UPES Centre for Continuing Education

Strategic Technology Management to reduce Non-Technical Losses in Power Distribution (With specific reference to Tenaga Nasional Berhad, Kuala Lumpur, Malaysia)

By

Sudhansu Sekhar Pati

SAP ID – 500065844

Guided By

Devesh Joshi

HOD, Solution Delivery

Itron Asia Pacific

A DISSERTATION REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

MBA in Power Management

OF

CENTRE FOR CONTINUING EDUCATION

UNIVERSITY OF PETROLEUM & ENERGY STUDIES, DEHRADUN

Acknowledgement

This is to acknowledge with thanks the help, guidance and support that I have received during the Dissertation.
I have no words to express a deep sense of gratitude to the management of Itron Asia Pacific for giving me an opportunity to pursue my Dissertation, and in particular **Mr. Devesh Joshi** for his able guidance and support.

I must also thank **Mr. Kanith Boonthrapong** from Itron Thailand and **Mr. Chin-Han Ang**, from Itron Malaysia for their valuable support. My acknowledgment would be incomplete without my sincere thanks to Dr. Tiong for his constant guidance. I am also thankful to Ms. Suhana for teachings on a systematic process of research.

I also place on record my appreciation of the support provided by **Mr. Xiung Tepparit** and other staff of Bang Sa public Library of Thailand.

Further, I am grateful to all the respondents of the questionnaire, including staff of State Electricity Boards, Utility personnel's to answer the questionnaire and having interviews, by taking out time from their busy schedules.

My wife Sangeeta, my parents' mother Damayanti Pati and my father Mr. Ganesh Pati and my sisters have always been there for me as a rock support. Last but not the least; I want to thank my little daughter GIFT who has been a source of inspiration for me to do this research.

Sudhansu Sekhar Pati
SAP ID – 500065844
H-1107, Prateek Laurel
Sector 120, Noida
Cell: +91 858 689 7300
Email: PatiSudhansu@gmail.com

Declaration by the Guide

This is to certify that the Mr Sudhansu Sekhar Pati a student of MBA in Power Management, SAP ID 500065844 of UPES has successfully completed this dissertation report on "Strategic Technology Management to reduce Non-Technical Losses in Power Distribution (With specific reference to Tenaga Nasional Berhad, Kuala Lumpur, Malaysia)" under my supervision.

Further, I certify that the work is based on the investigation made, data collected and analyzed by him and it has not been submitted in any other University or Institution for award of any degree. In my opinion it is fully adequate, in scope and utility, as a dissertation towards partial fulfillment for the award of degree of MBA.

Devesh Joshi
Senior Delivery Manager
Itron India Pvt Ltd
C-7, Sector-3, Noida
UP-201301
Telephone: +91 120 4656666
Mobile: +91 7838281899
e-mail: Devesh.Joshi@itron.com

Date: Oct 18 2019
Place: Noida

Contents

Table of figures

Executive Summary

Executive Summary

In electricity supply to final consumers, losses refer to the amounts of electricity injected into the transmission and distribution grids that are not paid for by users. Total losses have two components: technical and non-technical. Technical losses occur naturally and consist mainly of power dissipation in electricity system components such as transmission and distribution lines, transformers and measurement systems. Non-technical losses are caused by actions external the power system and consist primarily of electricity theft, non-payment by customers and errors in accounting and record-keeping. These three categories of losses are respectively sometimes referred to as commercial, non-payment and administrative losses.

Non-technical losses in the power sector are almost non-existent or negligibly small in developed countries. In contrast, although mixed, the situation tends to be significantly different in developing countries. Many electricity utilities in developing countries succeeded in significantly reducing or eliminating non-technical losses in electricity supply on a sustainable manner, but others continue to show high losses.

The main effort of this thesis is to study the adoption and implementation of different strategic model management by different utilities to increase the effectiveness rate of the energy recovery process. The way to achieve this objective is by characterizing the variables that allow to model the behavior of a user of the electricity sector, so that according to the model it is possible to classify fraudulent and non-fraudulent users. The thesis focusses specifically on methods used by **Tenaga Nasional Berhad** to utilize this variable.

In all successful cases, a large share of non-technical losses was concentrated in users able to pay for cost-reflective tariffs. Thus, non-technical losses can be reduced with little loss of welfare, while their continuation jeopardizes the financial sustainability of the power sector and harms well-behaving-electricity consumers, taxpayers, socially disadvantaged segments, and the country. Elimination of those losses (with the exception unmetered consumption explicitly and transparently defined in the regulatory framework) should be a matter of high national priority for every country

Chapter 1

1. Introduction

1.1. Overview

The objectives of the research are the finding of Technology Adoption Index and Technology Usefulness Index. There is need to study the modern and innovative technologies which are being used in power sector in the comprehensive and integrated way. Each technology plays a pivotal role in determining Discoms performance and efficiency. Technology Adoption Index is for Utilities and it represents the intention and effort applied to adopt a new technology to reduce NTL. Whereas Technology Usefulness Index is for technologies and it represents whether the technology has any contribution in reducing NTL. The Index can be used to compare the level of technology implementation among the Discoms.

1.2. Background

Most of the power distribution companies in developing countries is suffering from manual process, poor customer satisfaction, and high revenue leakage. Many Discoms have developed and deployed many IT applications, but most of them are in form of islands and holistic integration approach of technology management and implementation. On the other hand, many Discoms implemented various modern and innovative technologies but still, they are at a very nascent stage in adoption and diffusion of technologies.

Three aspects of technology in power distribution sector

1. Introduction

1.1. Overview

The objectives of the research are the finding of Technology Adoption Index and Technology Usefulness Index. There is need to study the modern and innovative technologies which are being used in power sector in the comprehensive and integrated way. Each technology plays a pivotal role in determining Discoms performance and efficiency. Technology Adoption Index is for Utilities and it represents the intention and effort applied to adopt a new technology to reduce NTL. Whereas Technology Usefulness Index is for technologies and it represents whether the technology has any contribution in reducing NTL. The Index can be used to compare the level of technology implementation among the Discoms.

1.2. Background

Most of the power distribution companies in developing countries is suffering from manual process, poor customer satisfaction, and high revenue leakage. Many Discoms have developed and deployed many IT applications, but most of them are in form of islands and holistic integration approach of technology management and implementation. On the other hand, many Discoms implemented various modern and innovative technologies but still, they are at a very nascent stage in adoption and diffusion of technologies.

Three aspects of technology in power distribution sector

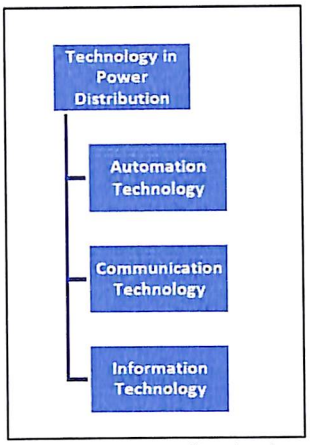


Figure 1 : Aspects of Technology

In most of the Discoms in developing countries, there has been limited interaction among these technologies mentioned in the fig above. Web technologies (IT) can be used to great extent as direct means to communicate with the customers. There is a serious need for all Discoms to focus on adoption of IT which can be a game changer in the transformation of their power utility business.

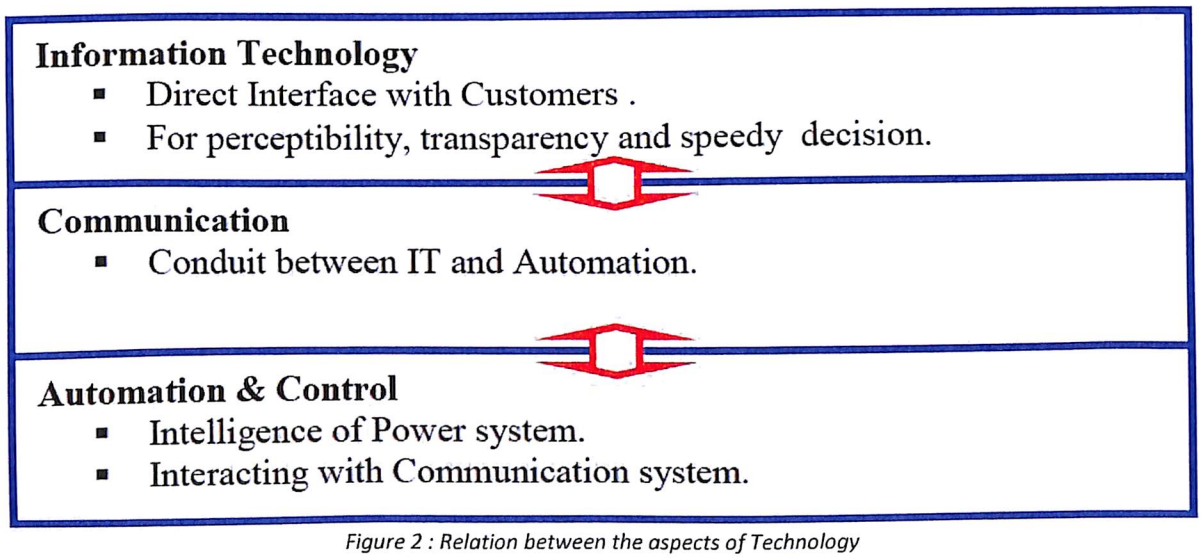


Figure 2 : Relation between the aspects of Technology

1.3. Purpose of the Study

Optimization of technical losses in electricity transmission and distribution grids is an engineering issue, involving classic tools of power systems planning and modeling. The driving criterion is minimization of the net present value (sum of costs over the economic life of the system discounted at a representative rate of return for the business) of the total investment cost of the transmission and distribution system plus the total cost of technical losses. Technical losses are valued at generation costs. Technical losses represent an economic loss for the country, and its optimization should beperformed from a country's perspective, regardless of the institutional organization of the sector and ownership of operating electricity utilities. Although each case has its specific characteristics,depending on the current and future values of generation costs, some general comments can be made. Energy experts agree that, in the next two decades, global prices of primary energyesources (oil and other fossil fuels) will be rising in real terms. In its *World Energy Outlook 2008*, the International Energy Agency forecasts world oil prices rebounding to about US$130(2007 U.S. dollars) per barrel in 2030. Other forecasts differ in absolute values, but not in the upward tendency of energy prices.

On the investment side, prices of equipment in the electricity sector (generation, transmission and distribution) steadily rose this decade until the global financial crisis that began in the 3rd quarter of 2008. Against these price trends, the total costs of technical losses tend to exceed investment costs of transmission and distribution equipmentr equired to reduce them to their optimum value, more so where a significant portion of generation is based on fossil fuels. This tendency is accentuated if environmental costs of power generation (harmful local pollutants as well as greenhouse gas emissions) and increasing difficulties in achieving social acceptance of new power plant construction (regardless of fuel type and technology) are considered.

Non-technical losses represent an avoidable financial loss for the utility. Although it is clear that the amounts of electricity involved in non-technical losses are being consumed by users that do not pay for them, experience shows that a significant percentage of those amounts (in some cases more than 50 percent) becomes reduced demand when those users have to pay for that electricity,2 because they adjust their consumption to their ability to pay for electricity services. That reduction in demand has the same effect as a reduction in technical losses: less

electricity needs to be generated. Thus, from the country's perspective, reductions in nontechnical losses are also positive. From a social point of view, non-technical losses have several perverse effects. Customers being billed for accurately measured consumption and regularly paying their bills are subsidizing those users who do not pay for electricity consumption. There is a wide range of situations creating non-technical losses. A classic case is a theft of electricity through an illegal connection to the grid or tampering of a consumption meter. But examples also include unmetered consumption by utility customers who are not accurately metered for a variety of reasons. In all the cases some level of poor management of the utility in execution of its operations is present.

1.4. Research Hypotheses

The Research Aim focusing on the factor and concern that underpin the most effective technology adoption by Power utilities for their various process and current issues faces which can be addressed. The Research Methodology is explained later which showcases description of the research methods used for various objectives mentioned earlier in this study.

The research endeavor was to understand how these power distribution companies have been using technologies, to what extent in their various operational process as well as business processes. The research design adopted in the study is Hypothesis-Testing Research Design. In this type of research design researcher test hypothesis between two or more variables.

The data collected then used for further analysis.
Hypothesis based on technological implementation level
H0- The level of technology implementation is same in power distribution companies.
H1- There is a difference in level of technology implementation in power distribution companies

Hypothesis based on relationship
H0: The technology implementation index & AT&C loss are correlated in power distribution companies.

H01: There is no correlation between technology implementation index & AT&C loss of power distribution companies.

Chapter 2

2. Literature Review

Kam-Hoong Cheong (1999), in his thesis on IT Strategy for utilities did extensive study of existing IT strategies within electric utilities, mostly formulated with the traditional mindset of the industrial-age that focused primarily on technology, is inadequate and ineffective in dealing with today's IT which has a broader implication on an organization's business activities. In order to enhance the effectiveness of future IT strategy, a framework that adopts system approach is proposed in this thesis.

United Energy (2013), a Victorian power distribution Company in Melbourne, in its Distribution Annual Planning Report, elaborates the company strategy in effective utilization of various technologies like GIS, SAP, SCADA, AMI etc. for better services to the customers and increasing efficiency of the organization.Robinson (2013) has effectively tried to give the brief idea regarding the impact of electricity on the economic status of people in Delhi and effectiveness of Delhi Electricity Regulatory Commission (DERC) regulation for annual/multiyear tariff rates calculation.

In July 2009, Mr. Pedro Antmann of the Energy Unit of the World Bank, published a background paper for World Bank Group energy sector strategy. In this paper, Mr. Pedro reviews the efforts taken by private and as well as state-owned electricity companies to reduce total losses in transmission and distribution in developing countries. He did a study on continents like Latin America, East Asia, South Asia, former Soviet Union and sub-Saharan Africa. He also published some relevant cases of successful and sustainable reduction of total losses in electricity distribution. The distribution companies included are AndhraPradesh State Electricity Board, North Delhi Power Corporation, DELSUR in El Salvador, Enersis of Chile, CODENSA (Bogotá, Colombia).

In 2017, Congrès International des Réseaux Electriques de Distribution (CIRED), France, published a report on 'Reduction of Technical and Non-Technical Losses in distribution network'. In this paper, CIRED focused on main challenges regarding losses assessment and reduction. The report involved evolutions in network, in a technical point of view and regulatory point of view. CIRED also included different methodologies to valuate both technical and non-technical losses, identifying main regulatory frames and corresponding incentives and roadblocks, benchmarking best practices for losses treatment and positioning emerging techniques to reduce losses. 5

Dr. Jawed Nagi, a Robotics and AI Researcher in New York University (NYU) USA published a paper in 2008 on analysis of Non-technical loss using support vector machines. In his paper, he introduced a new approach towards Non-Technical Loss (NTL) analysis for electric utilities using a novel intelligence-based technique, Support Vector Machine (SVM). The focus was to assist Tenaga Nasional Berhad (TNB) in Malaysia to reduce its NTLs in the distribution sector due to electricity theft. The proposed model preselects suspected customers to be inspected onsite for fraud based on irregularities and abnormal consumption behavior. This approach provides a method of data mining and involves feature extraction from historical customer consumption data.

Chapter 3

3. Research Design, Methodology and Plan

3.1. Data Sources

The major data sources for the research are:
1. Survey answers
2. Personal Interviews
3. History reports
4. Technology refresh reports

3.1.1. Survey

A set of 30 survey questions are prepared and sent to various discoms in Asia pacific regions excluding Australia. The major countries included are India, Malaysia, Thailand and Singapore. Most of the utilities have find it useful and replied. The survey includes questions related to technology awareness, intention to adopt new technology, organizational intention and financial implications. The survey includes utility personnel from various departments including metering, billing, IT, analytics. The survey also includes people from various level starting from General Managers to Graduate Engineers.

