RESEARCH ARTICLE OPEN ACCESS

Data Mining for Consumption Pattern Recognition of Electrical Utility in Nepal

Laxaman Mandal*[1], Dr. Rajesh Keshavrao Deshmukh*[2], Dr. Gajendra Sharma*[3]

*[1]Computer Science & Engineering Department, Kalinga University, Naya Raipur, India

*[2] Department of Computer Science & Engineering, Kalinga University, India

*[3] Department of Computer Science & Engineering, Kathmandu University, Nepal

(Phd Scholar, Kalinga University, Naya Raipur, India)

*Correspondent Author:blaxman2005@gmail.com

Abstract:

Electrical Utility in Nepal is having the problems in demand side management. For analyzing the issues, Utility has to recognize the consumption pattern of consumers as monthly, category, reading area and season wise consumptions. Monthly consumptions say that which month is having maximum and minimum demand of electricity. Similarly, which consumers' category require maximum and minimum electricity, same analysis is done for reading areas. Seasonal analysis of consumptions behavior of consumers may find the real demand of electricity at seasonal basis. NEA still has the electrical crisis in dry seasons. To fulfill the demand of electricity in dry season, NEA is importing the electrical power from India. Similarly,one can analyze the actual demand of consumer by reading area, category, month and season. Data mining is done by using ICT tools like Python and MySQL.

In thisarticle, consumption patterns of consumers are to be recognized for enhancing and upgradation of existing supply chain of electrical network which helps in proper demand side management in monthly, reading areas, consumers' category as well as seasonal basis. This helps on determination of actual demand of electrical energy so that NEA can manage the electrical crisis optimally.

Keywords: Electrical Utility, NEA, EE, Consumption Pattern, Dry and Wet Season, Electrical Crisis, ICT tool, Distribution Network.

INTRODUCTION

Electrical Utility of Nepal performs its work in distribution sector by manual approach for meter reading, maintaining the record of received and sales energy in sub-station in hand written log book which contains the information like incoming energy and outgoing energy as well as tripping occurrence of feeder at daily basis. Electrical Utility here in Nepal does the physical patrolling to identify the loss prone areas and performs loss reduction control activities in those areas physically.

Consumption pattern recognition is the important factor for demand side management. In Nepal, different categories of consumers are there. Some of the categories are domestic, non-domestic, commercial, non-commercial, industrial, irrigation, street light, temporary, water supply, transport, entertainment, drinking water, religious and culture, charging station etc. Among these categories, 92.32% of total consumers are of domestic [1]. Among domestic category, more consumers are of single phases and less of them are three phases. For analysis of category wise consumption, the sales data of each category should be analyzed monthly or seasonally. Now here are the requirements of analyzing the consumption behavior of customers monthly, categorically, reading area wise as well as seasonally. As in Nepal, electricity crises are even in dry season includes the months of January, February, March, April, May & June. Mostly, NEA is importing the electricity from India in dry season. But there is spill energy in wet seasons includes the month of July, August, September, October, November and December.

Billions of rupees are invested by NEA for manual support and maintenance in all distribution sectors every year. No such ICT mechanism is utilized to identify the technical fault in network, meter reading approach is still manual means door to door movement while doing meter reading for sales data collection. NEA can save that much time and money with less investment and less time if there will be use of ICT tools such as AMR, MDM and GIS mapping of field asset of distribution network sector.

For better knowledge of demand side management, it is necessary to mine the sales data at consumer end. This data can be acquired by billing system of electrical utility of Nepal.

I. METHODOLOGY

This section is for data mining of sales data of electrical utility in Nepal. This analyzes the consumption behaviors of consumers as per month, category, reading areas and seasons.Data mining of consumption pattern is done by using ICT tools like Python and MySQL.

