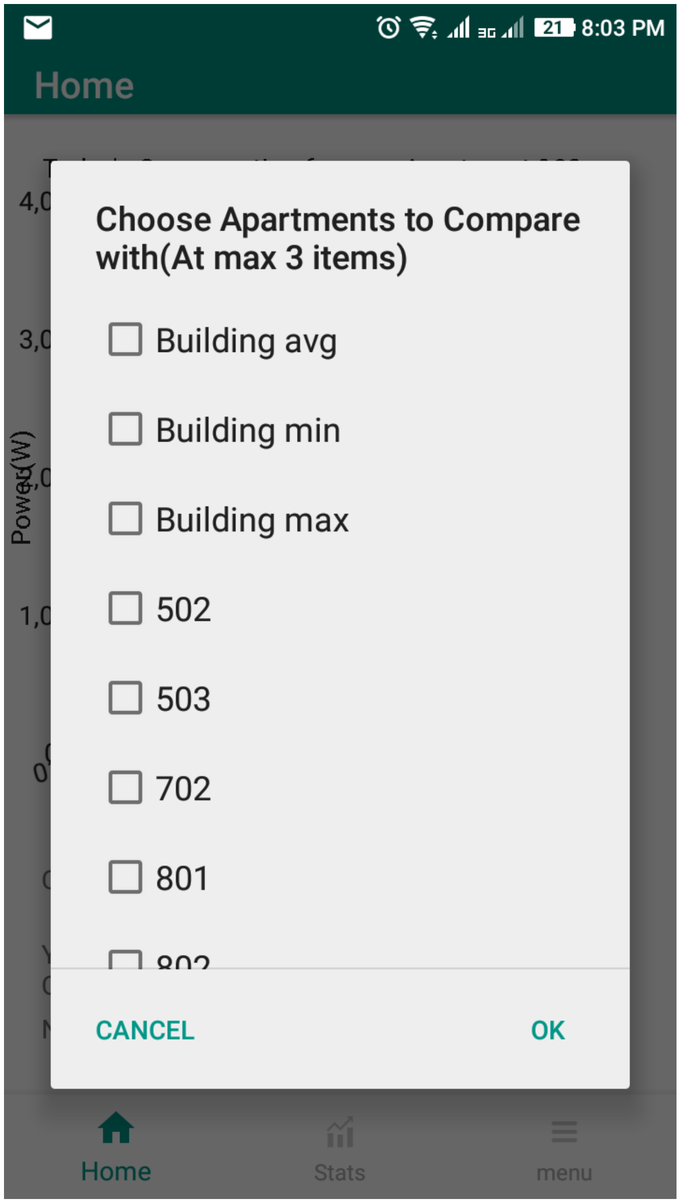

Figure 6.4: Comparison Dialog Box

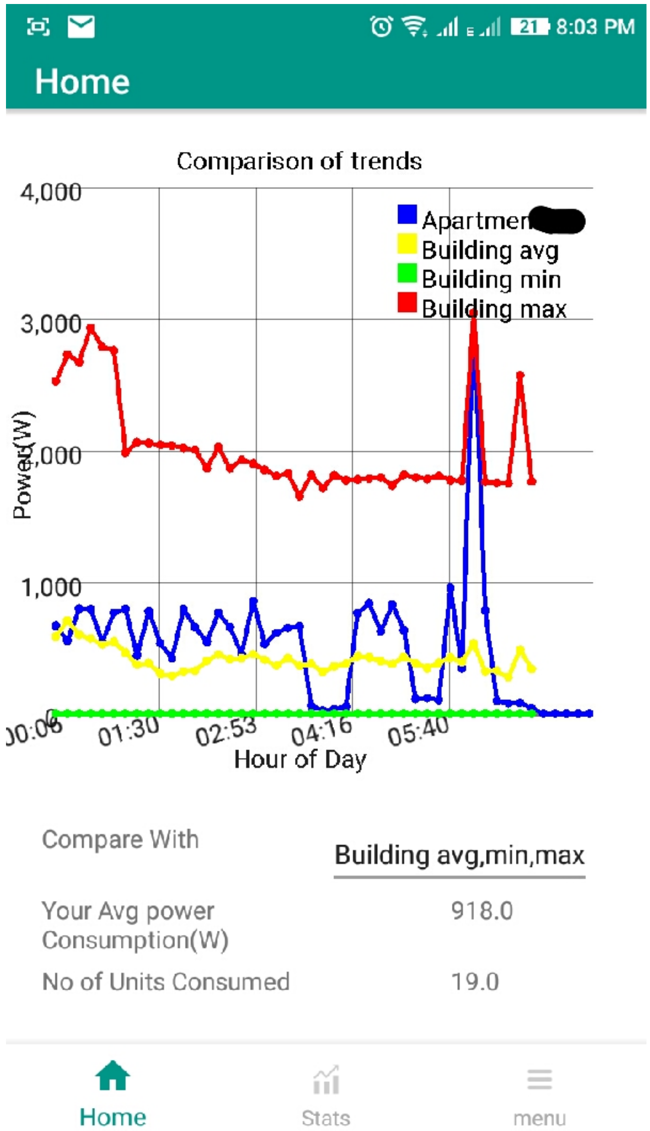

Figure 6.5: Comparison of trends

Figure 6.6 shows the screen where we can select a particular date and