3.1.2. Personal Interviews

Interviews conducted with utility personnel on the basis of survey response. The interview questions are designed to go deep and understand the pain points and the action items utility are working to reduce NTL. The interviews were designed to explore current ways Utility is adopting to protect revenues and the drawbacks in the used methods. The interview also helped to gather data on the need of utility in technology terms.

3.1.3. History reports:

Different history reports are followed to gather data on previous technology adoption methods and its success or failure. Government websites are studied to understand previous attempts in this field and its success criteria. RAPDRP reports from Govt of India websites has information about all RAPDRP initiatives and their status. Similarly, proof of concept reports from Utilities in Thailand are studies and very good insight has been captured on their technology adoption attempts. Technology refresh reports from utilities in Malaysia are followed to gather information about issues in different technology and the methods they adopted to recoup it.

3.2. Research Design

A **research design** is the set of methods and procedures used in collecting and analyzing measures of the variables specified in the problem research. The design of a study defines the study type (descriptive, correlation, semi-experimental, experimental, review, meta-analytic) and sub-type (e.g., descriptive-longitudinal case study), research problem, hypotheses, independent and dependent variables, experimental design, and, if applicable, data collection methods and a statistical analysis plan. A research design is a framework that has been created to find answers to research questions.

Research Design may be generally classified as:

i. Action Research Design
ii. Hypothesis Research Design
iii. Descriptive Design
iv. Exploratory Research Design

3.2.1. Action Research Design

The essentials of action research design follow a characteristic cycle whereby initially an exploratory stance is adopted, where an understanding of a problem is developed, and plans are made for some form of strategy. Then the intervention is carried out [the "action" in action research] during which time, pertinent observations are collected in various forms. The new interventional strategies are carried out, and this cyclic process repeats, continuing until a enough understanding of [or a valid implementation solution for] the problem is achieved. The protocol is iterative or cyclical in nature and is intended to foster deeper understanding of a given situation, starting with conceptualizing and particularizing the problem and moving through several interventions and evaluations.

These studies tell

1. This is a collaborative and adaptive research design that lends itself to use in work or community situations.
2. Design focuses on pragmatic and solution-driven research outcomes rather than testing theories.
3. When practitioners use action research, it has the potential to increase the amount they learn consciously from their experience; the action research cycle can be regarded as a learning cycle.
4. Action research studies often have direct and obvious relevance to improving practice and advocating for change.
5. There are no hidden controls or preemption of direction by the researcher.

These studies don't tell

1. It is harder to do than conducting conventional research because the researcher takes on responsibilities of advocating for change as well as for researching the topic.
2. Action research is much harder to write up because it is less likely that you can use a standard format to report your findings effectively [i.e., data is often in the form of stories or observation].
3. Personal over-involvement of the researcher may bias research results.
4. The cyclic nature of action research to achieve its twin outcomes of action [e.g. change] and research [e.g. understanding] is time-consuming and complex to conduct.
5. Advocating for change usually requires buy-in from study participants.

3.2.2. Hypothesis-Testing Research Design

The hypothesis is an explicit statement as to what we believe to be true about the observed phenomena. The scientific method then instructs researchers to test the hypothesis. Testing usually involves designing a method or protocol for collecting information (data) that will allow us to evaluate the hypothesis, and finally, to accept or reject it.

In this type of research design researcher test hypothesis between two or more variables. In this research study of the identified research variables requires the use of quantitative as well as qualitative methods. Thus Data were collected using a questionnaire based survey hours. The data collected were then used for further analysis.

3.2.3. Descriptive Design

Descriptive research designs help provide answers to the questions of who, what, when, where, and how associated with a research problem; a descriptive study cannot conclusively ascertain answers to why. Descriptive research is used to obtain information concerning the current status of the phenomena and to describe "what exists" with respect to variables or conditions in a situation.

These studies tell

1. The subject is being observed in a completely natural and unchanged natural environment. True experiments, whilst giving analyzable data, often adversely influence the normal behavior of the subject [a.k.a., the Heisenberg effect whereby measurements of certain systems cannot be made without affecting the systems].
2. Descriptive research is often used as a pre-cursor to more quantitative research designs with the general overview giving some valuable pointers as to what variables are worth testing quantitatively.
3. If the limitations are understood, they can be a useful tool in developing a more focused study.
4. Descriptive studies can yield rich data that lead to important recommendations in practice.
5. Approach collects a large amount of data for detailed analysis.

These studies don't tell

1. The results from a descriptive research cannot be used to discover a definitive answer or to disprove a hypothesis.
2. Because descriptive designs often utilize observational methods [as opposed to quantitative methods], the results cannot be replicated.
3. The descriptive function of research is heavily dependent on instrumentation for measurement and observation.

3.2.4. Exploratory Research Design

Exploratory research is a research conducted for a problem that has not been studied more clearly, intended to establish priorities, develop operational definitions and improve the final research design.[1] Exploratory research helps determine the best research design, data-collection method and selection of subjects. It should draw definitive conclusions only with extreme caution.

Given its fundamental nature, exploratory research often relies on techniques such as:

- secondary research - such as reviewing available literature and/or data
- informal qualitative approaches, such as discussions with consumers, employees, management or competitors
- formal qualitative research through in-depth interviews, focus groups, projective methods, case studies or pilot studies

The Internet allows for research methods that are more interactive in nature. For example:

- RSS feeds efficiently supply researchers with up-to-date information
- services such as Google Alerts may send major search-engine search results by email to researchers
- services such as Google Trends track comprehensive search results over lengthy periods of time
- researchers may set up websites to attract worldwide feedback on any subject

When research aims to gain familiarity with a phenomenon or to acquire new insight into it in order to formulate a more precise problem or to develop a hypothesis, exploratory studies (also known as formulative research) come in handy. If the theory happens to be too general or too specific, a hypothesis cannot be formulated. Therefore, a need for an exploratory research may be realized and instituted to gain experience that may help in formulating a relevant hypothesis for more definite investigation.

The results of exploratory research are not usually useful for decision-making by themselves, but they can provide significant insight into a given situation. Although the results of qualitative research can give some indication as to the "why", "how" and "when" something occurs, they cannot reveal "how often" or "how many".

Exploratory research is not typically generalizable to the population at large.

Social exploratory research "seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them. The goal is to learn 'what is going on here?' and to investigate social phenomena without explicit expectations."[3] This methodology is also at times referred to as a grounded theory approach to qualitative research or interpretive research, and is an attempt to unearth a theory from the data itself rather than from a predisposed hypothesis.

Earl Babbie identifies three purposes of social-science research: exploratory, descriptive and explanatory.

Exploratory research takes place when problems are in a preliminary stage.[4] Exploratory research is used when the topic or issue is new and when data is difficult to collect. Exploratory research is flexible and can address research questions of all types (what, why, how). Exploratory research is often used to generate formal hypotheses. Shields and Tajalli link exploratory research with the conceptual framework working hypothesis. Skeptics, however, have questioned the usefulness and necessity of exploratory research in situations where prior analysis could be conducted instead.

3.3. Survey Questions

The two most common types of survey questions are closed-ended questions and open-ended questions.

3.3.1. Closed-Ended Questions

- The respondents are given a list of predetermined responses from which to choose their answer.
- The list of responses should include every possible response and the meaning of the responses should not overlap.
- An example of a close-ended survey question would be, "Please rate how strongly you agree or disagree with the following statement: 'I feel good about my work on the job.' Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?"
- A Likert scale, which is used in the example above, is a commonly used set of responses for closed-ended questions.
- Closed-ended questions are usually preferred in survey research because of the ease of counting the frequency of each response.

3.3.2. Open-Ended Questions

- Survey respondents are asked to answer each question in their own words. An example would be, "In the last 12 months, what is the new technology you learned?" Another would be, "Please tell me why you chose this technology platform?"
- It is worth noting that a question can be either open-ended or close-ended depending on how it is asked. In the previous example, if the question on technology asked respondents to choose from a given set of new technology applications instead, it would be considered close-ended.
- Responses are usually categorized into a smaller list of responses that can be counted for statistical analysis.

A well-designed questionnaire is more than a collection of questions on one or more topics. When designing a questionnaire, researchers must consider several factors that can affect participation and the responses given by survey participants.

Some of the things that are considered helping ensure high rates of participation and accurate survey responses include:

- It is important to consider the order in which questions are presented.
    - Sensitive questions, such as questions about income, drug use, or sexual activity, should generally be placed near the end of the survey. This allows a level of trust or psychological comfort to be established with the respondent before asking questions that might be embarrassing or more personal.
    - Researchers also recommend putting routine questions, such as age, gender, and marital status, at the end of the questionnaire.
    - Questions that are more central to the research topic or question and that may serve to engage the respondent should be asked early. For example, a survey on children's early development that is administered to parents should ask questions that are specific to their children in the beginning or near the beginning of the survey.
- Double-barreled questions, which ask two questions in one, are not used in the survey. An example of a double-barreled question is, "Please rate how strongly you agree or disagree with the following statement: 'I feel this software application is helping me and the utility to do better analysis.'" This question is problematic because survey respondents are asked to give one response for their feelings about two conditions of their job.
- It was advised to avoid or limit the use of professional jargon or highly specialized terms, especially in surveys of the general population.
- Question and response option text should use words that are at the appropriate reading level for research participants.
- The use of complex sentence structures should be avoided.
- Researchers should avoid using emotionally loaded or biased words and phrases.
- The length of a questionnaire is always a consideration. There is a tendency to try and ask too many questions and cover too many topics. The questionnaire should be kept to a reasonable length and only include questions that are central to the research question(s). The length should be appropriate to the mode of administration. For example, in general, online surveys are shorter than surveys administered in-person.

Questionnaires and the procedures that are used to administer them are pretested (or field tested) before they are used in a main study. The goal of the pretest is to identify any problems with how questions are asked, whether they are understood by individuals similar to those who will participate in the main study, and whether response options in close-ended questions are adequate. For example, a parent questionnaire that will be used in a large study of preschool-age children may be administered first to a small (often non-random) sample of parents in order to identify any problems with how questions are asked and understood and whether the response options that are offered to parents are adequate.

Based on the findings of the pretest, additions or modifications to questionnaire items and administration procedures are made prior to their use in the main study.

3.4. Interview Procedures

An **interview** in qualitative research is a conversation where questions are asked to elicit information. The *interviewer* is usually a professional or paid researcher, sometimes trained, who poses questions to the *interviewee*, in an alternating series of usually brief questions and answers. They can be contrasted with focus groups in which an interviewer questions a group of people and observes the resulting conversation between interviewees, or surveys which are more anonymous and limit respondents to a range of predetermined answer choices. In phenomenological or ethnographic research, interviews are used to uncover the meanings of central themes in the life world of the subjects from their own point of view.

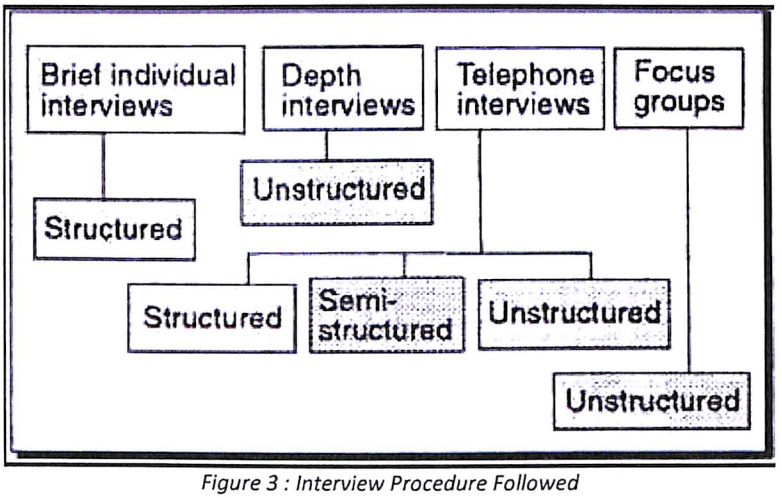


Figure 3 : Interview Procedure Followed

- Interviews are completed by the interviewer based on what the interviewee says to be conformed and done.
- Interviews are a far more personal form of research than questionnaires.
- In the personal interview, the interviewer works directly with the interviewee.
- Unlike with mail surveys, the interviewer has the opportunity to probe or ask follow up questions.
- Interviews are generally easier for the interviewee, especially if what is sought are opinions and/or impressions.
- Interviews are time-consuming and resource-intensive.
- The interviewer is considered a part of the measurement instrument and has to be well trained in how to respond to any contingency.

- Interviews provide an opportunity of face to face interaction between 2 persons; hence, they reduce conflicts.
- The two main types of interviews conducted in marketing research are structured and unstructured.

3.4.1. Unstructured informal interview

- The unstructured informal interview is normally conducted as a preliminary step in the research process to generate ideas/hypotheses about the subject being investigated so that these might be tested later in the survey proper. Such interviews are entirely informal and are not controlled by a specific set of detailed questions. Rather the interviewer is guided by a pre-defined list of issues. These interviews amount to an informal conversation about the subject.
- Informal interviewing is not concerned with discovering 'how many' respondents think in a particular way on an issue (this is what the final survey itself will discover). The aim is to find out how people think and how they react to issues, so that the ultimate survey questionnaire can be framed along the lines of thought that will be most natural to respondents.
- The respondent is encouraged to talk freely about the subject, but is kept to the point on issues of interest to the researcher. The respondent is encouraged to reveal everything that he/she feels and thinks about these points. The interviewer must note (or tape-record) all remarks that may be relevant and pursue them until he/she is satisfied that there is no more to be gained by further probing. Properly conducted, informal interviews can give the researcher an accurate feel for the subject to be surveyed. Focus groups, discussed later in this chapter, make use of relatively unstructured interviews.

3.4.2. Structured standardized interview

- With structured standardised interviews, the format is entirely different. A structured interview follows a specific questionnaire and this research instrument is usually used as the basis for most quantitative surveys. A standardised structured questionnaire is administered where specific questions are asked in a set order and in a set manner to ensure no variation between interviews.

3.4.3. Depth interviews

- Depth interviews are one-to-one encounters in which the interviewer makes use of an unstructured or semi-structured set of issues/topics to guide the discussion. The object of the exercises is to explore and uncover deep-seated emotions, motivations and attitudes. They are most often employed when dealing with sensitive matters and respondents are likely to give evasive or even misleading answers when directly questioned. Most of the techniques used in the conduct of depth interviews have been borrowed from the field of psychoanalysis. Depth interview are usually only successful when conducted by a well-trained and highly skilled interviewer.
- Other instances when depth interviewers can be particularly effective are: where the study involves an investigation of complex behavior or decision-making processes; when the target respondents are difficult to gather together for group interviewers (e.g. farmers, veterinary surgeons, haulage contractors, government officials); and where the interviewee is prepared to become an informant only if he/she is able to preserve his/her anonymity.
- Dillon et al[1]. believe that to be effective, the interviewer must adhere to six fundamental rules. These are:
    - he/she must avoid appearing superior or condescending and make use of only familiar words
    - he/she must put question indirectly and informatively
    - he/she must remain detached and objective
    - he/she must avoid questions and questions structure that encourage 'yes' or 'no' answers
    - he/she must probe until all relevant details, emotions and attitudes are revealed
    - he/she must provide an atmosphere that encourages the respondent to speak freely, yet keeping the conservation focused on the issue(s) being researched
- Depth interviews involve a heavy time commitment, especially on the part of the marketing researcher. Interview transcripts have to be painstakingly recovered, if they are to be accurate, either from terse interview notes or from tape-recordings of the interviews. This can take many hours of often laborious work. The transcripts then have to be read and re-read, possibly several times, before the researcher is able to begin the taxing process of analysing and interpreting the data.