There is the consideration of two aspects of power consumption. One is for maximumdemand and other one for minimum demand. In order to manage the demand of electricity, one has to reconfigure the distribution network that consists of substation, feeder as well as replacement of inappropriate size of conductorsused in distribution lines.
Consumption pattern recognition is essential for electricity demand forecasting that plays a significant role in distribution system power planning, management of power system. It is important that accurate electricity forecasting has major importance to the household sector, a major contributor to the peak loads in most electricity systems. Overestimating electricity demands misleads planners and wastes resources with expensive expansion plans that results in increasing of operating costs, since electricity cannot be stored on a large scale unlike other energy sources [2]. Thus, forecasting of short-term electricity demand in the residential sector is a complex problem because its rise and fluctuation is caused by the difference in demand from month to month. Except these, the consumption behavior is also influenced by many nonlinear variables, such as weather conditions, economics, and demographics [2].
In past few years, electrical energy analysis is emerging and lots of research on electricity consumption analysis like consumer segmentation, classification, forecasts and knowledge extraction from smart meter had been done.The data mining techniques mostly used classification,clustering of electricity demand patterns and cluster analysis of smart metering data [3].
The following approaches are involved in data mining of consumption pattern of electricity.

a) Data collection
This approach is used to collect data from computerized billing system of electrical utility in Nepal. The data is collected in excel format.

b) Data conversion
This approach uses the ICT tools as MySQL for importing excel data into MySQL database.

c) Create database, table and views under MySQL platform
d) Data sampling
This approach is filtering the data as per requirements means that original table contains multiple columns of data. Among which some the columns which are required ones only be selected. This is done by creating the MySQL views.
Syntax for creating views as:
Create view *viewname* as select column1, column2, …… columnN from *table_name*;
Syntax for creating stored procedure
Create or replace procedure procedure_name(in arg1 datatype,in arg2 datatype,……)
Begin
……..
……..
End;
For performing the task, stored procedure is called.
Syntax for calling stored procedure:
Call (ag1, arg2, …);

e) For data analysis programming language is used as Python
This is the programming language using which one can design the form. Now user can enter the sales data in this form and pattern of sales as monthly, categorically, seasonally and reading area wise are generated.
For this work, a sample code of Python is for graphical representation of result as here under.

```
import mysql.connector #importing library for making database connection
import tkinter as tk #importing library for creating forms
from tkinter import ttk
import matplotlib.pyplot as plt #importing library for plotting the graph
def fetch_data_from_sp(result1):
   try:   #for making database connection
      connection = mysql.connector.connect(
         host='localhost',
         user='user name',
         password='p@ssw0rd',
         database='database name'
      )
```



```
        cursor = connection.cursor()

# Assuming you have a stored procedure named 'sp_graph_plot' that takes an input parameter
        cursor.callproc('sp_graph_plot', [result1])
        print("Calling stored procedure with:", result1)
        for result in cursor.stored_results():
   resultaa = result.fetchall()
        cursor.close()
        connection.close()
return resultaa
except mysql.connector.Error as err:
        print("------Error:", err)
        return []
def submit_button_clicked():
    #selected_option = combo_box.get()
    result1 = combo_box.get()
    result = fetch_data_from_sp(result1)

    # Extract x and y data from the result
    x_data = [row[0] for row in result]
    y_data = [row[1] for row in result]
    print(x_data)
    print(y_data)

    # Plotting the graph
    plt.figure(figsize=(8, 6))
    plt.plot(x_data, y_data, marker='o')
    plt.xlabel('X')
    plt.ylabel('Y')
    plt.title('Monthly Consumption Analysis')
    plt.grid(True)
    plt.show()

# Create a simple GUI using tkinter
root = tk.Tk()
root.title("Stored Procedure Data Fetching and Plotting")

# Create a combo box with options
combo_box = ttk.Combobox(root, values=["Year","1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12"])
combo_box.set("Select an option")
combo_box.pack(padx=10, pady=10)

# Create a submit button
submit_button = tk.Button(root, text="Submit", command=submit_button_clicked)
submit_button.pack(padx=10, pady=10)

root.mainloop()
```

f) Represent these data as sales of pattern graphically for further analysis.

