Journal of Renewable Energy and Environment

Journal Homepage: www.jree.ir

Materials and Energy Research Center

MERC

Research Article

Analytical Approach to Exploring the Missing Data Behavior in Smart Home Energy Consumption Dataset

Kasaraneni Purna Prakash [a], Yellapragada Venkata Pavan Kumar [b*]

[a] School of Computer Science and Engineering, VIT-AP University, Amaravati-522237, Andhra Pradesh, India.
[b] School of Electronics Engineering, VIT-AP University, Amaravati-522237, Andhra Pradesh, India.

P A P E R I N F O

Paper history:
Received: 04 November 2021
Revised in revised form: 19 December 2021
Scientific Accepted: 12 December 2021
Published: 07 March 2022

Keywords:
Behavior Analysis,
Data Analysis,
Energy Consumption Data,
Missing Data Anomalies,
Smart Homes,
Smart Meter Data

A B S T R A C T

Smart homes are considered to be the subset of smart grids that have gained widespread popularity and significance in the present energy sector. These homes are usually equipped with different kinds of sensors that communicate between appliances and the metering infrastructure to monitor and trace the energy consumption details. The smart meters trace the energy consumption data continuously or in a period of intervals as required. Sometimes, these traces will be missed due to errors in communication channels, an unexpected breakdown of networks, malfunctioning of smart meters, etc. This missingness greatly impacts smart home operations such as load estimation and management, energy pricing, optimizing assets, planning, decision making, etc. Moreover, to implement a suitable precautionary measure to eliminate missing of data traces, it is required to understand the past behavior of the data anomalies. Hence, it is essential to comprehend the behavior of missing data in the smart home energy consumption dataset. In this regard, this paper proposes an analytical approach to detect and quantify the missing data instants in all days for all appliances. Using this quantification, the behavior of missing data anomalies is analyzed during the day. For the analysis, a practical smart home energy consumption dataset 'Tracebase' is considered. Initially, the existence and the count of missing instants are computed. From this, the appliance 'MicrowaveOven' is considered for further analysis as it comprises the highest count of missing instants (84740) in a day when compared to all other appliances. Finally, the proposed analysis reveals that the large number of missing instants is occurring during the daylight period of a day.

https://doi.org/10.30501/jree.2021.313536.1277

1. INTRODUCTION

In recent years, smart homes have become very popular and grabbed the attention of people around the world. Smart homes provide access to a modern style of living with greater comfort and security. Besides, smart homes enable control over the appliance functionality, energy usage, billing, etc. Hence, the consumers are ready to avail the benefits of smart homes and make their homes automated. This automation includes different kinds of sensors, communication channels, computer-controlled equipment, etc., which are formed as a controlled network. This installed equipment captures the energy consumption data 24 × 7 from all the appliances connected in a smart home. The analysis of this data is essential to understanding the functionality of appliances. For this purpose, the availability of high-quality data is always desired. But, the data capturing process is often associated with certain anomalies due to several problems and failures in the power and communication networks. Among such anomalies, missing data records is a major issue, which deludes the analysis and decision-making about energy consumption.

There are several literature works available on the analysis of smart home datasets and detection of various anomalies present in it, as described in Table 1. All these state-of-the-art literature works can be segregated as works related to general concepts, complexities, challenges, and advancements in smart homes; IoT role in the smart home application; smart home environment, technology, and energy management; data analytics in smart grids/homes; data anomalies and their detection. As per the description provided in Table 1 on these works, it is clearly understood that all these works represent the preliminary requirements or supports for the smart home deployments. Further, in the context of data anomalies and missing data, conventional works have focused on identification, preprocessing, and visualization. These approaches help to rectify the data anomalies, thereby improving the data quality.

Along with the preprocessing methods available in the literature, it is also important to have some precautionary measures to avoid data quality issues. To identify the cause of data quality issues or implement a suitable precautionary measure, it is important to know the behavior of the data

*Corresponding Author's Email: pavankumar.yv@vitap.ac.in (Y.V. Pavan Kumar)
URL: https://www.jree.ir/article_146038.html

Please cite this article as: Purna Prakash, K. and Pavan Kumar, Y.V., "Analytical approach to exploring the missing data behavior in smart home energy consumption dataset", *Journal of Renewable Energy and Environment (JREE)*, Vol. 9, No. 2, (2022), 37-48. (https://doi.org/10.30501/jree.2021.313536.1277).

anomaly. However, there is no work in the literature that discussed the behavioral analysis of missing data anomalies present in the smart home energy consumption data, to the best of the authors' knowledge. This is a major gap in the literature. To address this gap, this paper proposes an analytical approach to exploring the missing data behavior in the smart home energy consumption dataset. This proposed approach quantifies the missing data records in all days for different appliances and analyzes the behavior of these missing data anomalies. The proper identification and behavior analysis of missing data anomalies enables the engineers to implement preventive measures to stop the occurrence of missing data anomalies. This is the motivation, main idea, and novelty of the proposed work of this paper.

The remaining part of the paper is organized as follows. Section 2 presents the methodology of the proposed approach. Section 3 presents the simulation results with their analysis, Finally, Section 4 concludes the findings and achievements of this paper along with the future scope.

Table 1. Review of literature works

Key topic	Reference	Year	Author(s)	Description of the literature work carried
General concepts, complexities, challenges, and advancements in smart homes	[1]	2021	Zielonka et al.	Performed a study and extensive analysis on the recent trends and advancements in smart homes to learn that how they support the users.
	[2]	2021	DeFranco et al.	Emphasized that smart homes were advanced and complex systems. To cope up with this complexity and for further improvement of smart homes' functionality, a comprehensive review and analysis are carried out.
	[3]	2021	Pira	Presented the social issues associated with living in smart homes and made suggestions to reduce the effect of those issues.
	[4]	2020	Kim et al.	Focused on developing design solutions based on user-centered scenarios that include the health issues and daily activities of users.
	[5]	2020	Benjamin et al.	Discussed the pros and cons of smart home technologies by examining the real data drawn in the United Kingdom.
	[6]	2020	Diahovchenko et al.	Reviewed the development and challenges involved in distributed generation, energy storage technologies, deployment of smart meters, microgrids, etc.
IoT role in smart home application	[7]	2021	Wonyoung et al.	Performed a thorough bibliometric study to understand the key trends and the role of the internet of things (IoT) in smart homes.
	[8]	2020	Lin et al.	Discussed the utilization of IoT platforms in the development of smart home applications such as PlantTalk, FishTalk, BreathTalk, TheaterTalk, FrameTalk, and GardenTalk. All these applications were developed under the project 'HomeTalk' which facilitates the flexibility of using appliances.
	[9]	2019	Almusaylim et al.	Conducted a review on the current status and challenges incurred with the implementation of IoT in smart homes.
	[10]	2017	Chen et al.	Introduced a new version of the smart home i.e., Smart Home 2.0. This was designed and implemented using botanical IoT and emotional detection.
Smart home environment, technology, and energy management	[11]	2021	Rasha	Reviewed smart home energy management schemes and also discussed the challenges implicated in smart home power quality.
	[12]	2021	Zhibin et al.	Presented a Spatio-temporal graphical analysis method to understand the behavior of users' energy requirements based on the analytics of smart meter data.
	[13]	2020	Yamauchi et al.	Realized approaches to recognize users' behavior based on their activities and detect anomalies using sensor data in smart homes.
	[14]	2018	Darby	Emphasized the importance of understanding the viability of smart home technologies and users' roles in the smart home environment.
	[15]	2018	Barsocchi et al.	Presented an affordable, easily installable, and accessible smart home environment in turn to reduce the user efforts in managing and improving smart homes.
	[16]	2018	Albuquerque et al.	Suggested a model to maximize energy efficiency and optimize the level of comfort in smart homes.
	[17]	2017	Fan et al.	Discussed cutting-edge visualization techniques and analyzed their merits and demerits to enhance the efficiency of smart home electricity by perceiving the user habits.
	[18]	2017	Martinez-Pabon et al.	Suggested a methodology to forecast the customers who will be eligible for demand response programs using real-time smart meter data.
	[19]	2016	Hare et al.	Conducted a comprehensive review on different modes of faults occurring in microgrids. This review was carried out on both renewable and traditional energy generation systems.
Data analytics in smart grids/homes	[20]	2020	Kezunovic et al.	Discussed the importance of big data analytics to achieve goals in future power grids.
	[21]	2020	vom Scheidt et al.	Performed an extensive quantitative and qualitative literature review on data analytics in the areas of electric power generation, market, transmission, distribution, and utilization.