3.4.4. Telephone Interviews

- Whilst telephone interviews among consumers, are very common in the developed world, these are conducted with far less frequency in the developing world. The reason is somewhat obvious, i.e. only a relatively small proportion of the total population has a telephone in the house. Moreover, telephone owners tend to be urban dwellers and have above average incomes and are therefore unrepresentative of the population as a whole.
- To a greater extent, telephone interviewing has potential in surveys of businesses, government agencies and other organizations or institutions. Even then, it is still the case that telephone surveys are rarely without bias. Whilst it is true that many businesses have a telephone, small businesses and even medium-sized enterprises are far less likely to have access to telephones.
- Telephone interviews afford a certain amount of flexibility. It is possible, for example, for interviewers to put complex questions over the telephone. The interviewers can probe, skip questions that prove irrelevant to the case of a particular respondent and change the sequence of questions in response to the flow of the discussion, and earlier replies can be revisited. The interaction between interviewer and interviewee that is possible over the telephone simply is not achievable through a mailed questionnaire. In comparison to personal interviews, telephone interviews do not appear to enjoy any margin of advantage. Perhaps the only advantages are those of speed and cost. Even then, manpower costs in developing countries tend to be very low and so only speed remains as a potential advantage over personal interviews.
- In the developed world, the era of computer-assisted telephone interviewing (CATI) has begun. Researchers conduct the telephone interview whilst seated at a computer. Responses are entered directly into the computer, by the interviewer. The screen displays the questionnaire and any skipping of questions, due to earlier responses directing that some questions are not applicable in the case of the interviewee, is controlled automatically by the computer. Since the responses are entered directly into the computer the data is instantaneously processed. The computer can also be programmed to produce standardised marketing reports.

3.4.5. Focus group interviews

Focus group interviews are a survey research instrument which can be used in addition to, or instead of, a personal interview approach. It has particular advantages for use in qualitative research applications. The central feature of this method of obtaining information from groups of people is that the interviewer strives to keep the discussion led by a moderator focused upon the issue of concern. The moderator behaves almost like a psycho-therapist who directs the group towards the focus of the researcher. In doing so, the moderator speaks very little, and encourages the group to generate the information required by stimulating discussion through terse provocative statements.

Characteristics of focus group interviews

The groups of individuals (e.g. housewives, farmers, manufacturers, etc.) are invited to attend an informal discussion. Usually between 6 and 8 participants are involved and the discussion would last between 1 and 2 hours. Small groups tend to lose the mutual stimulation among participants, whilst large groups can be difficult to manage and may prevent some participants having the opportunity to get fully involved in the discussion.

The researcher raises issues for discussion, following a 'guide list of topics' rather than a structured questionnaire. The participants are encouraged to discuss the issues amongst themselves and with the researcher in an informal and relaxed environment. The researcher records comments made by the participants (usually utilizing a tape or video recorder).

In contrast to a personal interview survey, the number of interviews in a typical group interview survey is very small, usually between 3 and 4 would be sufficient for each type of respondent-sector (e.g. farmers or manufacturers). Generally from the first interview on an unfamiliar subject the researcher will learn a great deal. The second and third interviews will produce more information, but not all of it will not be new. By the fourth interview most of what is revealed will have been covered before, and the diminishing returns involved would generally not justify the cost of further groups.

The participants within a focus group are selected in such a way that they exhibit a high degree of homogeneity with respect to either background, behaviour or both. Consider, for example, a study carried out by a small African nation that is looking for a niche market for a new range of sparkling wines. It is decided that, as a first step, a series of focus groups be conducted. The researchers are keen to ensure that each group comprises people who are similar in age and behaviour with respect to wine consumption.

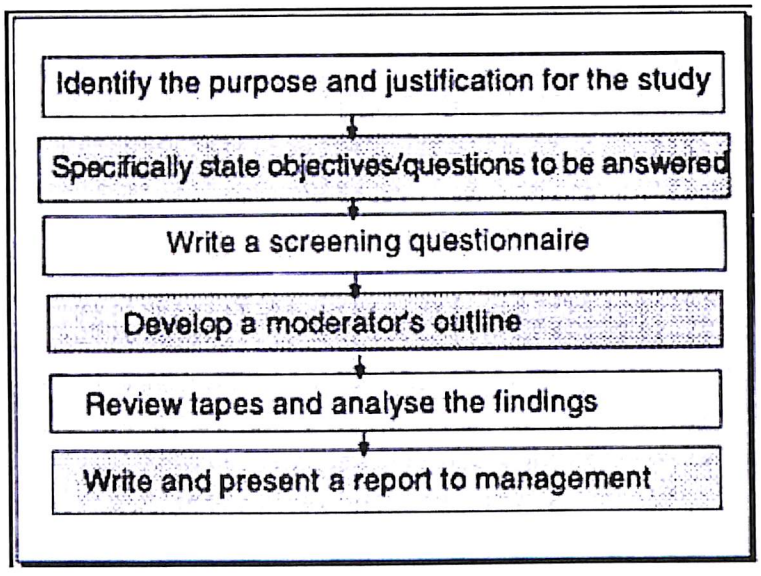


Figure 4 : Process of Interview

3.5. Data Analysis Procedures

There are differences between qualitative data analysis and quantitative data analysis. In qualitative researches using interviews, focus groups, experiments etc. data analysis is going to involve identifying common patterns within the responses and critically analyzing them in order to achieve research aims and objectives.

Data analysis for quantitative studies, on the other hand, involves critical analysis and interpretation of figures and numbers, and attempts to find rationale behind the emergence of main findings. Comparisons of primary research findings to the findings of the literature review are critically important for both types of studies – qualitative and quantitative.

Data analysis methods in the absence of primary data collection can involve discussing common patterns, as well as, controversies within secondary data directly related to the research area.

3.5.1. Qualitative data analysis

Qualitative data refers to non-numeric information such as interview transcripts, notes, video and audio recordings, images and text documents. Qualitative data analysis can be divided into the following five categories:

1. Content analysis. This refers to the process of categorizing verbal or behavioural data to classify, summarize and tabulate the data.

2. Narrative analysis. This method involves the reformulation of stories presented by respondents taking into account context of each case and different experiences of each respondent. In other words, narrative analysis is the revision of primary qualitative data by researcher.

3. Discourse analysis. A method of analysis of naturally occurring talk and all types of written text.

4. Framework analysis. This is more advanced method that consists of several stages such as familiarization, identifying a thematic framework, coding, charting, mapping and interpretation.

5. Grounded theory. This method of qualitative data analysis starts with an analysis of a single case to formulate a theory. Then, additional cases are examined to see if they contribute to the theory.

Qualitative data analysis can be conducted through the following three steps:

Step 1: Developing and Applying Codes. Coding can be explained as categorization of data. A 'code' can be a word or a short phrase that represents a theme or an idea. All codes need to be assigned meaningful titles. A wide range of non-quantifiable elements such as events, behaviours, activities, meanings etc. can be coded.

There are three types of coding:

1. *Open coding*. The initial organization of raw data to try to make sense of it.
2. *Axial coding*. Interconnecting and linking the categories of codes.
3. *Selective coding*. Formulating the story through connecting the categories.

Coding can be done manually or using qualitative data analysis software such as NVivo, Atlas ti 6.0, HyperRESEARCH 2.8, Max QDA and others.

When using manual coding you can use folders, filing cabinets, wallets etc. to gather together materials that are examples of similar themes or analytic ideas. Manual method of coding in qualitative data analysis is rightly considered as labour-intensive, time-consuming and outdated.

In computer-based coding, on the other hand, physical files and cabinets are replaced with computer based directories and files. When choosing software for qualitative data analysis you need to consider a wide range of factors such as the type and amount of data you need to analyse, time required to master the software and cost considerations.

Moreover, it is important to get confirmation from your dissertation supervisor prior to application of any specific qualitative data analysis software.

Step 2: Identifying themes, patterns and relationships. Unlike quantitative methods, in qualitative data analysis there are no universally applicable techniques that can be applied to generate findings. Analytical and critical thinking skills of researcher plays significant role in data analysis in qualitative

studies. Therefore, no qualitative study can be repeated to generate the same results.

Nevertheless, there is a set of techniques that you can use to identify common themes, patterns and relationships within responses of sample group members in relation to codes that have been specified in the previous stage.

Specifically, the most popular and effective methods of qualitative data interpretation include the following:

- *Word and phrase repetitions* - scanning primary data for words and phrases most commonly used by respondents, as well as, words and phrases used with unusual emotions;
- *Primary and secondary data comparisons* - comparing the findings of interview/focus group/observation/any other qualitative data collection method with the findings of literature review and discussing differences between them;
- *Search for missing information* - discussions about which aspects of the issue was not mentioned by respondents, although you expected them to be mentioned;
- *Metaphors and analogues* - comparing primary research findings to phenomena from a different area and discussing similarities and differences.

Step 3: Summarizing the data. At this last stage you need to link research findings to hypotheses or research aim and objectives. When writing data analysis chapter, you can use noteworthy quotations from the transcript in order to highlight major themes within findings and possible contradictions.

It is important to note that the process of qualitative data analysis described above is general and different types of qualitative studies may require slightly different methods of data analysis.

3.5.2. quantitative data analysis

In quantitative data analysis you are expected to turn raw numbers into meaningful data through the application of rational and critical thinking. Quantitative data analysis may include the calculation of frequencies of variables and differences between variables. A quantitative approach is usually associated with finding evidence to either support or reject hypotheses you have formulated at the earlier stages of your research process.

The same figure within data set can be interpreted in many different ways; therefore it is important to apply fair and careful judgement.

For example, questionnaire findings of a research titled "A study into the impacts of informal management-employee communication on the levels of employee motivation: a case study of Agro Bravo Enterprise" may indicate that the majority 52% of respondents assess communication skills of their immediate supervisors as inadequate.

This specific piece of primary data findings needs to be critically analyzed and objectively interpreted through comparing it to other findings within the framework of the same research. For example, organizational culture of Agro Bravo Enterprise, leadership style, the levels of frequency of management-employee communications need to be taken into account during the data analysis.

Moreover, literature review findings conducted at the earlier stages of the research process need to be referred to in order to reflect the viewpoints of other authors regarding the causes of employee dissatisfaction with management communication. Also, secondary data needs to be integrated in data analysis in a logical and unbiased manner.

A set of analytical software can be used to assist with analysis of quantitative data. The following table illustrates the advantages and disadvantages of three popular quantitative data analysis software: Microsoft Excel, Microsoft Access and SPSS.

	Advantages	Disadvantages
Excel Spreadsheet	Cost effective or Free of Charge Can be sent as e-mail attachments & viewed by most smartphones All in one program Excel files can be secured by a password	Big Excel files may run slowly Numbers of rows and columns are limited Advanced analysis functions are time consuming to be learned by beginners Virus vulnerability through macros
Microsoft Access	One of the cheapest amongst premium programs Flexible information retrieval Ease of use	Difficult in dealing with large database Low level of interactivity Remote use requires installation of the same version of Microsoft Access
SPSS	Broad coverage of formulas and statistical routines Data files can be imported through other programs Annually updated to increase sophistication	Expensive cost Limited license duration Confusion among the different versions due to regular update

Figure 5 : Advantages and Disadvantages of different application

Quantitative data analysis with the application of statistical software consists of the following stages:

1. Preparing and checking the data. Input of data into computer.
2. Selecting the most appropriate tables and diagrams to use according to your research objectives.
3. Selecting the most appropriate statistics to describe your data.
4. Selecting the most appropriate statistics to examine relationships and trends in your data.

It is important to note that while the application of various statistical software and programs are invaluable to avoid drawing charts by hand or undertake calculations manually, it is easy to use them incorrectly. In other words, quantitative data analysis is "a field where it is not at all difficult to carry out an analysis which is simply wrong, or inappropriate for your data or purposes. And the negative side of readily available specialist statistical software is that it becomes that much easier to generate elegantly presented rubbish".

Therefore, it is important for you to seek advice from your dissertation supervisor regarding statistical analyses in general and the choice and application of statistical software in particular.

3.5.3. Research variables

Factors affecting the assessment of Technology and adoption of GIS for power distribution sector are the research constructs that are required to be studied. These constructs namely Perceived Ease of Use, Attitude, Efficiency gain, Perceived Usefulness, Discom Organization Culture and Intention all were selected on the basis of literature review done earlier. All the variables are studied extensively because these would be factors determining the adoption of geospatial technologies, assessment of technology and website of Discoms.

Macro Variables also referred as Constructs with attached indicators, which can also be termed as micro variables. Indicators are the dimensions that help in measuring and quantifying the macro variables. These needs to be elaborated and explained. Literature review and understanding of the power distribution sector has helped in identification of the macro variables for the identified factors which would be further divided into the measurable dimensions. Variables for which dimensions are identified and can be measured quantitatively are measured and analyzed through the quantitative assessment approach. For other variables, dimensions are identified, measured and analyzed through further analysis utilizing qualitative analysis approach. Table-9 list macro variables identified macro variables and the assessment approach for each macro variables.