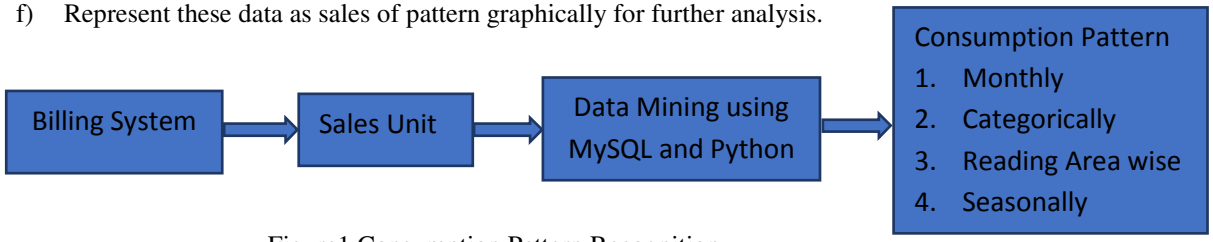

Figure1.Consumption Pattern Recognition

From this figure, it can be understood that sales unit is the data which is collected from computerized billing system running under NEA. Hence data collected is in the form of excel. This excel form of data is imported in MySQL.After importing, these sales data are kept under database having a table. Now data filtering process is performed by creating the views under MySQL database. The stored procedure is created for clustering the data in several form such as monthly sales, category wise sales, reading area wise sales and seasonal sales as consumption pattern. These pattern of sales data is presented graphically using python programming. This program helps to present such pattern of consumption graphically. Thus one can trace out the focal month, reading areas, consumer's category and season wise maximum and minimum sales.

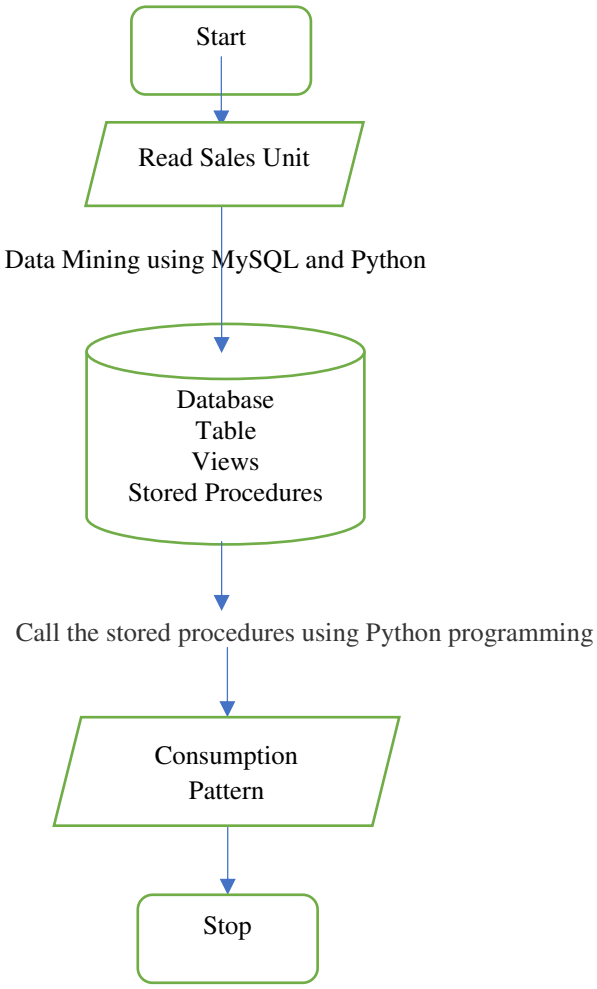

Figure2. Flowchart of Consumption Pattern Recognition

II. MODELING AND ANALYSIS

Modeling

This is the process of creating a conceptual picture of data and its relationships in order to simplify data management, analysis, and decision-making. It is a basic step in the database design and development process and is used in various domains, including business intelligence, data warehousing, and software development. In this paper, time series, clustering and seasonal decomposition model is adopted.

Data Sampling

For sampling the data, it is important to know about the pocket area, some portion of data among large data set, data size and statistical technique that will be suitable for drawing the valid conclusions. Here billing data is collected. Pocket area is considered here as ratnapark area of Kathmandu. Total data size is considered here as 5.2 lakhs of one year consisting of 14 different consumer's category of different reading areas with two different seasons like dry season as well as wet season. Among 12 months in a year, six months such as January to June is considered as dry season where July to December is considered as wet season by electrical utility of Nepal with respect to revenue billing.

Stratified Sampling: This type of sampling method is used here because the collected sample data set is divided into various sub groups such as monthly sales, category wise sales, reading area wise sales and seasonal sales. So that one can find minimum and maximum demand of electricity by month, category, reading area as well as season.