view the consumption on that day.

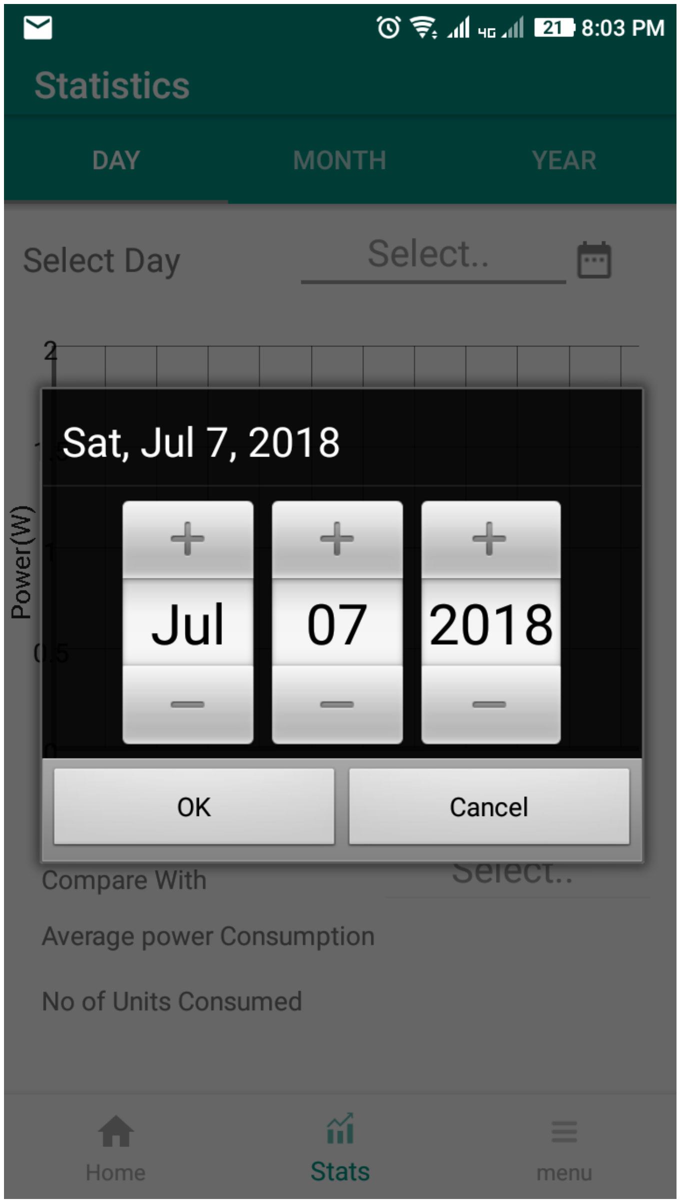

Figure 6.6: Day wise Stats

Figure 6.6 shows the screen where we can select a particular month and year and view the consumption on that day.