Macro Variables and its related Micro Variables identified and Study Required for Each Variable			
Sl No.	Macro Variables	Micro variables	Type of Study Required
	Website Assessment Factors		
1	Information Content	Not Required	Qualitative / Quantitative Assessment
2	Technical Details		Qualitative / Quantitative Assessment
3	Commercial		Qualitative / Quantitative Assessment
4	Communication Content		Qualitative / Quantitative Assessment
5	Sustainability		Qualitative / Quantitative Assessment
6	Customer Service		Qualitative / Quantitative Assessment
7	Payment Options		Qualitative / Quantitative Assessment
8	Regulator Information		Qualitative / Quantitative Assessment
9	Govt of India (Ministry of Power Initiatives)		Qualitative / Quantitative Assessment
10	Consumer Care Services		Qualitative / Quantitative Assessment
11	Me Details	Not Required	Qualitative / Quantitative Assessment
12	Contact Us	Not Required	Qualitative / Quantitative Assessment
13	Customer Information Centre	Not Required	Qualitative Assessment/ Quantitative Assessment

Figure 6 : Research Variables

	Technology Index Factors		Qualitative / Quantitative Assessment
17	Metering	Installation of Electronic Meters -LT Consumers	Qualitative / Quantitative Assessment
		AMR for HT Consumers	Qualitative / Quantitative Assessment
		AMR for Feeder Meters	Qualitative / Quantitative Assessment
		AMR for DTs	Qualitative / Quantitative Assessment
		CMRI Metering reading for HT/LT Consumers	Qualitative / Quantitative Assessment
		Prepaid Metering	Qualitative / Quantitative Assessment
18	IT Distribution Apps	Energy Accounting System	Qualitative / Quantitative Assessment
		SCADA	Qualitative / Quantitative Assessment
		Load Forecasting Applications	Qualitative / Quantitative Assessment
		Meter Data Management System	Qualitative / Quantitative Assessment
		Outage Management System	Qualitative / Quantitative Assessment
		Geographical Information System & Network Analysis	Qualitative / Quantitative Assessment
		Smart Grid Pilot Project	Qualitative / Quantitative Assessment

Figure 7 : Research variables 2

	Technology Adoption Factors		Qualitative / Quantitative Assessment
21	Perceived Usefulness Referred from Davis (1989); Davis et al. (1989); Moore and Benbasat, (1991)	GIS provides me better picture of my queries	Qualitative / Quantitative Assessment
		GIS Increase visualization of and extent of the problem.	Qualitative / Quantitative Assessment
		GIS is good feedback system to confirm reports.	Qualitative / Quantitative Assessment
		GIS improves in my decision making.	Qualitative / Quantitative Assessment
		GIS is overall useful in my day a today work	Qualitative / Quantitative Assessment
22	Perceived Ease of Use. Referred from Davis (1989); Davis et al. (1989); Moore and Benbasat, (1991)	My interface with the GIS is clear and understandable.	Qualitative / Quantitative Assessment
		Using a GIS application does not require a lot of skill.	Qualitative / Quantitative Assessment

Figure 8 : Research Variables 3

23	Social Influence. Referred from Thompson et al. (1991); Venkatesh et al. (2003)	Within Depts, the environment is good for use of GIS.	Qualitative / Quantitative Assessment
		Other Depts. personnel often asks for GIS-based reports	Qualitative / Quantitative Assessment
		Senior Management encourages use of GIS.	Qualitative / Quantitative Assessment
		Workforce think GIS as reliable decision support system	Qualitative / Quantitative Assessment
24	Efficiency gain Referred from Davis (1989); Davis et al. (1989); Moore and Benbasat, (1991)	Using GIS has reduced my field visit to large extend	Qualitative / Quantitative Assessment
		Using GIS saves my time.	Qualitative / Quantitative Assessment
		GIS has provided a single platform for field staff & supervisors to view the problem.	Qualitative / Quantitative Assessment
		GIS has helped in providing better services to customers	Qualitative / Quantitative Assessment
		GIS has reduced fault time and helped in faster fault restoration	Qualitative / Quantitative Assessment
25	Attitude Adapted from Taylor and Todd (1995); Thompson et al. (1991); Venkatesh et al. (2003)	I like working with user-friendly system like GIS	Qualitative / Quantitative Assessment
		Using GIS system is exciting	Qualitative / Quantitative Assessment
		GIS system makes work more attention-grabbing.	Qualitative / Quantitative Assessment
		Working with the GIS system is enjoyable.	Qualitative / Quantitative Assessment

Figure 9 : Research variables 4

3.5.3.1. Perceived Ease of Use

It discusses to use of technology in which user find how easy is the system to use without requiring much of effort. The technology introduced here with referred as GIS technology should be easy to reminisce and should be easy to absorb and disseminate. The interface of the GIS system should be

trouble-free to work upon and a person can start using these technologies after basic training.

3.5.3.2. Attitude

Attitude toward use is defined as "an individual's overall affective reaction to using system" (Venkatesh, et al., 2003, p. 455). Melone (1990) also defines user attitude in "a predisposition to respond favorably or unfavorably to a computer system, application, system staff member, or a process related to the use of that system of Application".

3.5.3.3. Efficiency Gain

Efficiency gain refers to how the GIS technology has helped them in their current daily activities. Also how the geospatial technologies have reduced their efforts while giving the same output and reduced field efforts. Since GIS empowers the field engineers about field conditions while sitting on their PC in the office, it used to measure the degree of which technology has helped them in their work.

3.5.3.4. Perceived Usefulness

This was defined by Fred Davis as "the degree to which a person believes that using a particular system would enhance his or her job performance".

Construct	Author of the scales	No of indicators	Type of scale
Perceived Usefulness	Davis (1989): Davis et al. (1989): Moore and Benbasat. (1991)	5	Likert Scale
Perceived Ease of Use	Davis (1989): Davis et al. (1989): Moore and Benbasat. (1991)	5	Likert Scale
Efficiency gain	Davis (1989): Davis et al. (1989): Moore and Benbasat. (1991)	5	Likert Scale
Attitude	Taylor and Todd (1995): Thompson et al. (1991): Venkatesh et al. (2003)	5	Likert Scale
Intention	Venkatesh & Davis (1996): Venkatesh & Davis (2000): Agarwal & Prasad (1997)	5	Likert Scale
Top Management Support & Govt Initiatives	Proposed by Research Scholar	5	Likert Scale

Figure 10 : Constructs

Chapter 4

4. Findings and Analysis

4.1. Descriptive Statistics

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the mean, median, and mode, while measures of variability include the standard deviation, variance, the minimum and maximum variables, and the kurtosis and skewness.

Descriptive statistics, in short, help describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data. The most recognized types of descriptive statistics are measures of center: the mean, median, and mode, which are used at almost all levels of math and statistics. The mean, or the average, is calculated by adding all the figures within the data set and then dividing by the number of figures within the set. For example, the sum of the following data set is 20: (2, 3, 4, 5, 6). The mean is 4 (20/5). The mode of a data set is the value appearing most often, and the median is the figure situated in the middle of the data set. It is the figure separating the higher figures from the lower figures within a data set. However, there are less-common types of descriptive statistics that are still very important.

All descriptive statistics are either measures of central tendency or measures of variability, also known as measures of dispersion. Measures of central tendency focus on the average or middle values of data sets; whereas, measures of variability focus on the dispersion of data. These two measures use graphs, tables, and general discussions to help people understand the meaning of the analyzed data.

Measures of central tendency describe the center position of a distribution for a data set. A person analyzes the frequency of each data point in the distribution and describes it using the mean, median, or mode, which measures the most common patterns of the analyzed data set.

Measures of variability, or the measures of spread, aid in analyzing how spread-out the distribution is for a set of data. For example, while the measures of central tendency may give a person the average of a data set, it does not describe how the data is distributed within the set. So, while the average of the data may be 65 out of 100, there can still be data points at both 1 and 100. Measures of variability help communicate this by describing the shape and spread of the data set. Range, quartiles, absolute deviation, and variance are all examples of measures of variability.

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

Descriptive statistics are typically distinguished from inferential statistics. With descriptive statistics you are simply describing what is or what the data shows. With inferential statistics, you are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to try to infer from the sample data what the population might think. Or, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study. Thus, we use inferential statistics to make inferences from our data to more general conditions; we use descriptive statistics simply to describe what's going on in our data.

Descriptive Statistics are used to present quantitative descriptions in a manageable form. In a research study we may have lots of measures. Or we may measure a large number of people on any measure. Descriptive statistics help us to simplify large amounts of data in a sensible way. Each descriptive statistic reduces lots of data into a simpler summary. For instance, consider a simple number used to summarize how well a batter is performing in baseball, the batting average. This single number is simply the number of hits divided by the number of times at bat (reported to three significant digits). A batter who is hitting .333 is getting a hit one time in every three at bats. One batting .250 is hitting one time in four. The single number describes a large number of discrete events. Or, consider the scourge of many students, the Grade Point Average (GPA). This single number describes the general performance of a student across a potentially wide range of course experiences.

4.2. Univariate Analysis

Univariate analysis involves the examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at:

- the distribution
- the central tendency
- the dispersion

In most situations, we would describe all three of these characteristics for each of the variables in our study.

4.2.1. The distribution

The Ditribution is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of persons who had each value. For instance, a typical way to describe the distribution of college students is by year in college, listing the number or percent of students at each of the four years. Or, we describe gender by listing the number or percent of males and females. In these cases, the variable has few enough values that we can list each one and summarize how many sample cases had the value. But what do we do for a variable like income or GPA? With these variables there can be a large number of possible values, with relatively few people having each one. In this case, we group the raw scores into categories according to ranges of values. For instance, we might look at GPA according to the letter grade ranges. Or, we might group income into four or five ranges of income values.

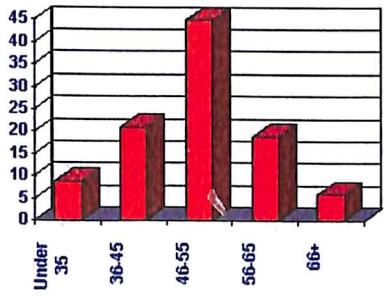


Figure 11 : Frequency Distribution

4.2.2. Dispersion.

Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. The range is simply the highest value minus the lowest value. The Standard Deviation is a more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range (as was true in this example where the single outlier value of 36 stands apart from the rest of the values. The Standard Deviation shows the relation that set of scores has to the mean of the sample.

$$\sqrt{\frac{\Sigma(X - \bar{X})^2}{(n - 1)}}$$

where:

X = each score
$\bar{X}$ = the mean or average
n = the number of values
Σ means we sum across the values

Figure 12 : Standard Deviation

In the top part of the ratio, the numerator, we see that each score has the mean subtracted from it, the difference is squared, and the squares are summed. In the bottom part, we take the number of scores minus 1. The ratio is the variance and the square root is the standard deviation. In English, we can describe the standard deviation as: the square root of the sum of the squared deviations from the mean divided by the number of scores minus one

The standard deviation allows us to reach some conclusions about specific scores in our distribution. Assuming that the distribution of scores is normal or bell-shaped (or close to it!), the following conclusions can be reached:

- approximately 68% of the scores in the sample fall within one standard deviation of the mean
- approximately 95% of the scores in the sample fall within two standard deviations of the mean
- approximately 99% of the scores in the sample fall within three standard deviations of the mean

4.3. Correlation/ Regression Analyses

Correlation and Regression are the two-analysis based on multivariate distribution. A multivariate distribution is described as a distribution of multiple variables. **Correlation** is described as the analysis which lets us know the association or the absence of the relationship between two variables 'x' and 'y'. On the other end, **Regression** analysis, predicts the value of the dependent variable based on the known value of the independent variable, if average mathematical relationship between two or more variables.

The difference between correlation and regression is one of the commonly asked questions in interviews. Moreover, many people suffer ambiguity in understanding these two. So, take a full read of this article to have a clear understanding on these two.

Basis for Comparison	Correlation	Regression
Meaning	Correlation is a statistical measure which determines co-relationship or association of two variables.	Regression describes how an independent variable is numerically related to the dependent variable.
Usage	To represent linear relationship between two variables.	To fit a best line and estimate one variable on the basis of another variable.
Dependent and Independent variables	No difference	Both variables are different.
Indicates	Correlation coefficient indicates the extent to which two variables move together.	Regression indicates the impact of a unit change in the known variable (x) on the estimated variable (y).
Objective	To find a numerical value expressing the relationship between variables.	To estimate values of random variable on the basis of the values of fixed variable.

Figure 13 : Comparison Chart

Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables. A model of the relationship is hypothesized and estimates of the parameter values are used to develop an estimated regression equation.

4.3.1. Regression model.

In simple linear regression, the model used to describe the relationship between a single dependent variable y and a single independent variable x is $y = a_0 + a_1x + k$. $a_0$and a_1 are referred to as the model parameters, and is a probabilistic error term that accounts for the variability in y that cannot be explained by the linear relationship with x. If the error term were not present, the model would be deterministic; in that case, knowledge of the value of x would be sufficient to determine the value of y.

4.3.2. Least squares method.

Either a simple or multiple regression model is initially posed as a hypothesis concerning the relationship among the dependent and independent variables. The least squares method is the most widely used procedure for developing estimates of the model parameters.

4.3.3. Correlation.

Correlation and regression analysis are related in the sense that both deal with relationships among variables. The correlation coefficient is a measure of linear association between two variables. Values of the correlation coefficient are always between -1 and +1. A correlation coefficient of +1 indicates that two variables are perfectly related in a positive linear sense, a correlation coefficient of -1 indicates that two variables are perfectly related in a negative linear sense, and a correlation coefficient of 0 indicates that there is no linear relationship between the two variables. For simple linear regression, the sample correlation coefficient is the square root of the coefficient of determination, with the sign of the correlation coefficient being the same as the sign of b1, the coefficient of x1 in the estimated regression equation.

Neither regression nor correlation analyses can be interpreted as establishing cause-and-effect relationships. They can indicate only how or to what extent variables are associated with each other. The correlation coefficient measures only the degree of linear association between two variables. Any conclusions about a cause-and-effect relationship must be based on the judgment of the analyst.

Correlation is a measure of association between two variables. The variables are not designated as dependent or independent. The two most popular correlation

coefficients are: Spearman's correlation coefficient rho and Pearson's product-moment correlation coefficient.

When calculating a correlation coefficient for ordinal data, select Spearman's technique. For interval or ratio-type data, use Pearson's technique.

The value of a correlation coefficient can vary from minus one to plus one. A minus one indicates a perfect negative correlation, while a plus one indicates a perfect positive correlation. A correlation of zero means there is no relationship between the two variables. When there is a negative correlation between two variables, as the value of one variable increases, the value of the other variable decreases, and vise versa. In other words, for a negative correlation, the variables work opposite each other. When there is a positive correlation between two variables, as the value of one variable increases, the value of the other variable also increases. The variables move together.

The standard error of a correlation coefficient is used to determine the confidence intervals around a true correlation of zero. If your correlation coefficient falls outside of this range, then it is significantly different than zero. The standard error can be calculated for interval or ratio-type data (i.e., only for Pearson's product-moment correlation).

The significance (probability) of the correlation coefficient is determined from the t-statistic. The probability of the t-statistic indicates whether the observed correlation coefficient occurred by chance if the true correlation is zero. In other words, it asks if the correlation is significantly different than zero. When the t-statistic is calculated for Spearman's rank-difference correlation coefficient, there must be at least 30 cases before the t-distribution can be used to determine the probability. If there are fewer than 30 cases, you must refer to a special table to find the probability of the correlation coefficient.

Simple regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known. Regression goes beyond correlation by adding prediction capabilities.

People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children's diet results in higher energy levels. The ease of waking up in the

morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

For example, a medical researcher might want to use body weight (independent variable) to predict the most appropriate dose for a new drug (dependent variable). The purpose of running the regression is to find a formula that fits the relationship between the two variables. Then you can use that formula to predict values for the dependent variable when only the independent variable is known. A doctor could prescribe the proper dose based on a person's body weight.

The regression line (known as the *least squares line*) is a plot of the expected value of the dependent variable for all values of the independent variable. Technically, it is the line that "minimizes the squared residuals". The regression line is the one that best fits the data on a scatterplot.

Using the regression equation, the dependent variable may be predicted from the independent variable. The slope of the regression line (b) is defined as the rise divided by the run. The y intercept (a) is the point on the y axis where the regression line would intercept the y axis. The slope and y intercept are incorporated into the regression equation. The intercept is usually called the constant, and the slope is referred to as the coefficient. Since the regression model is usually not a perfect predictor, there is also an error term in the equation.

In the regression equation, y is always the dependent variable and x is always the independent variable. Here are three equivalent ways to mathematically describe a linear regression model.

y = intercept + (slope ±x) + error

y = constant + (coefficient±x) + error

y = a + bx + e

The significance of the slope of the regression line is determined from the t-statistic. It is the probability that the observed correlation coefficient occurred by chance if the true correlation is zero. Some researchers prefer to report the F-ratio instead of the t-statistic. The F-ratio is equal to the t-statistic squared.

The t-statistic for the significance of the slope is essentially a test to determine if the regression model (equation) is usable. If the slope is significantly different than zero, then we can use the regression model to predict the dependent variable for any value of the independent variable.

On the other hand, take an example where the slope is zero. It has no prediction ability because for every value of the independent variable, the prediction for the dependent variable would be the same. Knowing the value of the independent variable would not improve our ability to predict the dependent variable. Thus, if the slope is not significantly different than zero, don't use the model to make predictions.