	[22]	2019	Wang et al.	Conducted an application-oriented review of data analytics in smart meter data in terms of descriptive, predictive, and prescriptive analytics. This review also discussed various challenges and applications concerned with smart meter data analytics.
	[23]	2016	Chou et al.	Developed a framework based on smart grid data analytics for conserving energy in residential buildings. The electricity cost reduction and optimal scheduling of operations depend on the decision made by the decision support system of this framework.
Data anomalies and their detection	[24]	2021	Prakash et al.	Implemented a simple approach to detecting and quantifying the missing data anomalies in smart home energy consumption data.
	[25]	2021	Gilani Fahad et al.	Implemented an approach to detect the anomalies in daily activities of smart home users.
	[26]	2019	Ariyaluran Habeeb et al.	Reviewed the state-of-the-art technologies for detecting anomalies and discussed the challenges of big data processing in real-time.
	[27]	2018	Moghaddass et al.	Designed a framework to detect anomalies in large volumes of smart meter data.
	[28]	2018	Hela et al.	Implemented an association-rule based approach to anticipate the risk of anomalies in the smart home with regard to the activities of users.
	[29]	2017	Wen et al.	Studied the data quality issues such as incomplete data, noisy data, and outliers in energy consumption data of smart grids.

2. METHODOLOGY

The implementation steps of the proposed analytical approach are shown in Figure 1. The process starts with data preparation. The original Tracebase dataset is available with a single column and in string format [30]. The analysis of such a kind of format is difficult. Hence, the dataset is split into the desired columns such as REC_DATE, REC_HOUR, REC_MINUTE, REC_SECOND, and READING. An appropriate type of conversion is applied. The required variables vec, hourly_missing, and day_missing are initialized. The Tracebase dataset consists of Comma-Separated Value (CSV) files in each subdirectory. Each CSV file represents a full day. To access these CSV files, read the directory and subdirectories. Read each CSV file and proceed with the calculation of the number of instants missing at each hour. To accomplish this, filter the data based on the REC_DATE, REC_HOUR, and REC_MINUTE at each hour 'h' and each minute 'm' [for (h in 0:23) and for (m in 0:59)]. These filtered data are saved into an object called 'instants_traces'. To verify whether any instants are missing in the dataset, compare the values of the variable 'vec' with the seconds of the variable REC_SECOND of instants_traces. This comparison gives the information of instants missing at each hour and saved into the variable hourly_missing. The number of instants on each day is calculated by using hourly_missing information and saved into the variable day_missing. Finally, calculate the maximum instants missing at each hour.