Thus, data filtering approach is applied for determining the consumption pattern of electricity sales at the consumer end. Sales data is characterized by monthly, seasonally, categorically and reading area wise. For this various views and stored procedure are created. These sales of data are filtered and represented graphically using python programing language.

Statistical analysis performed over very high temporal resolution data allows the description of the electricity consumption profiles. This permits the identification of significant differences and similarities within cluster groups that could be useful for market segmentation and tariff design for utilities and to improved knowledge on groups of consumers to feed specific energy reduction recommendations [4].

Data Analysis

a) Monthly Sales Analysis

In this analysis, sales data of every consumer is showed in monthly basis. So that one can analyze that which month there is maximum and minimum demand of electricity? Whether supplied energy is sufficient in the particular month or not. Following is the sample data set for general idea.

Monthly Sales Data			
Month ID	Month Name	Sales Unit	Remarks
1	April	1500	Minimum Sales= Least Demand
2	May	1600	
3	June	2000	
4	July	2100	
5	August	2250	
6	September	2050	
7	October	2100	
8	November	1950	
9	December	2150	
10	January	2350	
11	February	2400	Maximum Sales= Peak Demand
12	March	2200	

Table1: Monthly Sales Unit

Meter Reading Approach: - Meter reading process is very sensitive as it provides the sales unit which is further used to designate consumption pattern. If reading is accurate then only the sales will be correct otherwise there will be compromise in recognition of consumption pattern which results in wrong decision whether distribution network needs to be upgraded or not. So that meter reading process should be automated in such a way that there should not be intervention of manual approach to read consumption unit of meter. NEA is currently doing the meter reading manually as well as using hand held meter reading device. Almost consumers of NEA is having conventional meter.

In NEA, meter reader reads the consumer meter as per route assigned for every user on monthly basis. For example, consider the following table for different consumers with different consumer numbers. Some sample data set is hereunder.

S.No.	Consumer Name	Consumer Number	Route No. as Reading Date	MR Holidays in a month are
1	Rajesh	001.01.001	01	07,14,21,28
2	Mahesh	001.02.001	02	
3	Ramu	001.03.001	03	
4	Ritesh	001.04.001	04	
5	Mina	001.05.001	05	
6	Maya	001.06.001	06	
7	Gurmeet	001.08.001	08	

Table2: Consumer number details

Here, format of Consumer Number is Area.Rout.Consumer Number. First 3 digit represents Reading Areas, Second 2 digit represents Reading Date called as route and last 3 digit represents consumer serial numbers, this may also be of alphanumeric such as 001BA, 001BA1 etc. Meter Reader reads the consumer meters as per consumer numbers mentioned in the table. Maximum number of Meter Reading Areas may be of 999 in a Distribution Center, Maximum Routes may be of 28 in a month and Maximum number of consumers may be of 999 per route in a Distribution Center.

The consumer as Mahesh whose meter reading is done at 8^{th} of every month as mentioned in table above, but we consider the monthly sales for calculation of energy loss against received unit from feeder meter.

General practice of meter reading in Nepal is manual as well as using hand held device. Sample data set is hereunder as for general idea.

S.No.	Meter Reading Approach	Sales Accuracy
1	Manual reading	Less
2	Using hand held device	Medium
3	Using ICT tools	High

Table3: Meter reading approach

Solutions for acquiring the actual consumption unit from consumer meter are usage of smart meter with AMI (Advance Metering Infrastructure) & MDM (Meter Data Management) features facilitated with IoT (Internet of Things), replacement of faulty meters, smart meter for reading consumer meters etc. [5].