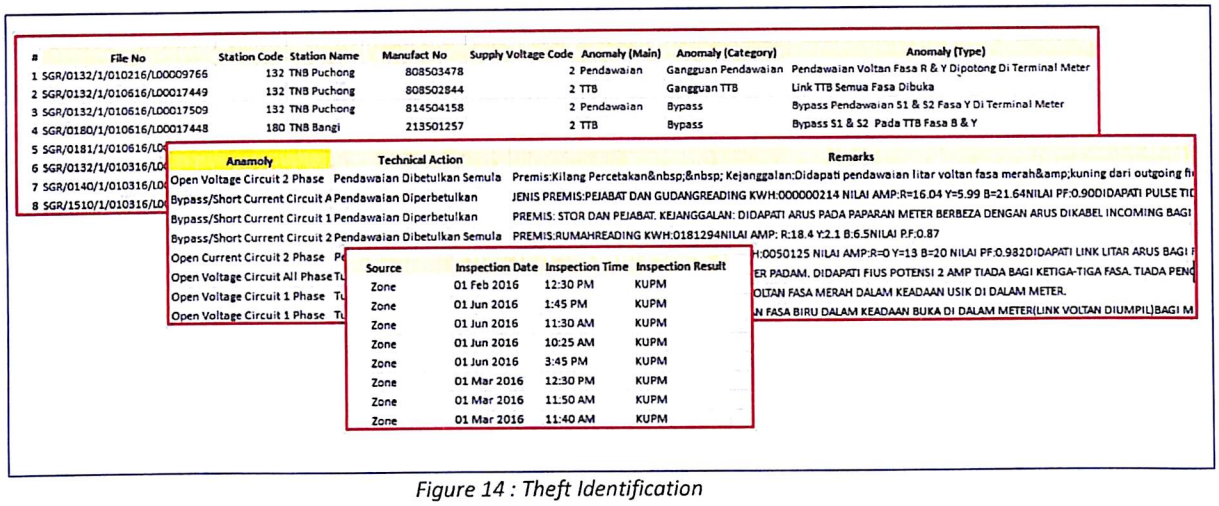
The coefficient of determination (r-squared) is the square of the correlation coefficient. Its value may vary from zero to one. It has the advantage over the correlation coefficient in that it may be interpreted directly as the proportion of variance in the dependent variable that can be accounted for by the regression equation. For example, an r-squared value of .49 means that 49% of the variance in the dependent variable can be explained by the regression equation. The other 51% is unexplained.

The standard error of the estimate for regression measures the amount of variability in the points around the regression line. It is the standard deviation of the data points as they are distributed around the regression line. The standard error of the estimate can be used to develop confidence intervals around a prediction.

4.4. Findings on Tenaga Nasional Berhad

It can be concluded from the study that Tenaga Nasional Berhad has always a fore front in adopting new technologies. The utility is also investing on its resources to learn and research new trends in theft reduction methods.

A sample of the customer base with Theft identification



#	File No	Station Code	Station Name	Manufact No	Supply Voltage Code	Anomaly (Main)	Anomaly (Category)	Anomaly (Type)
1	SGR/0132/1/010216/L00009766	132	TNB Puchong	808503478	2	Pendawaian	Gangguan Pendawaian	Pendawaian Voltan Fasa R & Y Dipotong Di Terminal Meter
2	SGR/0132/1/010616/L00017449	132	TNB Puchong	808502844	2	TTB	Gangguan TTB	Link TTB Semua Fasa Dibuka
3	SGR/0132/1/010616/L00017509	132	TNB Puchong	814504158	2	Pendawaian	Bypass	Bypass Pendawaian S1 & S2 Fasa Y Di Terminal Meter
4	SGR/0180/1/010616/L00017448	180	TNB Bangi	213501257	2	TTB	Bypass	Bypass S1 & S2 Pada TTB Fasa B & Y
5	SGR/0181/1/010616/L0							
6	SGR/0132/1/010316/L0							
7	SGR/0140/1/010316/L0							
8	SGR/1510/1/010316/L0							

Anamoly	Technical Action	Remarks
Open Voltage Circuit 2 Phase	Pendawaian Dibetulkan Semula	Premis:Kilang Percetakan Kejanggalan:Didapati pendawaian litar voltan fasa merah&kuning dari outgoing fi
Bypass/Short Current Circuit A	Pendawaian Diperbetulkan	JENIS PREMIS:PEJABAT DAN GUDANGREADING KWH:000000214 NILAI AMP:R=16.04 Y=5.99 B=21.64NILAI PF:0.90DIDAPATI PULSE TID
Bypass/Short Current Circuit 1	Pendawaian Diperbetulkan	PREMIS: STOR DAN PEJABAT. KEJANGGALAN: DIDAPATI ARUS PADA PAPARAN METER BERBEZA DENGAN ARUS DIKABEL INCOMING BAGI
Bypass/Short Current Circuit 2	Pendawaian Dibetulkan Semula	PREMIS:RUMAHREADING KWH:0181294NILAI AMP: R:18.4 Y:2.1 B:6.5NILAI P.F:0.87
Open Current Circuit 2 Phase	P	H:0050125 NILAI AMP:R=0 Y=13 B=20 NILAI PF:0.982DIDAPATI LINK LITAR ARUS BAGI F
Open Voltage Circuit All Phase	Tu	ER PADAM. DIDAPATI FIUS POTENSI 2 AMP TIADA BAGI KETIGA-TIGA FASA. TIADA PENC
Open Voltage Circuit 1 Phase	Tu	OLTAN FASA MERAH DALAM KEADAAN USIK DI DALAM METER.
Open Voltage Circuit 1 Phase	Tu	N FASA BIRU DALAM KEADAAN BUKA DI DALAM METER(LINK VOLTAN DIUMPIL)BAGI M

Source	Inspection Date	Inspection Time	Inspection Result
Zone	01 Feb 2016	12:30 PM	KUPM
Zone	01 Jun 2016	1:45 PM	KUPM
Zone	01 Jun 2016	11:30 AM	KUPM
Zone	01 Jun 2016	10:25 AM	KUPM
Zone	01 Jun 2016	3:45 PM	KUPM
Zone	01 Mar 2016	12:30 PM	KUPM
Zone	01 Mar 2016	11:50 AM	KUPM
Zone	01 Mar 2016	11:40 AM	KUPM

Figure 14 : Theft Identification

TNB uses data mining to find the correlation of different meter events and its implications to theft

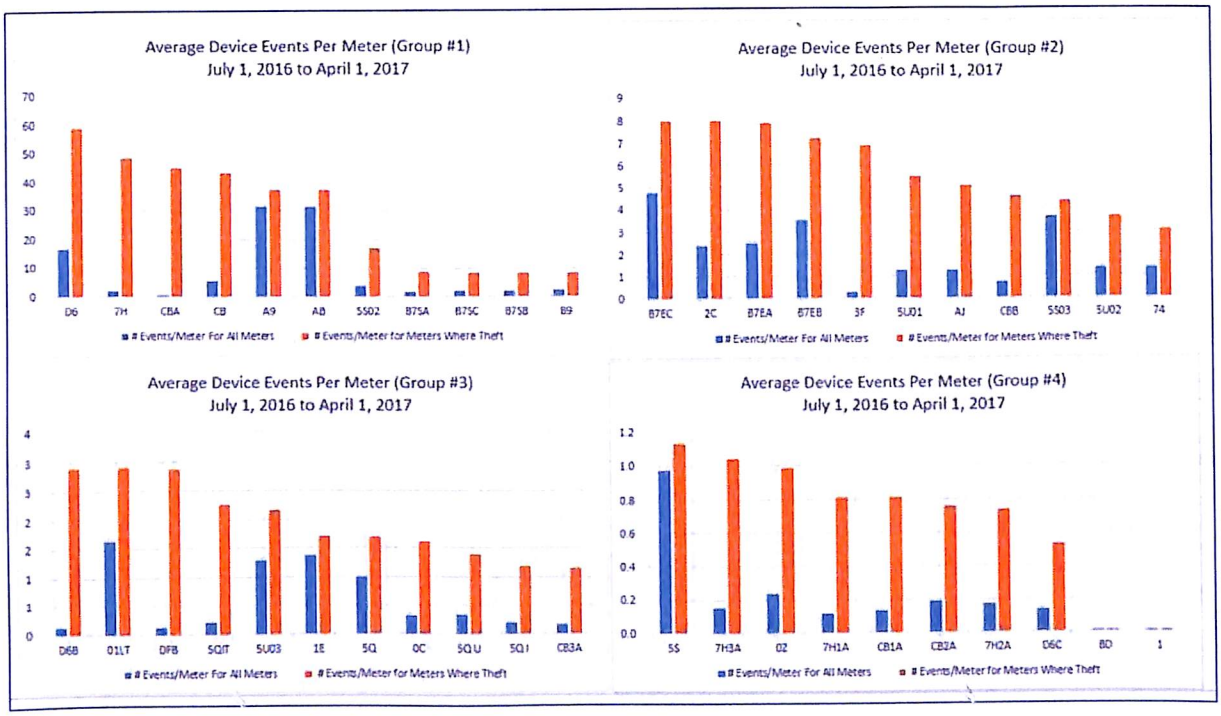


Figure 15 : Correlation of Meter Events

TNB uses its proprietary logics to assign a weight to different device events so that the cases can be prioritized and have a short turn around time.

	Device Event Type	Description	Weight
23	7H3A	Reverse active energy phase C Cleared	0.54
24	7H1A	Reverse active energy phase A Cleared	0.52
25	B7EB	Current Missing on Phase B Stop	0.48
26	5Q I	Zero Sequence I	0.48
27	CB1A	Reverse active energy phase A	0.46
28	0C	Watchdog Timer Error	0.43
29	B7EC	Current Missing on Phase C Stop	0.40
30	5U02	Voltage Cut Phase 2	0.37
31	5Q U	Zero Sequence U	0.35
32	0Z	Phase Sequence Error	0.32
33	7H2A	Reverse active energy phase B Cleared	0.31
34	CB2A	Reverse active energy phase B	0.28
35	74	Unrecognized Recorder Event Code	0.28
36	D6C	Phase Failure on Phase C	0.26
37	01LT	Battery Failure - Latched	0.20
38	5U03	Voltage Cut Phase 3	0.14
39	5Q	Zero Sequence U	0.13
40	5S03	Current Reversal Phase 3	0.09
41	1E	Attempted Access With Invalid Password	0.05
42	5S	Current Reversal 01	0.03
43	1	Battery Failure	0.02