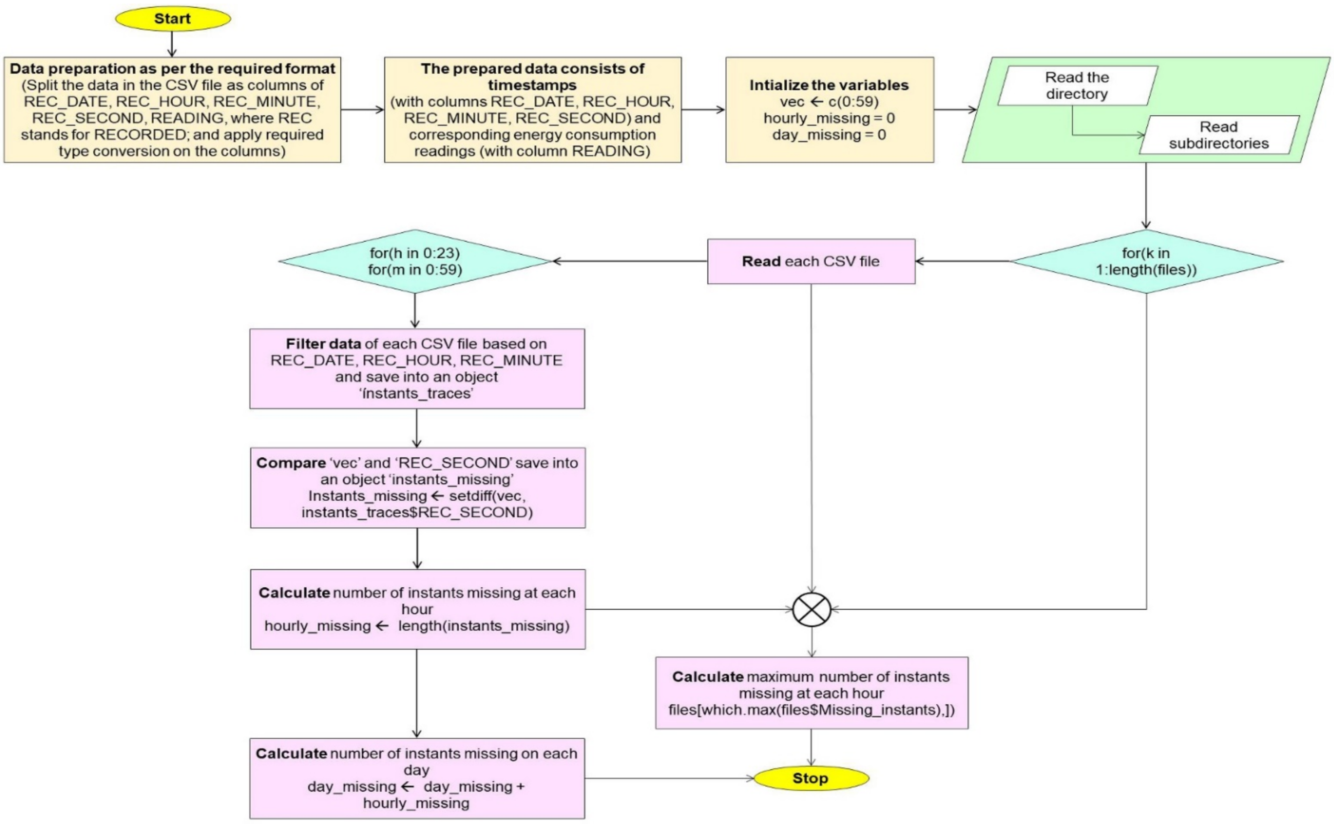


Figure 1. Implementation flow of the proposed analytical approach

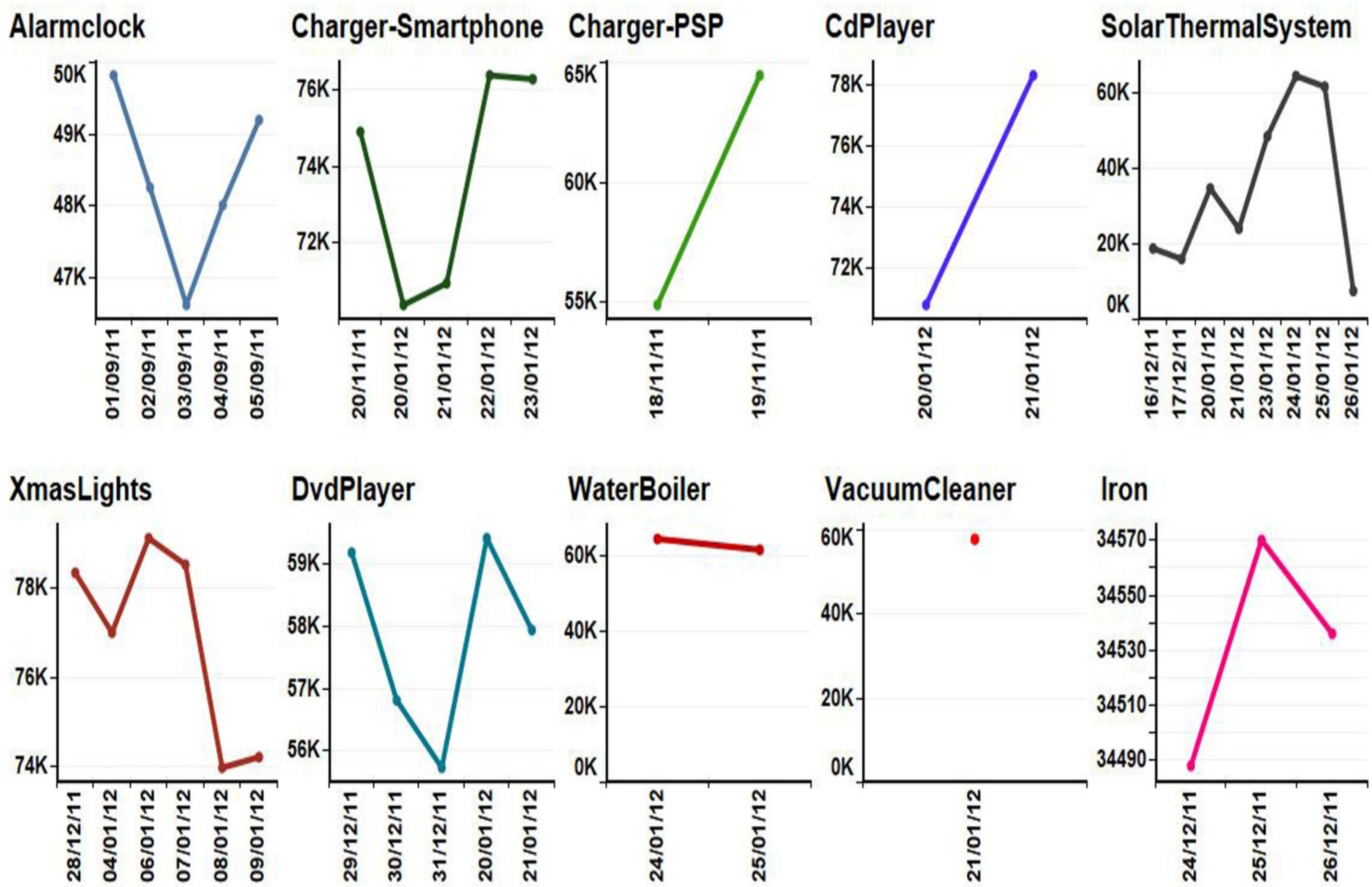
3. RESULTS AND ANALYSIS

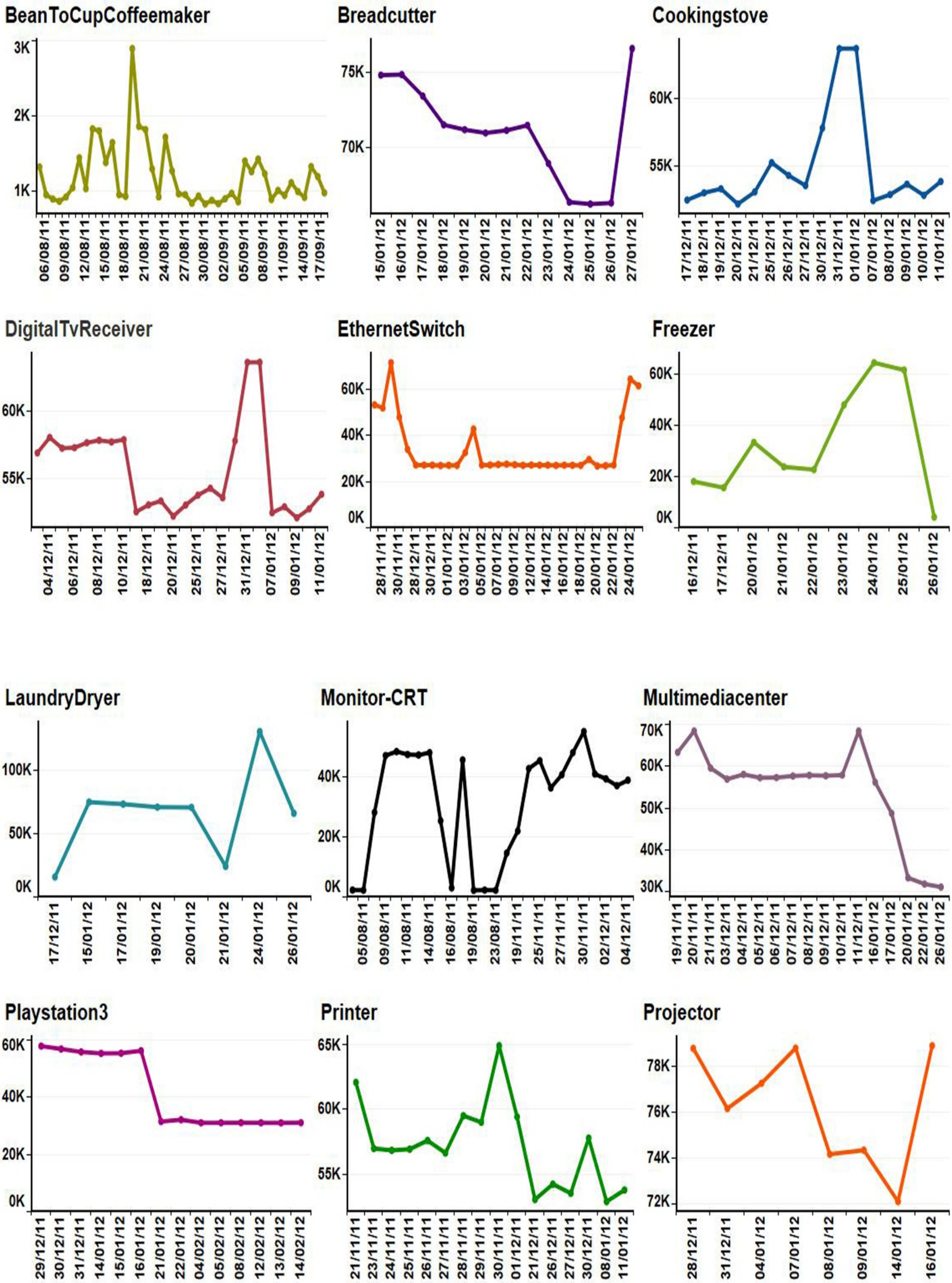
The subplots represented in Figure 2. showcase the variation in the count of missing instants in different appliances of the smart home energy consumption dataset. These subplots are drawn based on the date (the day where the appliance is connected) on the x-axis and the count of missing instants on the y-axis. The observations made from Figure 2. are given below.