MDM manages and stores the reading data to be transmitted as per user request. It also validates the reading data as well as outage events if any.

b) Consumer's category wise sales analysis

In this analysis, sales unit is analyzed by each category so that one can recognize consumption pattern by category such that one among the various categories is having maximum demand where as other one is having least demand among all categories. In this paper, the categories considered as per Nepal scenario as domestic, non-domestic, commercial, non-commercial, industrial, entertainment, religious and cultural, charging station, transport, water supply etc.

c) Reading area wise sales analysis

In this analysis, sales data is analyzed as per reading area so that one can find which area is high demand and which one is least demand. On the basis of these data, it will be easier to identify the feeder which is having high demand and which one is under low demand. Thus, management can find the overloaded as well as under loaded feeder. This result suggests the management for upgrading the feeder which is overloaded provided that energy is sufficient. Load diversion of such feeder having under loaded can be done for the high demanding areas of electricity.

d) Seasonal sales analysis

This analysis is done for tracing the demand of electricity at seasonal basis. As NEA is suffering from the energy crises in dry season and energy spill in wet season this kind of demand of energy sales analysis becomes very important. This approach of consumption pattern analysis helps the management such that how much energy is required at which season. Proper demand side management could be made possible by this approach of data mining. In Nepal, electrical utility designates January, February, March, April, May, June as dry season and July, August, September, October, November, and December as wet season.

III. RESULTS AND DISCUSSION

Results

The following sample data isof tabular form with respect to sales data are as follows:

Month ID	Month Name	Sales Unit
1	April	18687650
2	May	19662025
3	June	21388988
4	July	22674850
5	August	22221146
6	September	21507380
7	October	18799986
8	November	18131868
9	December	20207353
10	January	23489067
11	February	19907853
12	March	19004746

Table4: Monthly Sales Unit

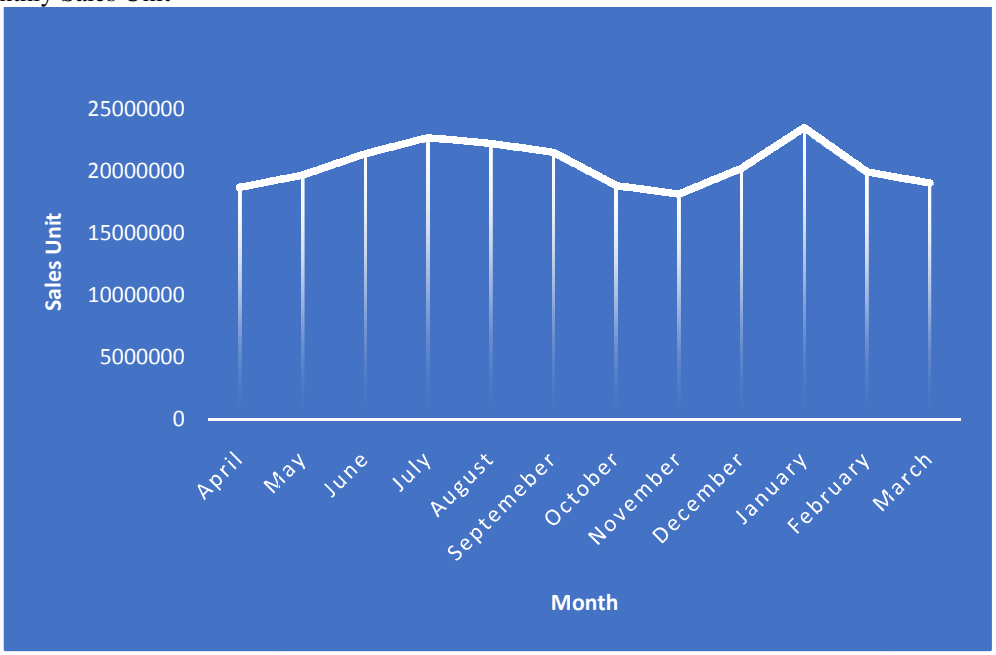

Chart1: Monthly Sales Graph

In the above table4, there are two months such as January belongs to dry season having maximum and November belongs to wet season having minimum demand of electricity. The data shown in table4 are of 2022/2023 Fiscal Year. These data are of

sales unit of the particular month mentioned over there in table. This chart1shows the demand of electricity for both wet and dry season also.These sample data belong to the sales unit of Ratnapark region of Kathmandu. This region is situated at the heart of Kathmandu valley having monthly sales throughout the year.