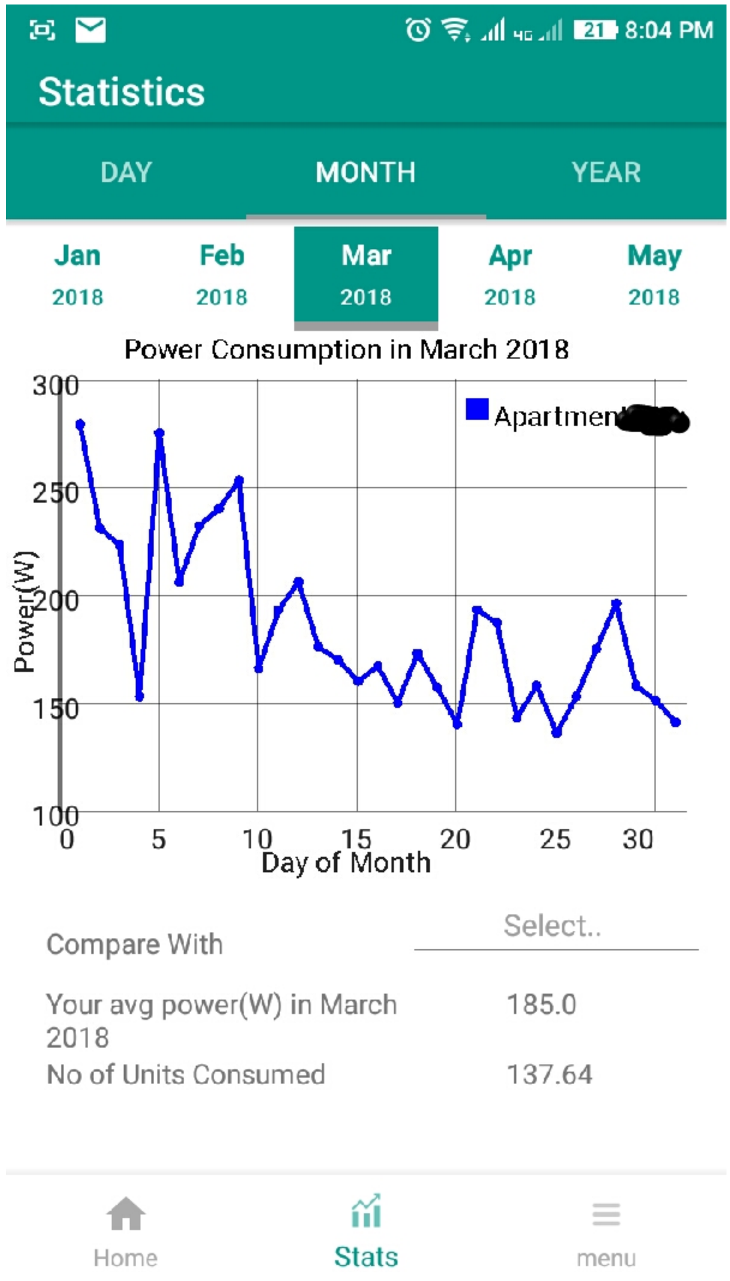

Figure 6.7: Month Wise stats

Figure 6.8 shows the option to select an year to view the historical stats.