Figure 16 : Weight assigned to different events

Load reduction logics in TNB

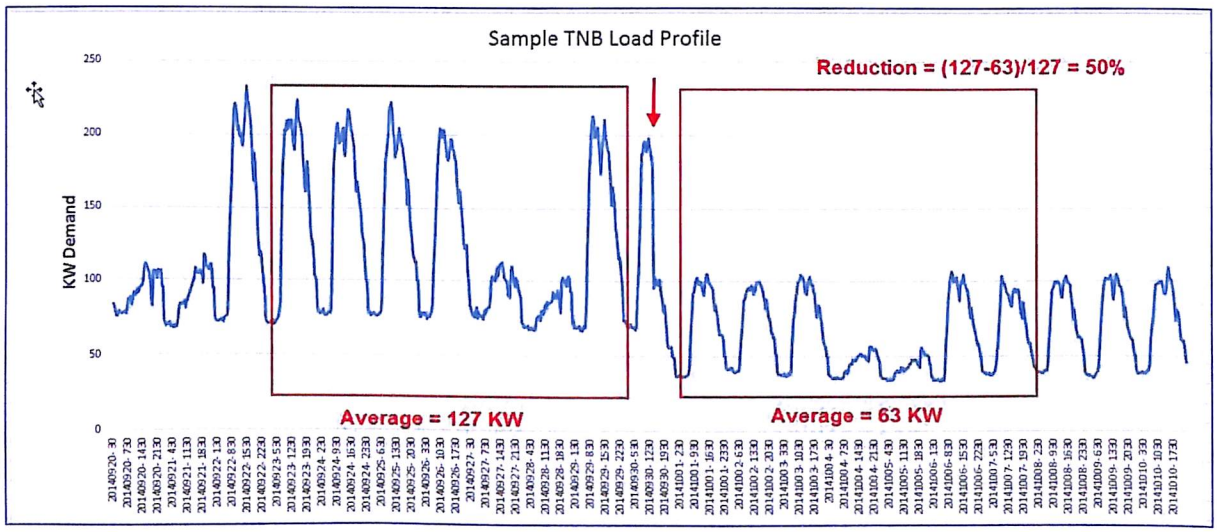


Figure 17 : 30 Day Load Reduction Graph

Load reduction for 30 day before and 7 day after

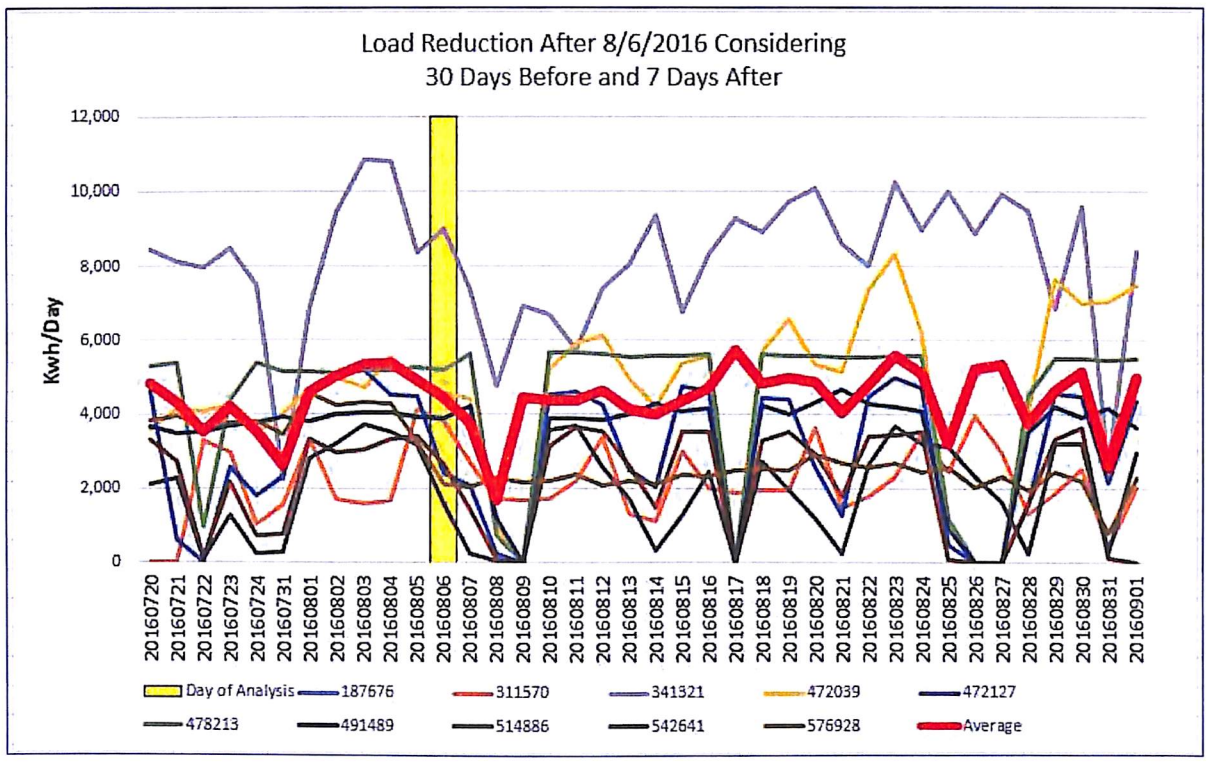


Figure 18 : Load reduction graph

Load reduction for 30 day before and 30 day after

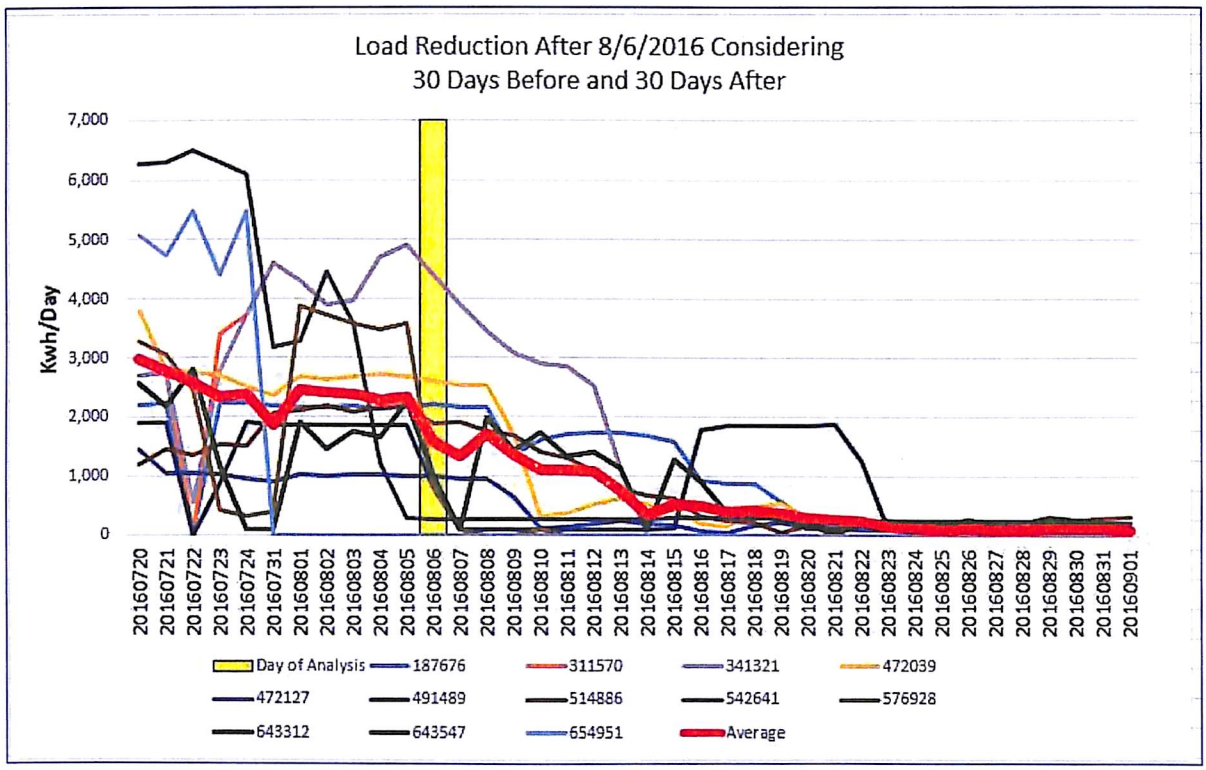


Figure 19 : Load reduction

Steps to identify new theft cases:

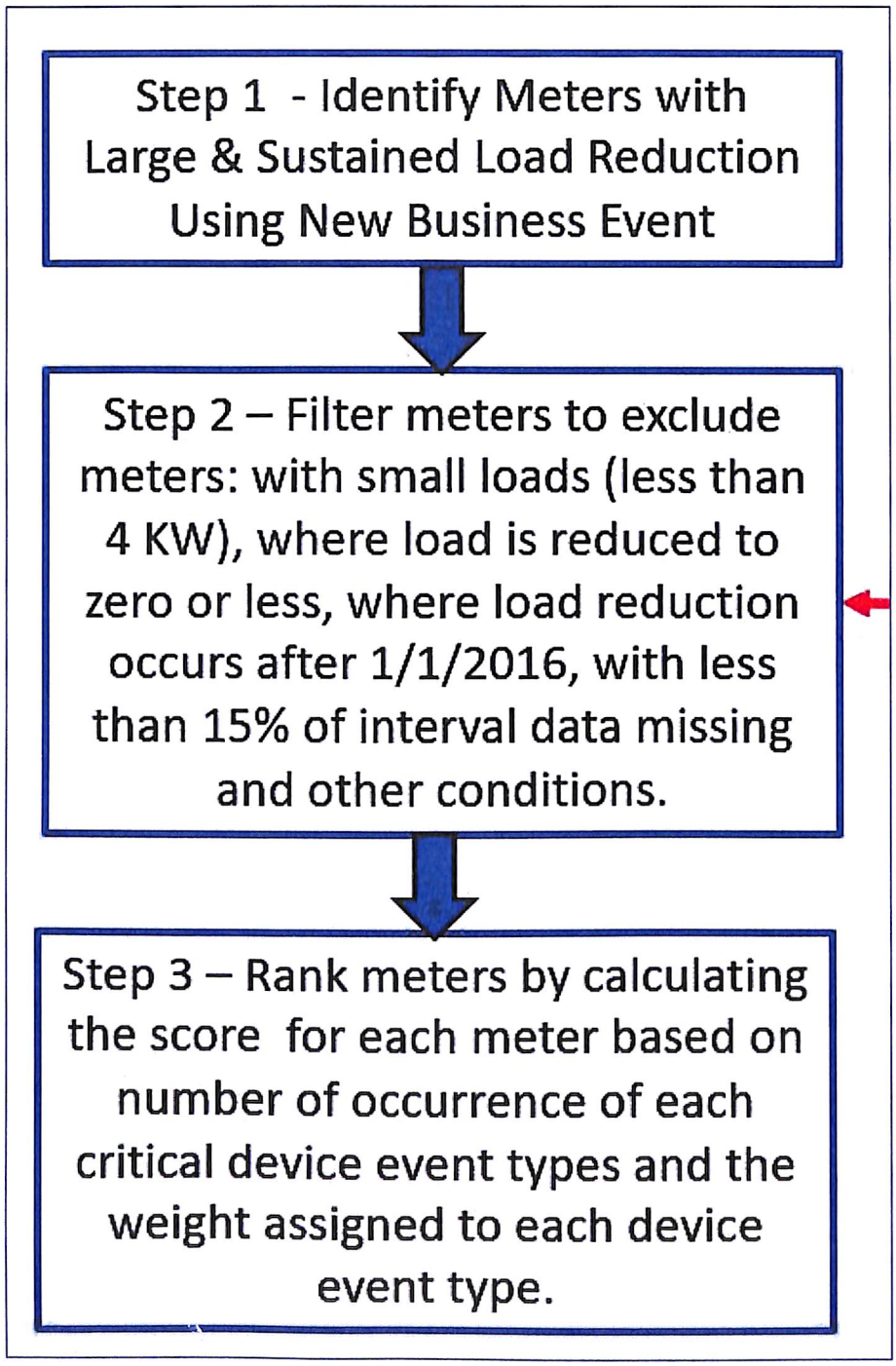


Figure 20 : Flowchart for Theft identification

Theft case 1

TNB finds theft cases through consumption profiles of customers:

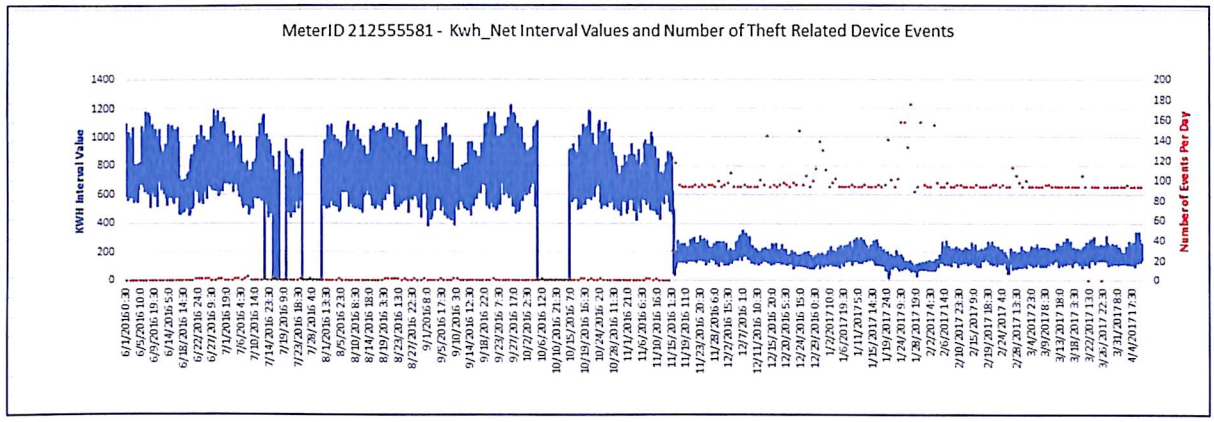


Figure 21 : Theft Case 1

Theft Case 2

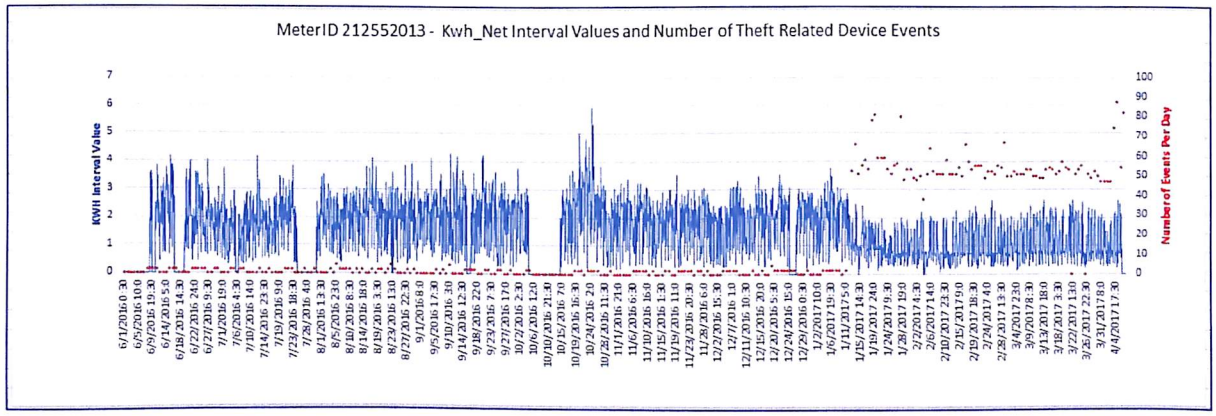


Figure 22 : Theft Case 2

Theft case 3

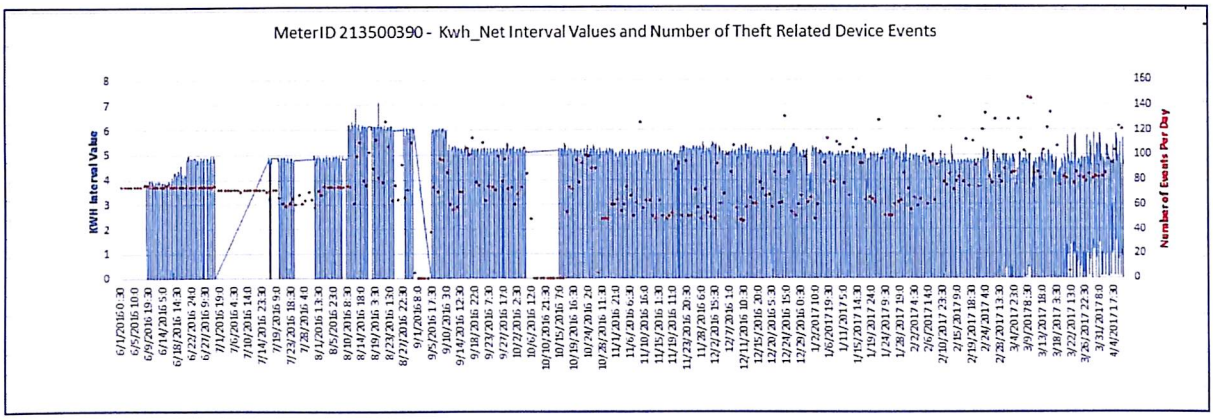


Figure 23 : Theft Case 3

Theft Case 4

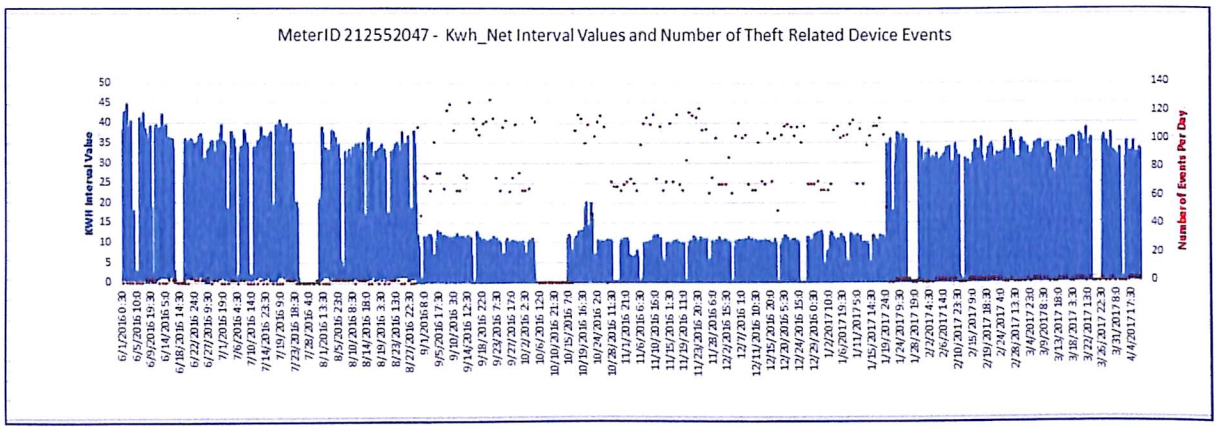


Figure 24 : Theft Case 4

Excluded case for unavailability of data

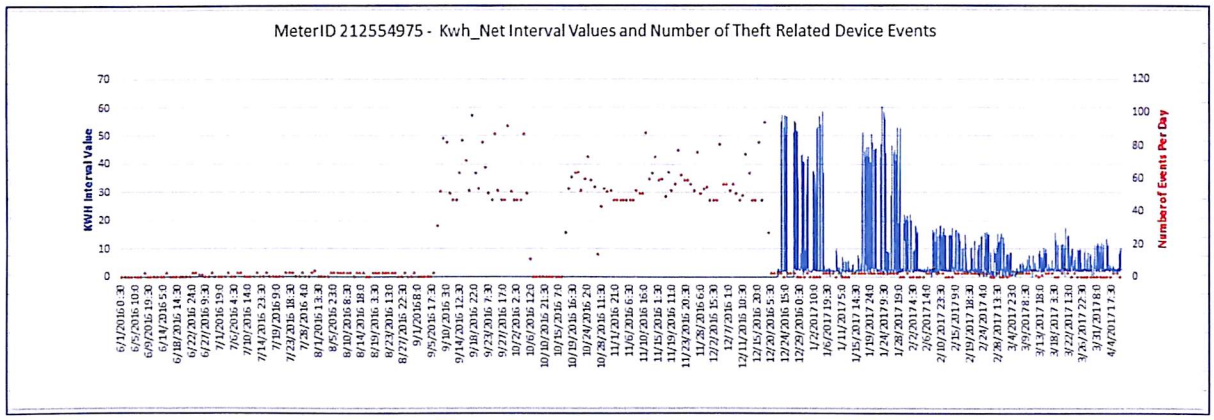


Figure 25 : Probable theft case

Analytics application, TNB uses for revenue protection

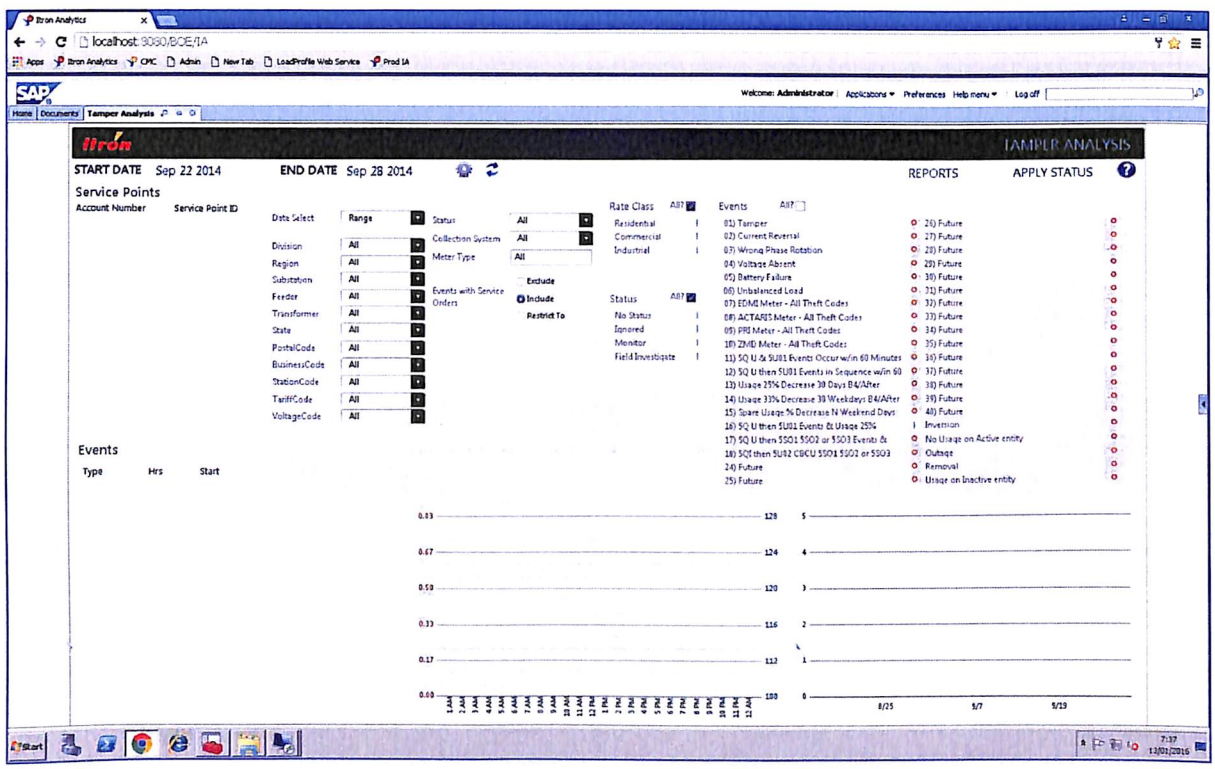


Figure 26 : Analytics application

Moving Average Analysis used by TNB

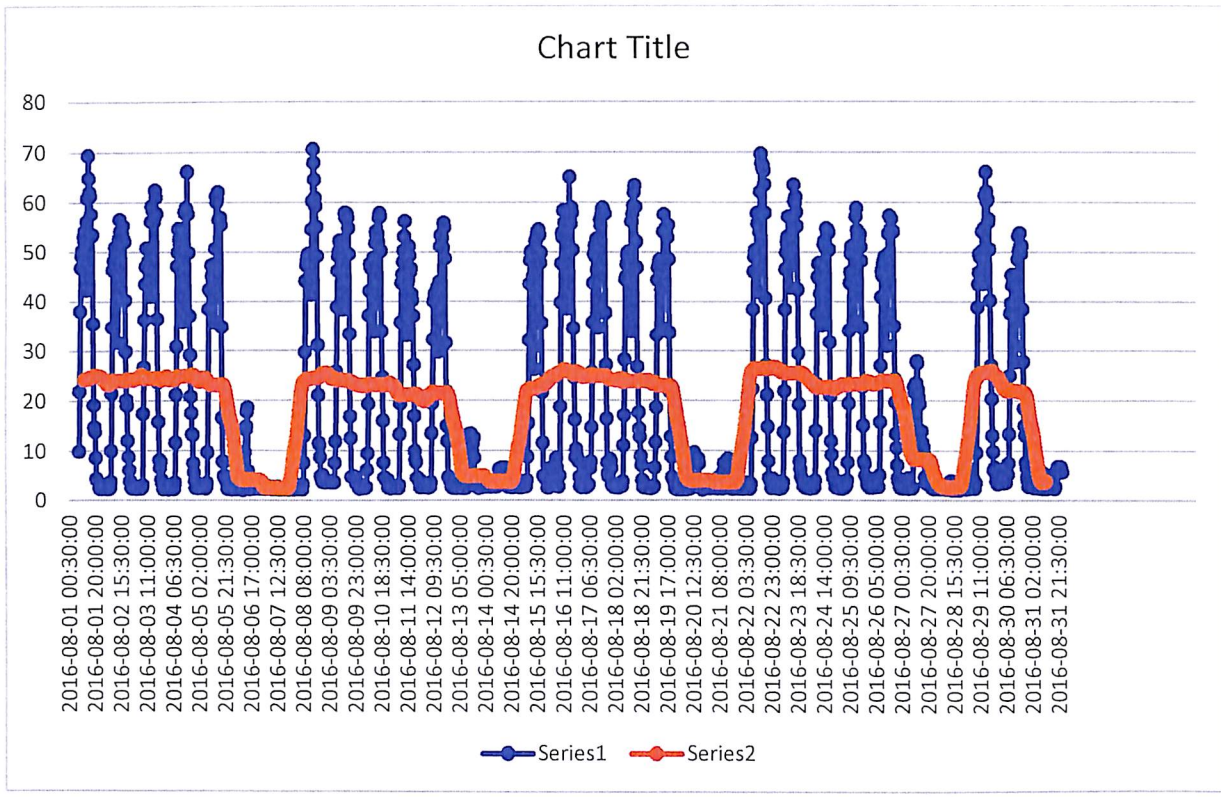


Figure 27 : Moving Avg Analysis

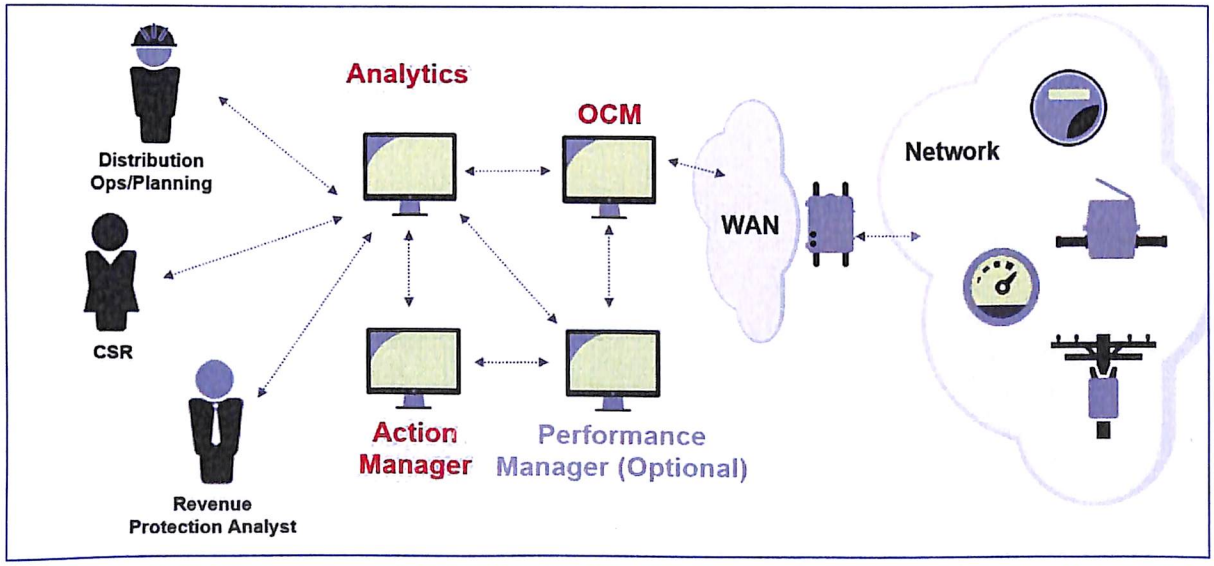


Figure 28 : Data collection Process at TNB

TNB uses network hierarchy to identify first level of theft by comparing consumption on distribution side and load side.

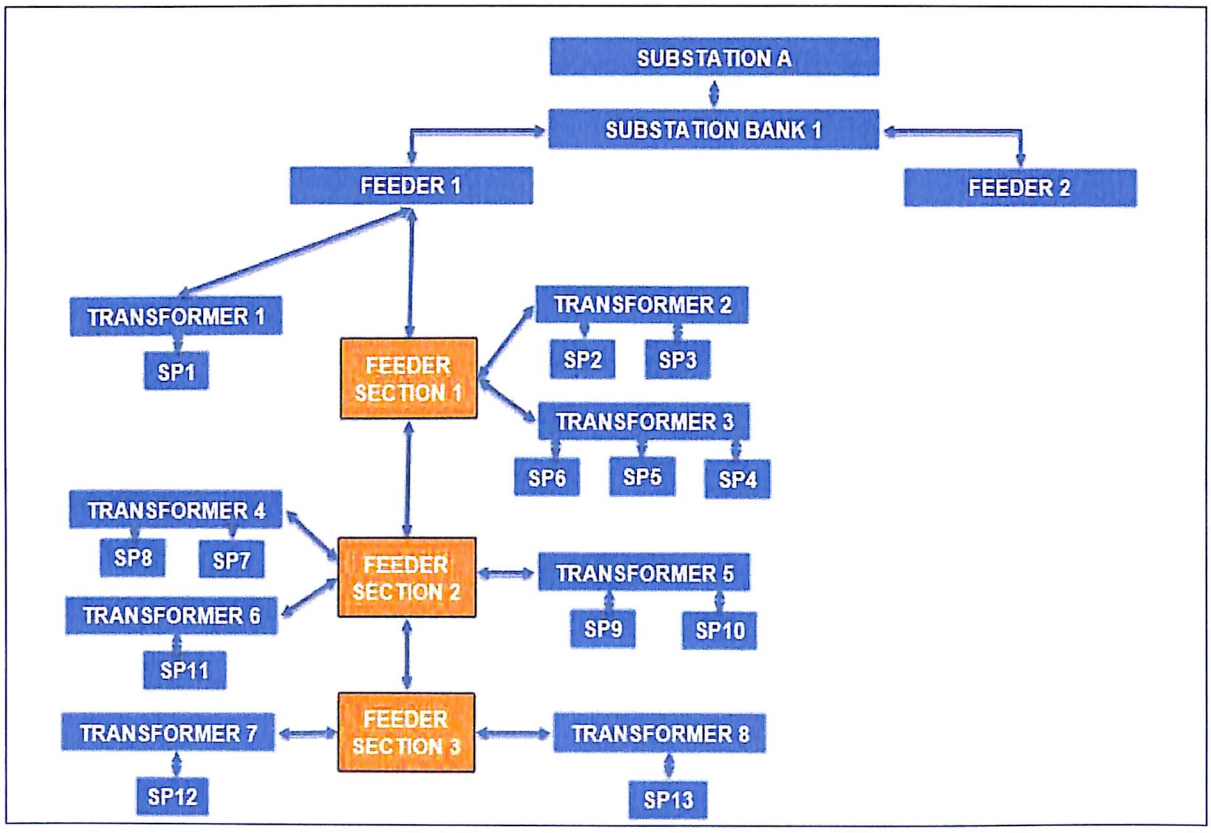


Figure 29 : Sample Network Hierarchy

Figure 30 : KPIs in Load analysis

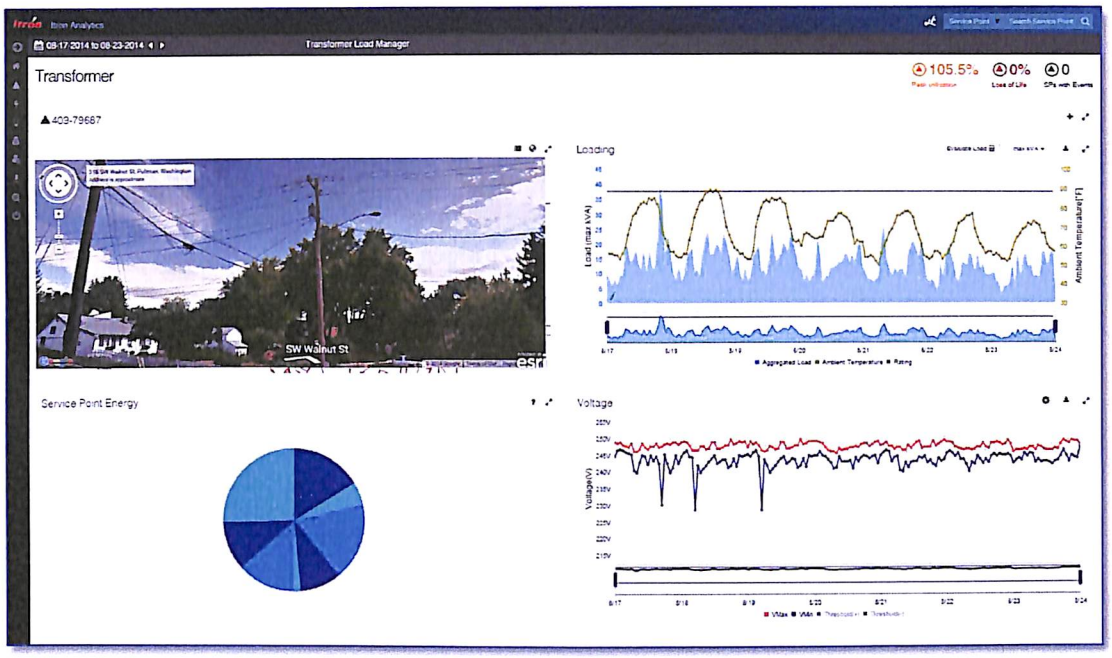


Figure 31 : Application used for TLM

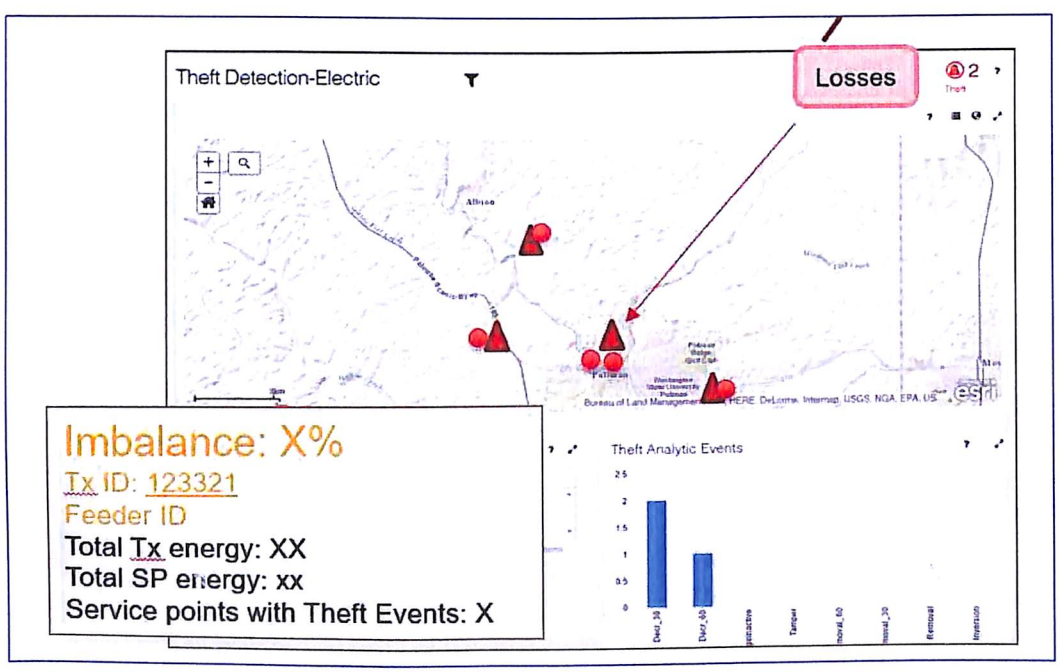


Figure 32 : Loss identification

Chapter 5

5. Interpretation of Results

5.1. Interpretation of Results

The data is collected in two forms:

Primary Data Collection: The methods of primary data collection were in form of survey, focused and unstructured interview. The researcher made use of various forums, seminars and workshop for conducting interviews and data from the Discoms officials.