Category	Sales Unit
Transport	53175
Drinking Water	92142
Religious Culture	153161
WaterSupply	233018
ChargingStation	261800
TempSupply	413796
Entertainment	877095
IntrnlConsumption	1040964
Industrial	5185984
StreetLight	6141006
NonCommercial	33320007
Commercial	62946855
Domestic	65656978
NonDomestic	69306931

Table5: Category wise Sales Unit

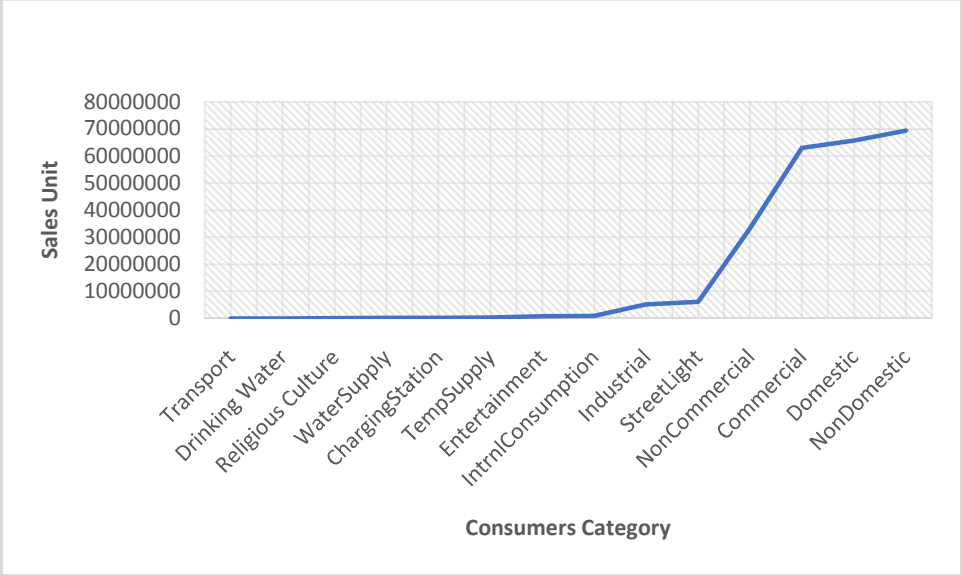

Chart2: Category wise Sales Graph

The table5 represents the consumer's category wise sales unit of whole year. According to the chart2, Non-Domestic category's consumer is having maximum demand of electricity as compared to others type of consumer's categories. Transport like consumer's category is having least demand of electricity as per data mentioned in the table5.

Reading Area	Sales Unit
002	5585333
003	11628553
004	9169492
005	2534560
006	4111274
009	2972447
011	9577323
013	1907941
016	15937588
022	160393997

033	4688191
037	17176213

Table6: Reading Area wise Sales Unit

Reading Area	Category	Sales Unit
002	NonDomestic	348847
003	NonDomestic	592932
004	NonDomestic	958906
005	NonDomestic	84744
006	NonDomestic	374958
009	NonDomestic	143631
011	NonDomestic	432284
013	NonDomestic	0
016	NonDomestic	307535
022	NonDomestic	65762357
033	NonDomestic	49512
037	NonDomestic	251225

Table7: Reading Area-Category wise Sales Unit

Chart3: Reading Area Wise Sales Graph

The table6 represents reading area wise sales. According to the chart3, area number 22 is having maximum demand of electricity as compared to other areas. This is also verified by the sales unit shown in table7 having reading area as 22 with Non-Domestic consumer having maximum demand of electricity.

Season	Sales Unit
Dry Season	120958694
Wet Season	124724218

Table8: Seasonal Sales Unit

The table7 represents reading-area,category,season wise sales. According to the chart4, wet season is having maximum demand of electricity as compared to other areas. In ratnapark area of Kathmandu valley, Non_domestic category consumers are maximum demand of electricity as compared to other categories of consumers. While analyzing the consumptions data

regarding seasons, wet season is having maximum demand of electricity as compared to dry season because wet season including the months like Jun, July, August, September and December is having larger amount of consumptions than that of five months except January of dry season according to the resulted data of table4. Thus, wet season is having maximum demand of electricity as compared to dry season.