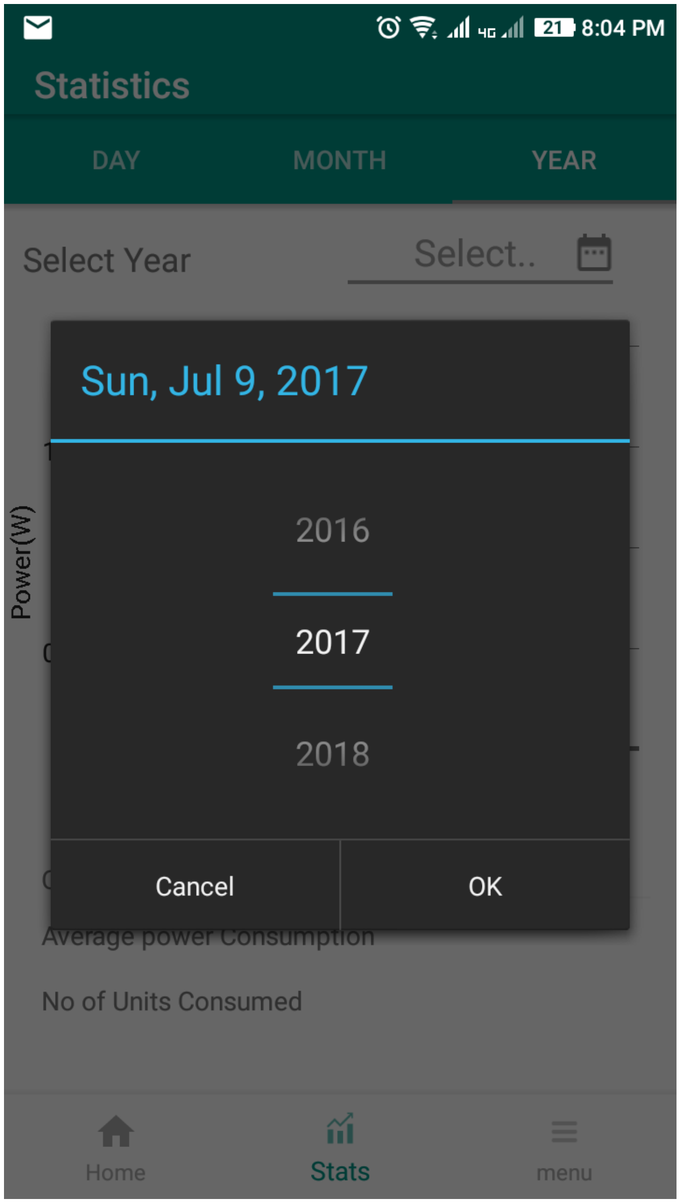

Figure 6.8: Year wise statistics

Figure 6.9 where the user has the option to update the apartment details and can see any notifications he got.

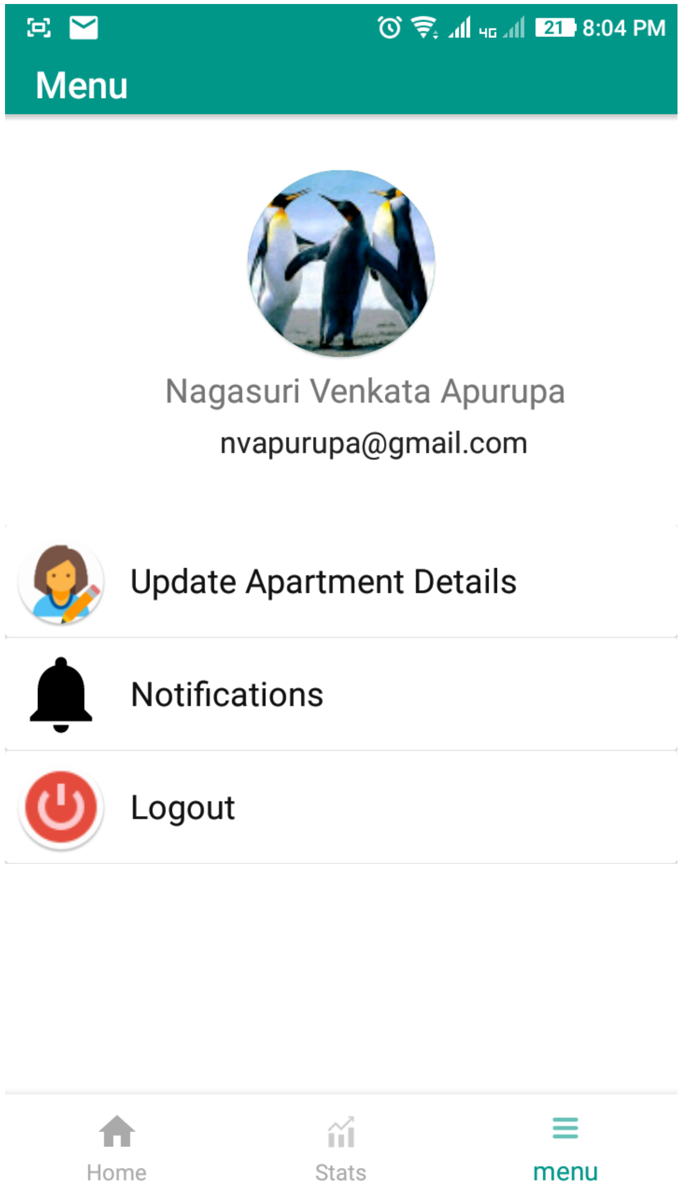

Figure 6.9: Menu Screen

6.4 Conclusions

In this chapter, we have explained the features of our android app, EMonitor and the technologies and databases used. We have deployed our app and scheduled a script to continuously update the database with the real time data.

Chapter 7

Conclusions and Future Work

Our investigations confirm that most of the apartments to be having relatively higher consumption patterns in the months of May, June, July compared to other months.

From our clustering analysis, we found that the apartments with working couple to be having lower power consumption and also the family size is instrumental in power consumption.

From our regression analysis, we found that floor level, orientation, family size, the presence of a working couple in the residential plot are shown to have a significant impact on the power consumption. While the floor level has a positive impact on the power consumption, the families having a working couple are having less consumption patterns compared to others. Our study has used only 27 apartment's power consumption details for 4-5 years. In order to have deeper insights of the power consumption patterns of Indian residential building's tenants, we need to collect power consumption data of few thousands of tenants along with their climatic details such as rainfall, temperature, humidity, wind speed, wind direction, the appliances they are hosting along with their wattage, when they have switched on/offed them(usage details), etc. Also, we believe that the power consumption might be related to the social background of tenants also. Thus, we need their socio-economical details also to exploit the possible associations of them with power consumption.