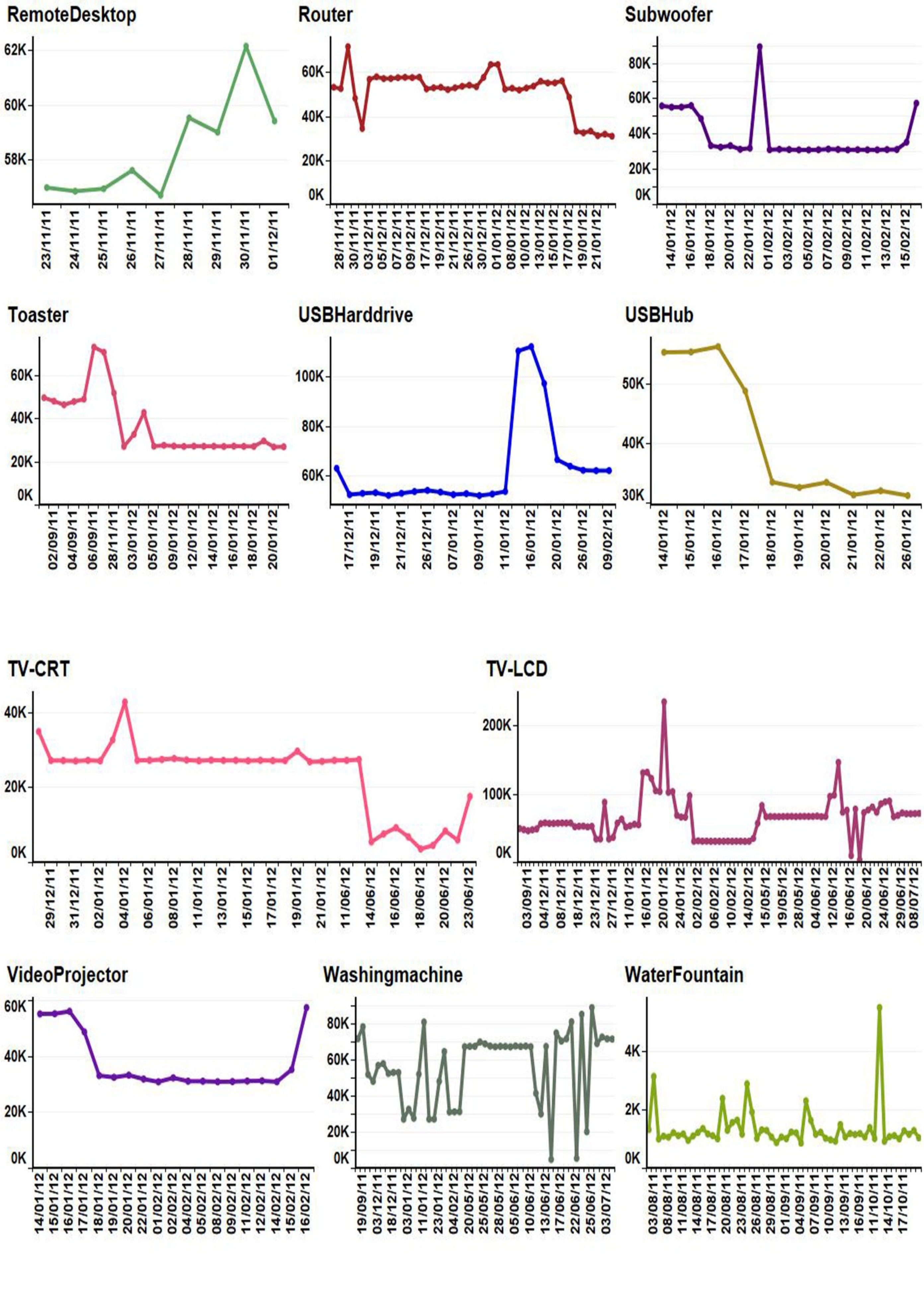
Alarmclock is connected for 5 days and the highest count of missing instants (49826) was observed on 01/09/11, while the lowest count of missing instants (46608) was observed on 03/09/11. Charger-Smartphone is connected for 5 days and the highest count of missing instants (76364) was observed on 22/01/12, while the lowest count of missing instants (70389) was observed on 20/01/12.

Charger-PSP is connected for 2 days and the highest count of missing instants (64442) was observed on 19/11/11, while the lowest count of missing instants (54905) was observed on 18/11/11. CdPlayer is connected for 2 days and the highest count of missing instants (78289) was observed on 21/01/12, while the lowest count of missing instants (70781) was observed on 20/01/12. SolarThermalSystem is connected for 8 days and the highest count of missing instants (64553) was observed on 24/01/12, while the lowest count of missing instants (7560) was observed on 26/01/12. XmasLights is connected for 6 days and the highest count of missing instants (79128) was observed on 06/01/12, while the lowest count of missing instants (73968) was observed on 08/01/12. DvdPlayer is connected for 5 days and the highest count of missing instants (59412) was observed on 20/01/12, while the lowest count of missing instants (55728) was observed on 31/12/11. WaterBoiler is connected for 2 days and the highest count of missing instants (64490) was observed on 24/01/12, while the lowest count of missing instants (61618) was observed on 25/01/12. VacuumCleaner is connected for 1 day and the count of missing instants (57830) was observed on 21/01/12. Iron is connected for 3 days and the highest count of missing instants (34570) was observed on 25/12/11, while the lowest count of missing instants (34488) was observed on 24/12/11.

BeanToCupCoffeemaker is connected for 44 days and the highest count of missing instants (2899) was observed on 19/08/11, while the lowest count of missing instants (827) was observed on 30/08/11. Breadcutter is connected for 13 days and the highest count of missing instants (76612) was observed on 27/01/12, while the lowest count of missing instants (66214) was observed on 25/01/12. Cookingstove is connected for 16 days and the highest count of missing instants (63638) was observed on 01/01/12, while the lowest count of missing instants (52199) was observed on 20/12/11. DigitalTvReceiver is connected for 24 days and the highest count of missing instants (63638) was observed on 01/01/12, while the lowest count of missing instants (52070) was observed on 09/01/12. EthernetSwitch is connected for 33 days and the highest count of missing instants (71657) was observed on 29/11/11, while the lowest count of missing instants (26802) was observed on 20/01/12. Freezer is connected for 9 days and the highest count of missing instants (64565) was observed on 24/01/12, while the lowest count of missing instants (4130) was observed on 26/01/12.




BeanToCupCoffeemaker
3K
2K
1K
Breadcutter
75K
70K
Cookingstove
60K
55K
DigitalTvReceiver
60K
55K
EthernetSwitch
60K
40K
20K
0K
Freezer
60K
40K
20K
0K
LaundryDryer
100K
50K
0K
Monitor-CRT
40K
20K
0K
Multimediacenter
70K
60K
50K
40K
30K
Playstation3
60K
40K
20K
0K
Printer
65K
60K
55K
Projector
78K
76K
74K
72K


RemoteDesktop
62K
60K
58K
23/11/11 24/11/11 25/11/11 26/11/11 27/11/11 28/11/11 29/11/11 30/11/11 01/12/11
Router
60K
40K
20K
0K
28/11/11 30/11/11 03/12/11 05/12/11 07/12/11 09/12/11 17/12/11 19/12/11 21/12/11 26/12/11 30/12/11 01/01/12 08/01/12 10/01/12 13/01/12 15/01/12 17/01/12 19/01/12 21/01/12
Subwoofer
80K
60K
40K
20K
0K
14/01/12 16/01/12 18/01/12 20/01/12 22/01/12 01/02/12 03/02/12 05/02/12 07/02/12 09/02/12 11/02/12 13/02/12 15/02/12
Toaster
60K
40K
20K
0K
02/09/11 04/09/11 06/09/11 28/11/11 03/01/12 05/01/12 09/01/12 12/01/12 14/01/12 16/01/12 18/01/12 20/01/12
USBHarddrive
100K
80K
60K
17/12/11 19/12/11 21/12/11 26/12/11 07/01/12 09/01/12 11/01/12 16/01/12 20/01/12 26/01/12 09/02/12
USBHub
50K
40K
30K
14/01/12 15/01/12 16/01/12 17/01/12 18/01/12 19/01/12 20/01/12 21/01/12 22/01/12 26/01/12
TV-CRT
40K
20K
0K
29/12/11 31/12/11 02/01/12 04/01/12 06/01/12 08/01/12 11/01/12 13/01/12 15/01/12 17/01/12 19/01/12 21/01/12 11/06/12 14/06/12 16/06/12 18/06/12 20/06/12 23/06/12
TV-LCD
200K
100K
0K
03/09/11 04/12/11 08/12/11 18/12/11 23/12/11 27/12/11 11/01/12 16/01/12 20/01/12 24/01/12 02/02/12 06/02/12 10/02/12 14/02/12 15/05/12 19/05/12 28/05/12 04/06/12 12/06/12 16/06/12 20/06/12 24/06/12 29/06/12 08/07/12
VideoProjector
60K
40K
20K
0K
14/01/12 15/01/12 16/01/12 17/01/12 18/01/12 19/01/12 20/01/12 22/01/12 01/02/12 02/02/12 04/02/12 05/02/12 08/02/12 09/02/12 11/02/12 12/02/12 14/02/12 15/02/12 16/02/12
Washingmachine
80K
60K
40K
20K
0K
19/09/11 03/12/11 18/12/11 03/01/12 11/01/12 23/01/12 04/02/12 20/05/12 25/05/12 28/05/12 05/06/12 10/06/12 13/06/12 17/06/12 22/06/12 25/06/12 03/07/12
WaterFountain
4K
2K
0K
03/08/11 08/08/11 11/08/11 14/08/11 17/08/11 20/08/11 23/08/11 26/08/11 29/08/11 01/09/11 04/09/11 07/09/11 10/09/11 13/09/11 16/09/11 11/10/11 14/10/11 17/10/11