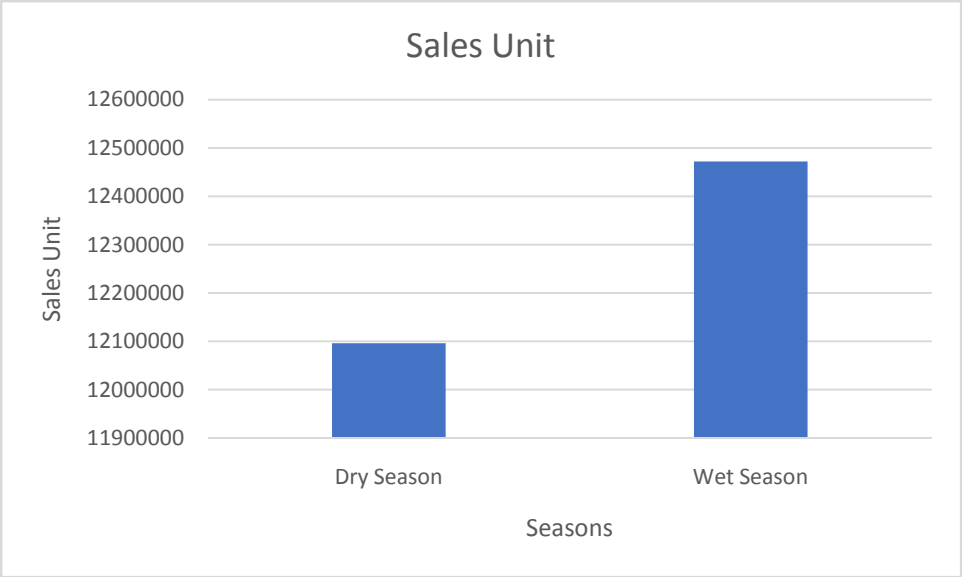

Chart4: Seasonal Sales Graph

Discussion

The above graphical representations of various data sets of billing data show that maximum and minimum demand of electricity by month, consumer'scategory, reading areas as well as seasons. This visual approach of data analysis helps management of electrical utility for proper management of electricity demand by month, category, reading area and seasons. So that management can optimize the supply of electricity as per demand of electricity in the particular domain. But here in Nepal, meter reading approach is manual as meter reader is going door to door for meter reading. As manual approach of meter reading produces inaccuratesales as compared to smart meter reading approach.

IV. CONCLUSION

With the study of various tools and technique, it is concluded that one has to do the actual reading of energy meter regularly by using smart meter featured with ICT tools like AMI(Advanced Metering Infrastructure) and MDM(Meter Data Management). This tool can store and retrieved the meter reading data as and when required by the management. If there is actual meter reading, then only actual sales can be obtained and analyzed the real sales unit of the particular distribution center. The result shown above proves that month as January having maximum demand of electricity, category as Non-Domestic having maximum demand of electricity, reading area as 022 having maximum demand of electricity and wet season having maximum demand of electricity as compared to dry season. For better analysis of consumptions, one has to do the followings.

- use of smart meter.
- GIS Mapping of the entire distribution network that consists of substation, feeder, transformer, poles, conductor, consumer meter

Without using ICT tools like smart meter and GIS Application in distribution network of NEA it will be difficult toanalyze the demand of electricity reading area wise, category wise, month wise and season wise as well as upgradation and maintenance of distribution network.

V. REFERENCES

[1] Org.np. [Online]. Available: https://nea.org.np/admin/assets/uploads/annual_publications/DCSD_Final_Book_2080.pdf. [Accessed: 19-Aug-2023].
[2] H. Son and C. Kim, "Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization," Procedia Eng., vol. 118, pp. 1162–1168, 2015.
[3] U. Ali, C. Buccella, and C. Cecati, "Households electricity consumption analysis with data mining techniques," in IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016.
[4] J. P. Gouveia, J. Seixas, S. Luo, N. Bilo, and A. Valentim, "Understanding electricity consumption patterns in households through data fusion of smart meters and door-to-door surveys," Eceee.org.,2015.
[5]J. S. Mc Menamin, "Closing the books: Forecasting unbilled energy and sarbanes-Oxley compliance," Itron.com. [Online]. Available: https://www.itron.com/-/media/feature/products/documents/white-paper/closing-the-books--forecasting-the-unbilled-energy-and-sarbanesoxley-compliance.pdf. [Accessed: 08-Jan-2023].