By studying the impact of these features, we can also develop a model that given the family size, floor level and orientation of the building that and other occupancy details, we could predict whether the family consumes higher amount of energy or not during a given period of year. We could also develop a system that could predict given the family details, which apartments are likely to be more energy efficient while choosing an apartment in a similar building structure. Including the impact of occupant behaviour in our studies could help us more in designing energy efficient buildings.

Bibliography

[1] Aditya Chunekar, S Varshney, and S Dixit. Residential electricity consumption in india: What do we know. *Prayas (Energy Group), Pune*, 4, 2016.

[2] Gesche M Huebner, Ian Hamilton, Zaid Chalabi, David Shipworth, and Tadj Oreszczyn. Explaining domestic energy consumption–the comparative contribution of building factors, socio-demographics, behaviours and attitudes. *Applied energy*, 159:589–600, 2015.

[3] Rory V Jones Kevin J Lomas. Determinants of high electrical energy demand in uk. *Database*, 60:462–480, 2015.

[4] Koen Steemers and Geun Young Yun. Household energy consumption: a study of the role of occupants. *Building Research & Information*, 37(5-6):625–637, 2009.

[5] Yufei Han, Xiaolan Sha, Etta Grover-Silva, and Pietro Michiardi. On the impact of socio-economic factors on power load forecasting. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 742–747. IEEE, 2014.

[6] A. Druckman and Thomas Jackson. Household energy consumption in the uk: A highly geographically and socio-economically disaggregated model. *Energy Policy*, 36(8):3167–3182, 2008.

[7] Gwendolyn Brandon and Alan Lewis. Reducing household energy consumption: a qualitative and quantitative field study. *Journal of Environmental Psychology*, 19(1):75–85, 1999.

[8] Xavier Labandeira, José M Labeaga, and Miguel Rodríguez. A residential energy demand system for spain. *The Energy Journal*, pages 87–111, 2006.

[9] Maya Yashwanth Pai et al. Effect of building orientation and window glazing on the energy consumption of hvac system of an office building for different climate zones. *International Journal of Engineering Research & Technology (IJERT)*, 4(9):838–843, 2015.

[10] Daniel Wiesmann, Inês Lima Azevedo, Paulo Ferrão, and John E Fernández. Residential electricity consumption in portugal: Findings from top-down and bottom-up models. *Energy Policy*, 39(5):2772–2779, 2011.

[11] Amir Kavousian, Ram Rajagopal, and Martin Fischer. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, 55:184–194, 2013.

[12] Yigzaw G Yohanis, Jayanta D Mondol, Alan Wright, and Brian Norton. Real-life energy use in the uk: How occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings*, 40(6):1053–1059, 2008.

[13] Anxiao Zhang, Regina Bokel, Andy van den Dobbelsteen, Yanchen Sun, Qiong Huang, and Qi Zhang. The effect of geometry parameters on energy and thermal performance of school buildings in cold climates of china. *Sustainability*, 9(10):1708, 2017.

[14] Haroon Rashid and Pushpendra Singh. Energy disaggregation for identifying anomalous appliance. In *Proceedings of the 4th ACM International*

Conference on Systems for Energy-Efficient Built Environments, page 41. ACM, 2017.

[15] Michail Vlachos, Christopher Meek, Zografoula Vagena, and Dimitrios Gunopulos. Identifying similarities, periodicities and bursts for online search queries. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 131–142. ACM, 2004.

[16] Meng Ma, Arjan Van Genderen, and Peter Beukelman. Developing and implementing peak detection for real-time image registration. In *Proceedings of the 16th Annual Workshop on Circuits, Systems & Signal Processing (ProRISC2005)*, pages 641–652, 2005.

[17] Girish Palshikar et al. Simple algorithms for peak detection in time-series. In *Proc. 1st Int. Conf. Advanced Data Analysis, Business Analytics and Intelligence*, pages 1–13, 2009.

[18] `https://www.datascience.com/blog/k-means-clustering/`.

[19] `https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm/`.

[20] Analytics vidhya. `https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#/`.

[21] `http://quantpsy.org/interact/interactions.htm/`.

[22] `https://www.r-bloggers.com/r-tutorial-series-simple-linear-regression//`.