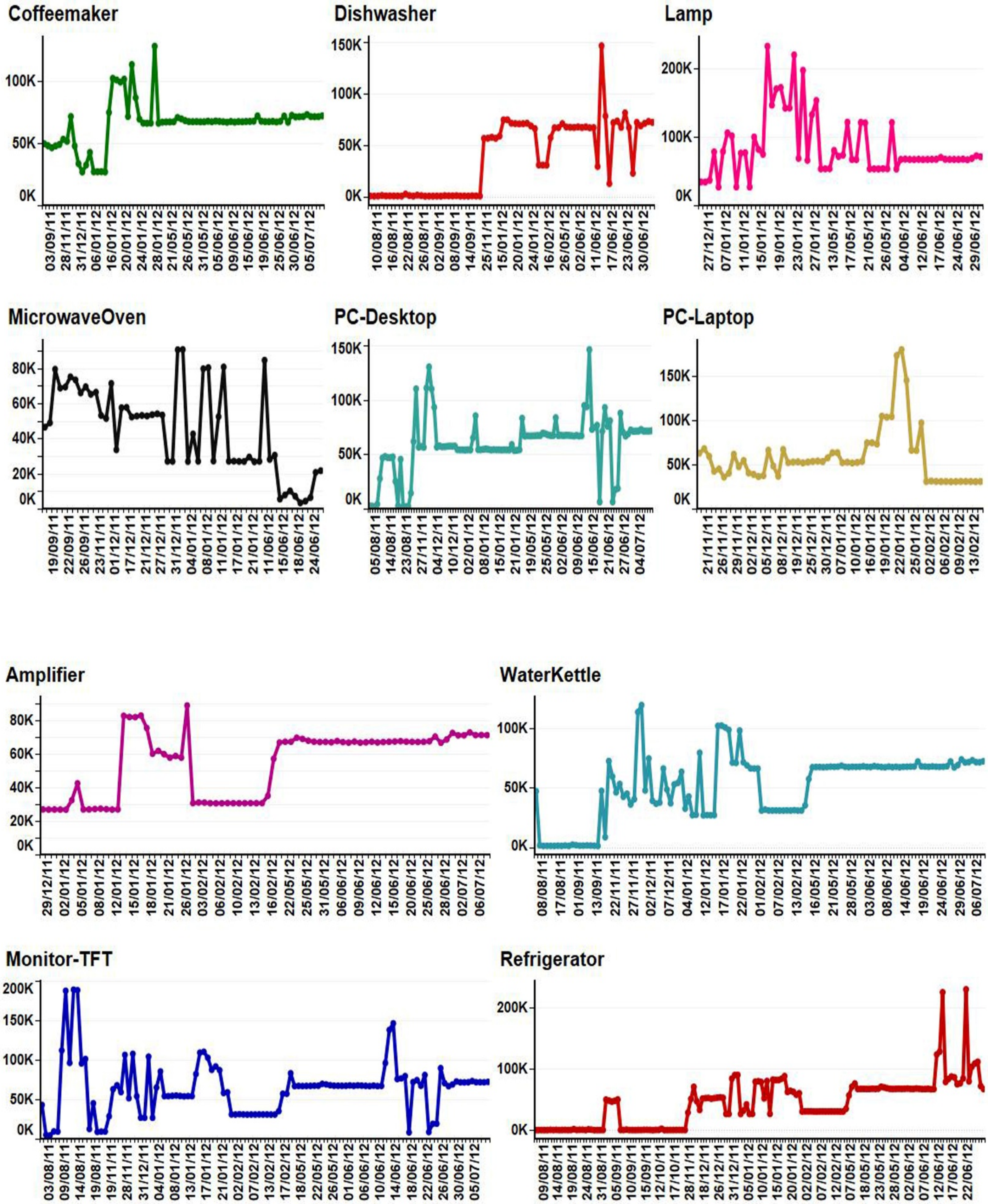


Figure 2. Quantification of missing data anomalies in different appliances (x-axis represents the date and y-axis represents the count of missing instants corresponding to the date)

LaundryDryer is connected for 9 days and the highest count of missing instants (130995) was observed on 24/01/12, while the lowest count of missing instants (15863) was observed on 17/12/11. Monitor-CRT is connected for 26 days and the highest count of missing instants (54858) was observed on 30/11/11, while the lowest count of missing instants (2195) was observed on 19/08/11. Multimediacenter is connected for 17 days and the highest count of missing instants (68409) was observed on 20/11/11, while the lowest count of missing instants (31085) was observed on 26/01/12. Playstation3 is connected for 14 days and the highest count of missing instants (57798) was observed on 29/12/11, while the lowest count of missing instants (30981) was observed on 13/02/12. Printer is connected for 16 days and the highest count of missing instants (64877) was observed on 30/11/11, while the lowest count of missing instants (52919) was observed on 08/01/12. Projector is connected for 8 days and the highest count of missing instants (78873) was observed on 16/01/12, while the lowest count of missing instants (72117) was observed on 14/01/12.

RemoteDesktop is connected for 9 days and the highest count of missing instants (62149) was observed on 30/11/11, while the lowest count of missing instants (56722) was observed on 27/11/11. Router is connected for 40 days and the highest count of missing instants (71741) was observed on 29/11/11, while the lowest count of missing instants (31101) was observed on 26/01/12. Subwoofer is connected for 28 days and the highest count of missing instants (89521) was observed on 26/01/12, while the lowest count of missing instants (31042) was observed on 05/02/12. Toaster is connected for 25 days and the highest count of missing instants (73077) was observed on 06/09/11, while the lowest count of missing instants (26886) was observed on 20/01/12. USBHarddrive is connected for 30 days and the highest count of missing instants (112335) was observed on 16/01/12, while the lowest count of missing instants (52133) was observed on 09/01/12. USBHub is connected for 10 days and the highest count of missing instants (56175) was observed on 16/01/12, while the lowest count of missing instants (31293) was observed on 26/01/12. TV-CRT is connected for 36 days and the highest count of missing instants (42852) was observed on 04/01/12, while the lowest count of missing instants (3654) was observed on 18/06/12. TV-LCD is connected for 119 days and the highest count of missing instants (234864) was observed on 20/01/12, while the lowest count of missing instants (4090) was observed on 18/06/12. VideoProjector is connected for 19 days and the highest count of missing instants (57509) was observed on 16/02/12, while the lowest count of missing instants (30950) was observed on 01/02/12.

Washingmachine is connected for 56 days and the highest count of missing instants (88890) was observed on 25/06/12, while the lowest count of missing instants (4913) was observed on 14/06/12. WaterFountain is connected for 56 days and the highest count of missing instants (5502) was observed on 12/10/11, while the lowest count of missing instants (864) was observed on 04/09/11. Coffeemaker is connected for 82 days and the highest count of missing instants (127944) was observed on 27/01/12, while the lowest count of missing instants (27090) was observed on 31/12/11. Dishwasher is connected for 76 days and the highest count of missing instants (146443) was observed on 13/06/12, while the lowest count of missing instants (944) was observed on 26/08/11. Lamp is connected for 86 days and the highest count of missing instants (232192) was observed on 16/01/12, while the lowest count of missing instants (27237) was observed on 05/01/12. MicrowaveOven is connected for 60 days and the highest count of missing instants (90825) was observed on 01/01/12, while the lowest count of missing instants (3557) was observed on 18/06/12. PC-Desktop is connected for 151 days and the highest count of missing instants (146357) was observed on 13/06/12, while the lowest count of missing instants (2251) is observed on 04/08/11. PC-Laptop is connected for 67 days and the highest count of missing instants (179830) was observed on 22/01/12, while the lowest count of missing instants (30981) was observed on 07/02/12. Amplifier is connected for 89 days and the highest count of missing instants (89375) was observed on 26/01/12, while the lowest count of missing instants (27094) was observed on 02/01/12. WaterKettle is connected for 134 days and the highest count of missing instants (119492) was observed on 29/11/11, while the lowest count of missing instants (1378) was observed on 13/09/11. Monitor-TFT is connected for 190 days and the highest count of missing instants (189047) was observed on 11/08/11, while the lowest count of missing instants (4836) was observed on 03/08/11. Refrigerator is connected for 206 days and the highest count of missing instants (230645) was observed on 21/06/12, lowest count of missing instants (700) is observed on 17/09/11. The observations made on the count of missing instants in different appliances are summarized in Table 2.