Secondary Data Collection: The secondary data collection mainly comprised of regulatory and audit reports, information from Discoms websites etc.

SL #	Factors Considered	Details	Total No.
1	Utility Customers Survey	Number of Customers Surveyed	33
2	Utility reports	Number of reports	7
3	Energy Audit Reports	Number of reports	5
4	Technology Refresh Reports	Number of reports	2
5	Press releases	Number of press releases	3

Factors and scores of Utilities surveyed

Sl	Category	Technologies	Score	Weights	Delhi Discom	Haryana Discom	Delhi Discom	Haryana Discom
1	Metering	Installation of Electronic Meters -LT Consumers	5	0.25	4	4	1	1
		AMR for Feeder Meters	5		4	3	1	0.75
		Prepaid Metering	5		4	1	1	0.25
2	IT Distribution Apps	Energy Accounting System	5	0.25	5	2	1.25	0.5

		SCADA	5		5	1	1.25	0.25
		Load Forecasting Applications	5		4	1	1	0.25
		Meter Data Management System	5		5	1	1.25	0.25
		Outage Management System	5		5	1	1.25	0.25
		Geographical Information System & Network Analysis	5		5	2	1.25	0.5
		Smart Grid Pilot Project	5		1	0	0.25	0
3	IT Retail Applications	Customer Information System	5	0.3	5	4	1.5	1.2
		Centralized Call Centre	5		5	3	1.5	0.9
		Online Bill payment	5		5	4	1.5	1.2
		E-Complaint	5		4	3	1.2	0.9
		Mobile Apps	5		4	1	1.2	0.3
		Spot Billing	5		5	4	1.5	1.2
4	Enterprise level Apps	ERP Module for Finance	5	0.2	5	1	1	0.2
		ERP Module for HR	5		4	1	0.8	0.2
		ERP Module for HR						
		Performance Management System of employees	5		5	1	1	0.2
		ERP for Assets Management	5		5	1	1	0.2
			100	1	89	39	22.7	10.5

After evaluation of technologies being used by power distribution companies using power distribution technology index, the researcher tried to find out whether there is any relation between AT&C losses incurred by Discoms and level of technologies being implemented.

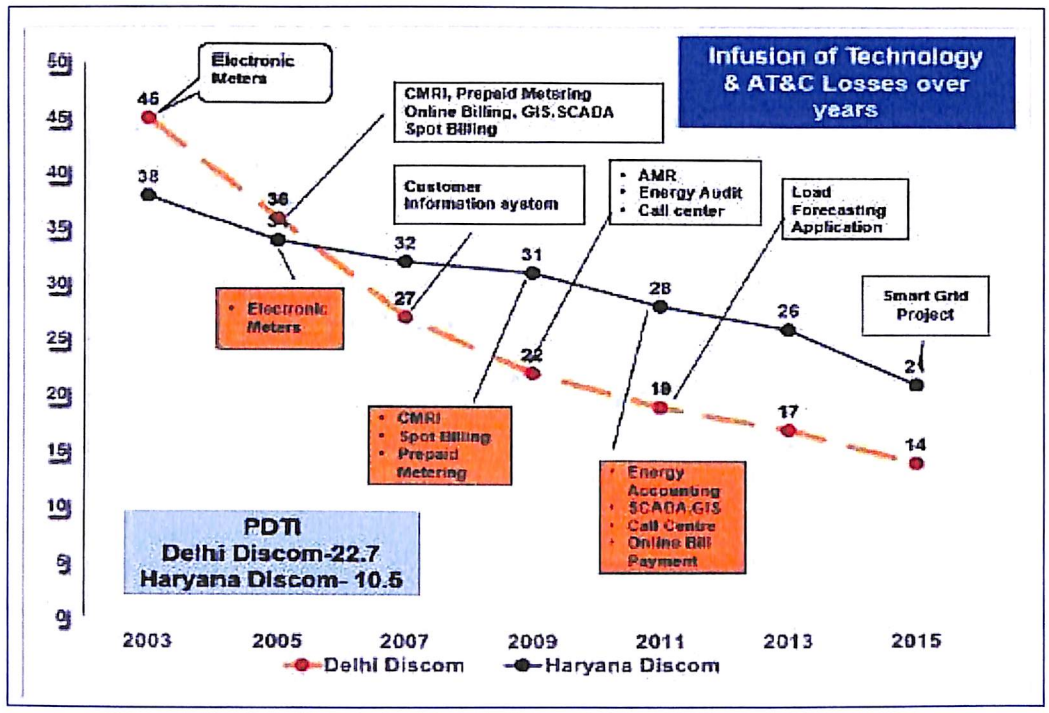


Figure 33 : Technology infusion graph for two Indian utilities

5.2. Comparison of Results with Assumptions (Hypotheses)

Tenaga Nasional Berhad has always a foreforont in adopting new technologies. The utility is also investing on its resources to learn and research new trends in theft reduction methods.

Chapter 6

6. Conclusions and Scope for Future Work

The study has been conducted by collecting both quantitative and qualitative. The extensive literature survey is done before the finalization of the research objectives by identifying the gaps. After research objectives are finalized, literature related to specific objective has been done in order to understand the insights of the earlier research. The research variables are identified for the research construct and data collection. Then questionnaire is framed based on research variables. The data analysis has been done using both qualitative as well quantitative methods. The results are analyzed and based on that inferences are drawn.

COMPLETE RESEARCH PROCESS

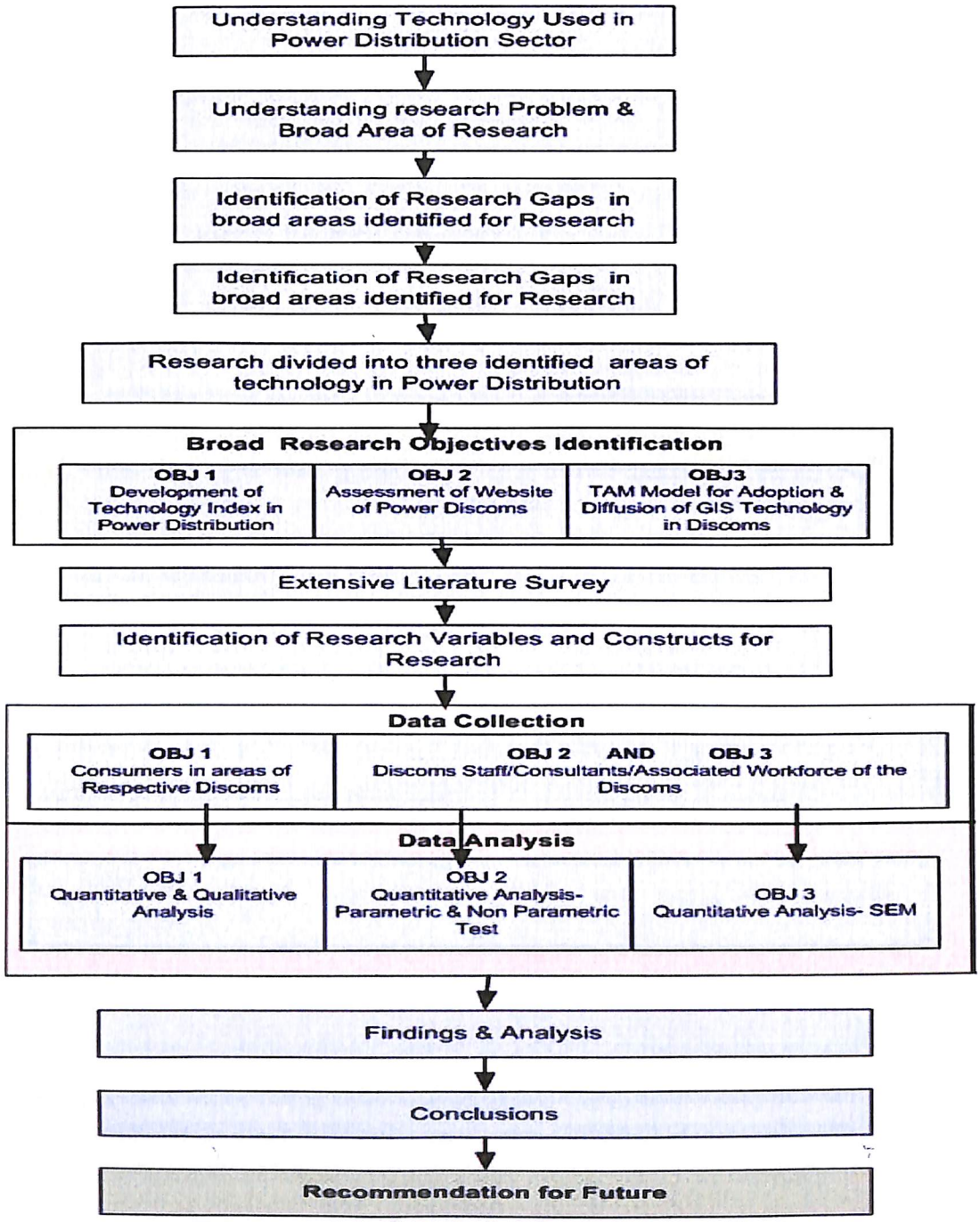


Figure 34 : Study Process

The distribution utilities in India have already implemented some IT and automation technologies and the rest are in the different stages of implementation of these technologies. Based on the data collected through survey and interviews and further analyze the results were derived, there is the difference in the level of technology implementation in power distribution companies. The power distribution technology index is a clear indicator of technology implementation in Discoms which clearly shows PPP Discoms score better than Govt Discoms in National Capital region.

The distribution utilities in Malaysia are way ahead than their peers in other Asia Pacific regions. The utilities are already having state of the Art data centers, DR centers on IT side. The utilities already have Automated Meter Readings implemented and working on Smart metering and Automated Metering Infrastructure. These utilities already started using data mining technology to do con current analysis based on last 7 years of data.

The utilities in Thailand have also moved ahead with Automated Metering Infrastructure and started investing heavily on IT. The utilities are evaluating newer technologies for revenue protection and demand response.

As per power distribution technology index, TNB (22.7) score better than Thailand (10.5), which means technology deployment in Malaysia is much better than Thailand and India Discoms. For Period 2005-2017, AT&C losses for TNB from 9 to 5%, whereas for India Discoms it got reduced from 34 to 21%. Singapore Discoms has implemented IT Apps (86%) for Power Distribution (GIS,SCADA, Analytics etc) whereas India implemented them (27%) .
Metering for all initiative and conversion of mechanical meters to electronic meters brought down AT&C level drastically in 2003-2005, the loss levels from 61 % to 36 %. In Delhi Discoms Implementation of Information Technology was directly responsible for the drastic reduction in AT&C losses. Technology gradually marks acceptance among users over a period of time and reduces AT&C losses. AT&C depends directly on many factors, especially commercial losses and social economic conditions where sometimes technology has to play little role.
A better technology deployment at India Discoms has contributed significantly towards larger reduction in AT&C losses.

7. Bibliography

I. Unyelioglu, K. (2018). *Data Clustering Using Apache Spark - DZone Big Data.* [online] dzone.com. Available at: https://dzone.com/articles/cluster-analysis-using-apache-spark-exploring-colo

II. Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), the handbook of attitudes (pp.173-221).

III. Chaturvedi, K. (2017). The Characteristics Of Big Data Platforms And Their Importance For Data Exploration | Linkedin. [online] Linkedin. Available at: https://www.linkedin.com/pulse/characteristics-big-data-platforms-importance-kamal-chaturvedi/

IV. Journal of Computer Science 9 (6): 763-770, 2013

V. doi:10.3844/jcssp.2013.763.770 Published Online 9 (6) 2013 (http://www.thescipub.com/jcs.toc)

VI. McNish, M., 2001. Guidelines for managing change: A study of their effects on the implementation of new information technology projects in organisations. J. Change Manage., 2: 201-211. DOI: 10.1080/738552754

VII. Chau, P.Y.K. and P.J.H. Hu, 2001. Information technology acceptance by individual professionals: A model comparison approach. Decision Sci., 32:699-719. DOI: 10.1111/j.1540-5915.2001.tb00978.x

VIII. Depuru, S. S. S. R., Wnag, L., Devabhaktuni, V., Green, R. C. "High performance computing for detection of electricity theft." International Journal of Electrical Power & Energy Systems, 47. 21-30.

IX. Méffe, A., Oliveira, C. C. B., Kagan N., Jonathan, S., Caparroz, S., Cavaretti, J. L. "A new method for the computation of technical losses in electrical power distribution systems." in Electricity Distribution, (Amsterdam, Netherlands, 2001), IET

8. Appendix

Sample Survey Questionnaire:

Survey – 1	
Date : Sept 12 2019	Title: Technology Adoption (Metering)

Name :	
Utility :	
Address with Country :	

SL #	Question	Response

Figure 35 : Survey Questionnaire

5	Are you aware of newer technologies like LORA, WISUN?	☐ Yes ☐ No
6	When it comes to technology, what best describes you?	○ I am skeptical of new technologies and use them only when I have to ○ I am usually one of the last people I know to use new technologies ○ I usually use new technologies when most people I know do ○ I like new technologies and use them before most people I know ○ I love new technologies and am among the first to experiment with and use them

Figure 36 : Survey Questionnaire 2

7	How do you feel about this statement: "Technology has made the Utility industry change too fast."	○ Strongly agree ○ Agree to some extent ○ Neither agree nor disagree ○ Disagree to some extent ○ Strongly disagree ○ Don't know
8	How do you learn about new tools and technologies? (Pick all that apply)	☐ From my supervisor ☐ From my company training department ☐ From my company's technology department ☐ From my own research ☐ From social media ☐ From my peers ☐ From friends or family ☐ Other Other (please specify) _____

Figure 37 : Survey Questionnaire 3