The highest count of missing instants for each appliance is plotted as shown in Figure 3. Using this plot, the highest count of missing instants at a particular device in each appliance revealed that the appliance 'MicowaveOven' had the highest count of missing instants. In total, 84740 instants were missing at the device with identifier 'dev_768D06' on 20/05/12. Hence, all the days of MicrowaveOven appliance are considered for further analysis to know how many hours are there with the highest counts of instants missing. For this purpose, the frequency of hours with the highest missing instants in MicrowaveOven appliance is plotted as shown in Figure 4. During the analysis, it is observed that all the hours except hours 1, 4, and 5 are containing the highest counts of instants missing in the considered 60 days. Out of these hours, hour '0' has the highest frequency with the value 8 and represents the occurrence of the highest count of data instants missing.

Table 2. Summary of observations on the count of missing instants

S.No.	Appliance	No. of days connected	Observation on highest missing instants counts		Observation on lowest missing instants counts	
			Date(s) with highest missing instants	Corresponding missing instants count	Date(s) with lowest missing instants	Corresponding missing instants count
1	Alarmclock	5	01/09/11	49826	03/09/11	46608
2	Charger-Smartphone	5	22/01/12	76364	20/01/12	70389
3	Charger-PSP	2	19/11/11	64442	18/11/11	54905
4	CdPlayer	2	21/01/12	78289	20/01/12	70781
5	SolarThermalSystem	8	24/01/12	64553	26/01/12	7560

6	XmasLights	6	06/01/12	79128	08/01/12	73968
7	DvdPlayer	5	20/01/12	59412	31/12/11	55728
8	WaterBoiler	2	24/01/12	64490	25/01/12	61618
9	VacuumCleaner	1	21/01/12	57830	-	-
10	Iron	3	25/12/11	34570	24/12/11	34488
11	BeanToCupCoffeemaker	44	19/08/11	2899	30/08/11	827
12	Breadcutter	13	27/01/12	76612	25/01/12	66214
13	Cookingstove	16	01/01/12	63638	20/12/11	52199
14	DigitalTvReceiver	24	01/01/12	63638	09/01/12	52070
15	EthernetSwitch	33	29/11/11	71657	20/01/12	26802
16	Freezer	9	24/01/12	64565	26/01/12	4130
17	LaundryDryer	9	24/01/12	130995	17/12/11	15863
18	Monitor-CRT	26	30/11/11	54858	19/08/11	2195
19	Multimediacenter	17	20/11/11	68409	26/01/12	31085
20	Playstation3	14	29/12/11	57798	13/02/12	30981
21	Printer	16	30/11/11	64877	08/01/12	52919
22	Projector	8	16/01/12	78873	14/01/12	72117
23	RemoteDesktop	9	30/11/11	62149	27/11/11	56722
24	Router	40	29/11/11	71741	26/01/12	31101
25	Subwoofer	28	26/01/12	89521	05/02/12	31042
26	Toaster	25	06/09/11	73077	20/01/12	26886
27	USBHarddrive	30	16/01/12	112335	09/01/12	52133
28	USBHub	10	16/01/12	56175	26/01/12	31293
29	TV-CRT	36	04/01/12	42852	18/06/12	3654
30	TV-LCD	119	20/01/12	234864	18/06/12	4090
31	VideoProjector	19	16/02/12	57509	01/02/12	30950
32	Washingmachine	56	25/06/12	88890	14/06/12	4913
33	WaterFountain	56	12/10/11	5502	04/09/11	864
34	Coffeemaker	82	27/01/12	127944	31/12/11	27090
35	Dishwasher	76	13/06/12	146443	26/08/11	944
36	Lamp	86	16/01/12	232192	05/01/12	27237
37	MicrowaveOven	60	01/01/12	90825	18/06/12	3557
38	PC-Desktop	151	13/06/12	146357	04/08/11	2251
39	PC-Laptop	67	22/01/12	179830	07/02/12	30981
40	Amplifier	89	26/01/12	89375	02/01/12	27094
41	WaterKettle	134	29/11/11	119492	13/09/11	1378
42	Monitor-TFT	190	11/08/11	189047	03/08/11	4836
43	Refrigerator	206	21/06/12	230645	17/09/11	700

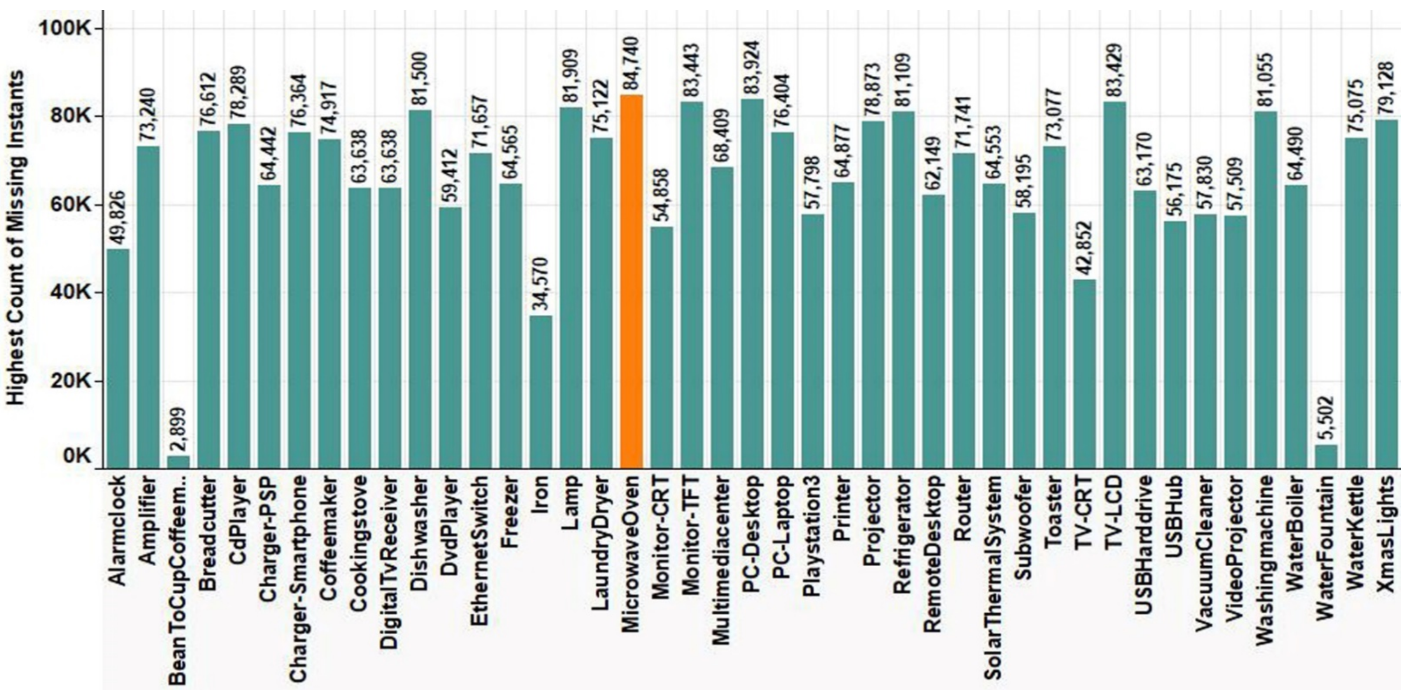


Figure 3. Highest count of missing instants in the readings of different appliances

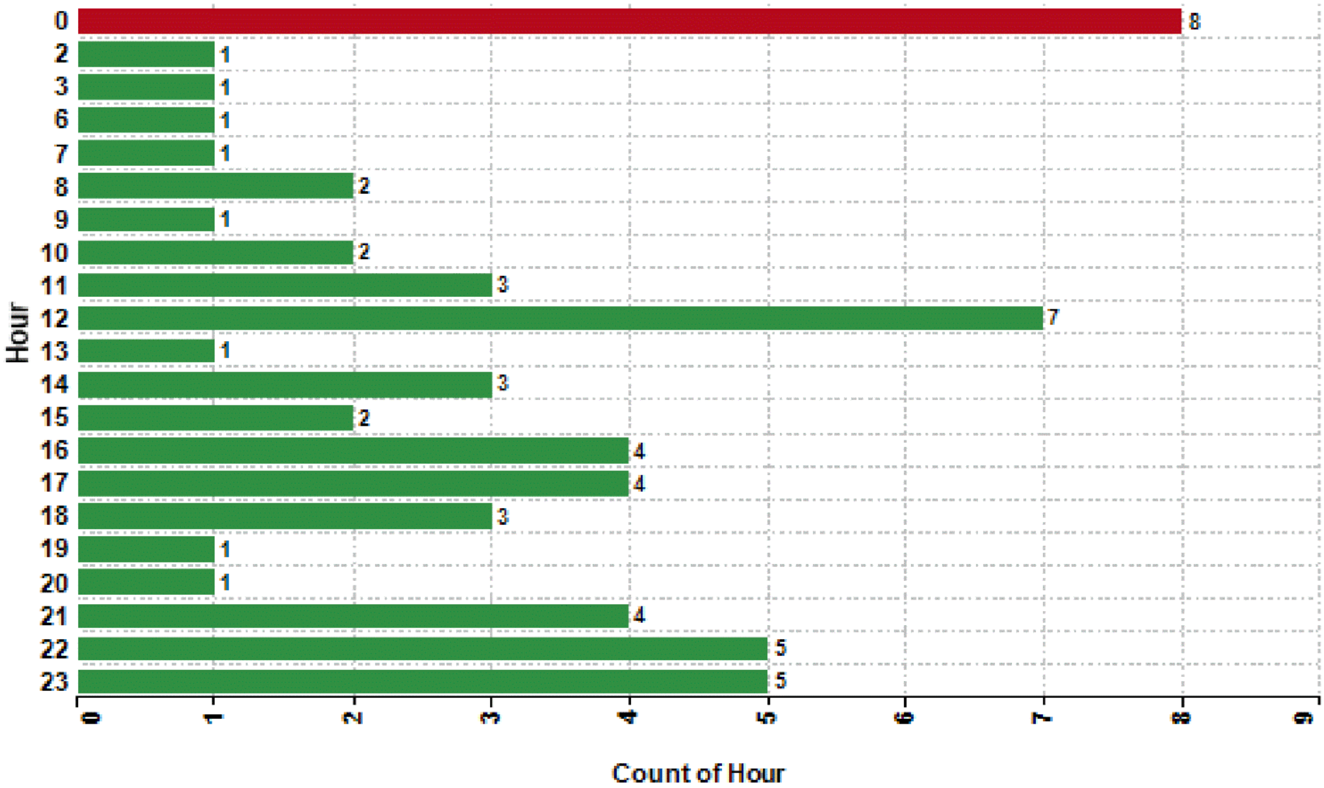


Figure 4. Frequency of hours with highest missing instants in MicrowaveOven

The parts of the day in Darmstadt, Germany are considered Night, Twilight, and Daylight [31]. The hours 00:00 to 06:00 represent Night time. The Astro, Twilight, Nautical Twilight, and Civil Twilight are together considered as Twilight. The hours 07:00, and 20:00 to 23:00 represent Twilight time. The hours 08:00 to 19:00 represent Daylight. All files (60 days) of the appliance "MicrowaveOven" are considered for further analysis as it has the highest count of missing instants. From this analysis, the behavior of missing data in MicrowaveOven appliance during various parts of a day is plotted as shown in Figure 5. From this, it is observed that the highest missing is on Daylight time (33 hours), the next highest is on Twilight time (16 hours), and the lowest is on Nighttime (11 hours).

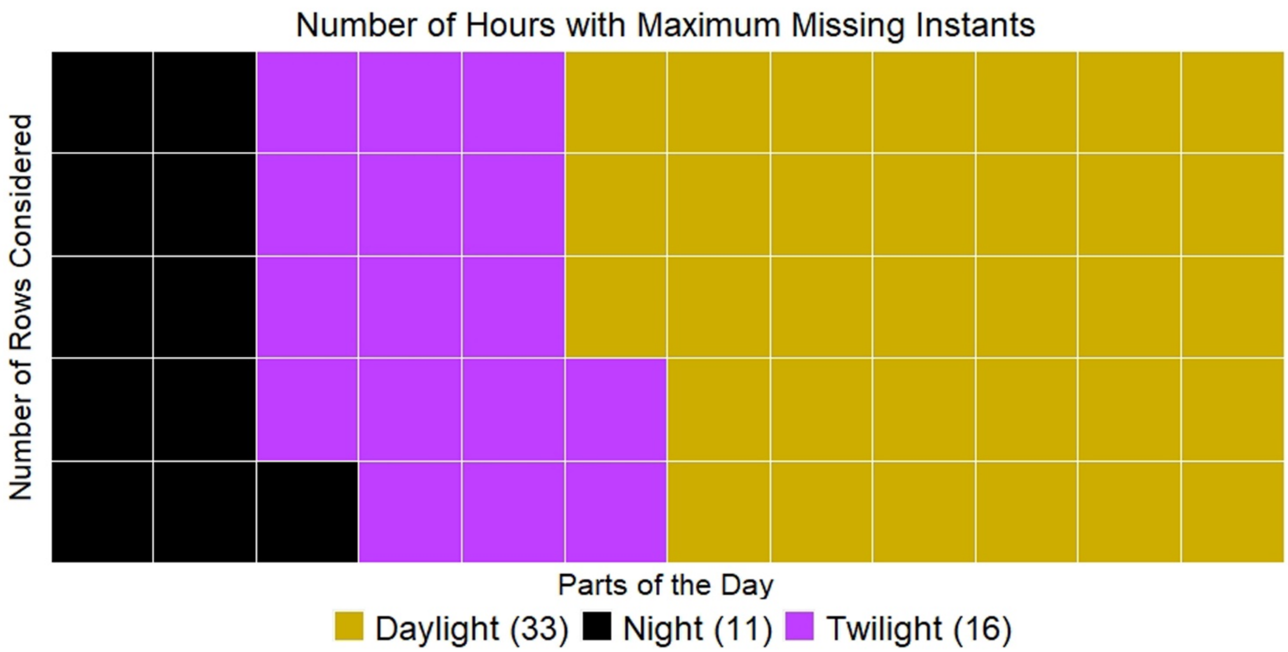


Figure 5. Behavior of missing data in MicrowaveOven appliance during various parts of a day

4. CONCLUSIONS AND FUTURE SCOPE

This paper proposed an analytical approach to exploring the missing data behavior in the smart home energy consumption dataset. The proposed approach successfully explored and quantified the missing data anomalies on all days for all the given appliances in the considered dataset. This analysis revealed that the appliance 'MicrowaveOven' had the highest count (84740) of missing instants. Further, this proposed approach finds the behavior of missing data anomalies by considering the appliance 'MicrowaveOven' as a test case. The conclusions drawn from the implementation of the proposed approach are given as follows:

- In some appliances, the devices have the same count of missing instants on the same day. For e.g., the device 'dev_D33097' of 'CookingStove' appliance and the device 'dev_D330A3' of 'DigitalTvReceiver' appliance consists of the same count (63638) of missing instants on the same day (01/01/2012). This analysis may help the engineers to suspect and identify some common factors that cause the same count of missing data records across various devices/appliances at the same instants of the time.
- Unexpectedly, in some appliances, more than one lakh missing instants are observed. The reason for this is explained below.
  - In general, the expected number of records in a day is 86400 (24 h × 60 m × 60 s) as there is one trace per second is desired. But, due to the redundancy in the energy consumption data records, the total number of records exceeds the ideal expected count (86400). This further increases the count of missing data instants than the actual count. Hence, these redundant records increase the complexity of the missing data analysis and further delude the identification of missing records correctly. So, it is expected that the dataset be free from such redundant records to have an accurate behavioral analysis of missing data.
  - This opens up a new investigation requirement on the redundant data anomalies to further enhance the data quality and purification process.
- The highest count of missing instants is observed during the Daylight period of a day.

Therefore, it is concluded that the proposed comprehensive exploration of missing data anomalies helps the engineers and researchers to understand the presence and the behavior of missing data anomalies that help for accurate analytics.

5. ACKNOWLEDGEMENT

This work was supported in part by Project Grant No: SRG/2019/000648, sponsored by the Start-up Research Grant (SRG) scheme of the Science and Engineering Research Board (SERB), a statutory body under the Department of Science and Technology (DST), Government of India.

REFERENCES

1. Zielonka, A., Woźniak, M., Garg, S., Kaddoum, G., Piran, Md. J. and Muhammad, G., "Smart homes: How much will they support us? A research on recent trends and advances", ***IEEE Access***, Vol. 9, (2021), 26388-26419. (https://doi.org/10.1109/ACCESS.2021.3054575).
2. DeFranco, J.F. and Kassab, M., "Smart home research themes: An analysis and taxonomy", ***Procedia Computer Science***, Vol. 185, (2021), 91-100. (https://doi.org/10.1016/j.procs.2021.05.010).
3. Pira, S., "The social issues of smart home: A review of four European cities' experiences", ***European Journal of Futures Research***, Vol. 9, No. 3, (2021), 1-15. (https://doi.org/10.1186/s40309-021-00173-4).
4. Kim, M.J., Cho, M.E. and Jun, H.J., "Developing design solutions for smart homes through user-centered scenarios", ***Frontiers in Psychology***, Vol. 11, (2020), 1-12. (https://doi.org/10.3389/fpsyg.2020.00335).
5. Benjamin, K.S. and Dylan, D.F.D.R., "Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies", ***Renewable and Sustainable Energy Reviews***, Vol. 120, (2020), 109663. (https://doi.org/10.1016/j.rser.2019.109663).
6. Diahovchenko, I., Kolcun, M., Čonka, Z., Savkiv, V. and Roman, M., "Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads", ***Iranian Journal of Science and Technology, Transactions of Electrical Engineering***, Vol. 44, (2020), 1319-1333. (https://doi.org/10.1007/s40998-020-00322-8).
7. Wonyoung, C., Kim, J., SangEun, L. and Park E., "Smart home and internet of things: A bibliometric study", ***Journal of Cleaner Production***, Vol. 301, (2021), 126908. (https://doi.org/10.1016/j.jclepro.2021.126908).
8. Lin, Y.W., Tseng, S.K., Liao, J.K. and Hsu, T.H., "Developing smart home applications", ***Mobile Networks and Applications***, (2020), 1-15. (https://doi.org/10.1007/s11036-020-01639-8).
9. Almusaylim, Z.A. and Zaman, N., "A review on smart home present state and challenges: Linked to context-awareness internet of things (IoT)", ***Wireless Networks***, Vol. 25, (2019), 3193-3204. (https://doi.org/10.1007/s11276-018-1712-5).
10. Chen, M., Yang, J., Zhu, X., Wang, X., Mengchen, L. and Jeungeun, S., "Smart home 2.0: Innovative smart home system powered by botanical IoT and emotion detection", ***Mobile Networks and Applications***, Vol. 22, (2017), 1159-1169. (https://doi.org/10.1007/s11036-017-0866-1).
11. Rasha, E.A., "Smart homes: Potentials and challenges", ***Clean Energy***, Vol. 5, No. 2, (2021), 302-315. (https://doi.org/10.1093/ce/zkab010).
12. Zhibin, N., Junqi, W., Liu, X., Huang, L. and Nielsen, P.S., "Understanding energy demand behaviors through spatio-temporal smart meter data analysis", ***Energy***, Vol. 226, (2021), 120493. (https://doi.org/10.1016/j.energy.2021.120493).
13. Yamauchi, M., Ohsita, Y., Murata, M., Ueda, K. and Kato, Y., "Anomaly detection in smart home operation from user behaviors and home conditions", ***IEEE Transactions on Consumer Electronics***, Vol. 66, No. 2, (2020), 183-192. (https://doi.org/10.1109/TCE.2020.2981636).
14. Darby, S.J., "Smart technology in the home: Time for more clarity", ***Building Research & Information***, Vol. 46, No. 1, (2018), 140-147. (https://doi.org/10.1080/09613218.2017.1301707).
15. Barsocchi, P., Calabrò, A., Ferro, E., Gennaro, C., Marchetti, E. and Vairo, C., "Boosting a low-cost smart home environment with usage and access control rules", ***Sensors***, Vol. 18, No. 6, (2018), 1-22. (https://doi.org/10.3390/s18061886).
16. Albuquerque, P.U.B., Ohi, D.K.d.A., Pereira, N.S., Prata, B.d.A. and Barroso, C.G., "Proposed architecture for energy efficiency and comfort optimization in smart homes", ***Journal of Control, Automation and Electrical Systems***, Vol. 29, (2018), 718-730. (https://doi.org/10.1007/s40313-018-0410-y).
17. Fan, X., Qiu, B., Liu, Y., Zhu, H. and Han, B., "Energy visualization for smart home", ***Energy Procedia***, Vol. 105, (2017), 2545-2548. (https://doi.org/10.1016/j.egypro.2017.03.732).
18. Martinez-Pabon, M., Eveleigh, T. and Tanju, B., "Smart meter data analytics for optimal customer selection in demand response programs", ***Energy Procedia***, Vol. 107, (2017), 49-59. (https://doi.org/10.1016/j.egypro.2016.12.128).
19. Hare, J., Shi, X., Gupta, S. and Bazzi, A., "Fault diagnostics in smart micro-grids: A survey", ***Renewable and Sustainable Energy Reviews***, Vol. 60, (2016), 1114-1124. (http://dx.doi.org/10.1016/j.rser.2016.01.122).
20. Kezunovic, M., Pinson, P., Obradovic, Z., Grijalva, S., Hong, T. and Bessa, R., "Big data analytics for future electricity grids", ***Electric Power Systems Research***, Vol. 189, (2020), 106788. (https://doi.org/10.1016/j.epsr.2020.106788).
21. vom Scheidt, F., Medinová, H., Ludwig, N., Richter, B., Staudt, P. and Weinhardt, C., "Data analytics in the electricity sector –A quantitative and qualitative literature review", ***Energy and AI***, Vol. 1, (2020), 100009. (https://doi.org/10.1016/j.egyai.2020.100009).
22. Wang, Y., Chen, Q., Hong, T. and Kang, C., "Review of smart meter data analytics: Applications, methodologies, and challenges", ***IEEE Transactions on Smart Grid***, Vol. 10, No. 3, (2019), 3125-3148. (https://doi.org/10.1109/TSG.2018.2818167).
23. Chou, J.S. and Ngo, N.T., "Smart grid data analytics framework for increasing energy savings in residential buildings", ***Automation in Construction***, Vol. 72, No. 3, (2016), 247-257. (http://dx.doi.org/10.1016/j.autcon.2016.01.002).

24. Purna Prakash, K. and Pavan Kumar, Y.V., "Simple and effective descriptive analysis of missing data anomalies in smart home energy consumption readings", ***Journal of Energy Systems***, Vol. 5, No. 3, (2021), 199-220. (https://doi.org/10.30521/jes.878318).

25. Gilani Fahad, L. and Tahir, S.F., "Activity recognition and anomaly detection in smart homes", ***Neurocomputing***, Vol. 423, (2021), 362-372. (https://doi.org/10.1016/j.neucom.2020.10.102).

26. Ariyaluran Habeeb, RA., Nasaruddin, F., Gani, A., Abaker Targio Hashem, I., Ahmed, E. and Imran, M., "Real-time big data processing for anomaly detection: A Survey", ***International Journal of Information Management***, Vol. 45, (2019), 289-307. (https://doi.org/10.1016/j.ijinfomgt.2018.08.006).

27. Moghaddass, R. and Wang, J., "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data", ***IEEE Transactions on Smart Grid***, Vol. 9, No. 6, (2018), 5820-5830. (https://doi.org/10.1109/TSG.2017.2697440).

28. Hela, S., Amel, B. and Badran, R., "Early anomaly detection in smart home: A causal association rule-based approach", ***Artificial Intelligence in Medicine***, Vol. 91, (2018), 57-71. (https://doi.org/10.1016/j.artmed.2018.06.001).

29. Wen, C., Zhou, K., Yang, S. and Cheng, W., "Data quality of electricity consumption data in a smart grid environment", ***Renewable and Sustainable Energy Reviews***, Vol. 75, (2017), 98-105. (http://dx.doi.org/10.1016/j.rser.2016.10.054).

30. Andreas, R., "The tracebase appliance-level power consumption dataset", (2020). (https://github.com/areinhardt/tracebase/tree/master/complete), (Accessed: 28 February 2022).

31. Time and Date AS | Darmstadt, HE, Germany — Sunrise, Sunset, and Moon Times for Today, Parts of a Day in Darmstadt, Germany, (2022). (https://www.timeanddate.com/astronomy/germany/darmstadt). (Accessed: 28 February 2022).