Framework to Identify Fast Moving Items in the Service Automation Division of ABC Company

A dissertation submitted for the Degree of Master of Information Technology

N. C. De Jeodth
University of Colombo School of Computing
2017

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: Miss N C De Jeodth

Registration Number: 2014/MIT/011

Index Number: 14550117

_____

Signature: Date:

This is to certify that this thesis is based on the work of

~~Mr.~~/Ms. N. C. De Jeodth

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name:

_____

Signature: Date:

Abstract

Identification of fast moving consumable items of photocopiers is the main research area of this project. Mainly there are four main consumable item types has been discussed in the document. Nowadays new businesses are emerging and the competitiveness in the business industry is very high. It is very important to identify business opportunities in-order to survive in the business industry. Identifying new business opportunities and come up with new strategies for business make the business solid in the industry. New strategies always make the business to keep on track. In-order to come up with new strategies it is essential to identify trends of item moving for that industry.

In-order to identify market trends or item moving patterns in the industry need to analyze the company data. These company data help to make effective decision makings. Decisions that made by a company decides the future of that company.

Day by day data been stacked up in company databases. Large amount of data need to be analyzed properly for make effective decisions like regular customers, customers buying patterns, seasonal trends, fast moving items etc. These decisions evolve with increase company sales. In- order to analysis, large amount of data and complex data need to use techniques which are suitable for that such as, business intelligence (BI) tools, data mining techniques etc.

This research has been carried out for a rental business in-order to identify fast moving items and rental business's business trends and forecast trends in the company. Because of large amount of data, data mining techniques have been used to analyze the data sets. Drill down methods has used to identify variations of number of copies have been taken from each photocopier machine and the impact of customers for the high amount of usage of photocopies. K- means cluster analysis used to identify items with similar characteristics. Hence that, K-means data mining technique has applied to identify fast moving items in the dataset. From the identification of fast moving items and high usages (no of copies) suggest new business strategies and forecast for usage volumes of photocopiers and fast moving items.

Key Words: Pattern Recognition, Clustering, Fast Moving Items, Forecasting, Business Trends, Data Mining

Acknowledgement

Firstly, I would like to thank my supervisor, Dr. H. A. Caldera for taking the initiative of guiding me throughout this project by providing with the necessary information in-order to make this project a success. Without his support, guidance and invaluable supervision, all efforts could have been shortsighted.

I would also like to thank my friends who helped me with various issues during the entire project, and for been around to support me morally. Also, special thanks are extended to the library staff for helping and providing with necessary facilities.

Last but not the least I would like to thank my parents for being right through supporting and encouraging me and made sure I receive everything that helpful to complete this project in a successful manner.

Finally, I am responsible for any errors that remain in this report.

TABLE OF CONTENTS

List of Tables

Table of Figures

List of Abbreviations

CRISP -DM - Cross Industry Standard Process for Data Mining

FM – Fast Moving

MFP – Most Frequent Pattern

FMCG- Fast Moving Consumer Goods

1.0 Chapter 1: Introduction

This chapter gives a general background about the working environment of the gathered dataset. This chapter consists of an introduction to photocopier business, main functions of it and consumable items involve with photocopiers. The motivation factor of reduce the wastage of consumable items and increase the profits in the division by identifying the patterns in photocopier usage and moving pattern of consumable items. Based on identified fast moving items patterns do a demand forecast and suggest strategies to increase the profits in the division. In addition to this objective and the scope of the project and structure of the project include in the chapter.

1.1 Company Background

ABC is a well-known company with a prestige name having more than 100 years of history [1]. ABC Company manages different kind of businesses such as investments, agriculture and plantation, power generation, marine and manufacturing, home and office solutions, pharmaceuticals and healthcare.

This project is based on the business sector of home and office solutions which handle by the Service Automation division. Service Automation division of ABC Company does several leading businesses including photocopier rental business, air conditioner business etc. They have wide range of items sell/rent by them. Under the Service Automation division there is another subdivision called Doclink Rental Business which handles photocopier rental business.

1.1.1 Doclink – Rental Business

Doclink service offers a range of office management solutions which includes facsimiles and photocopies, multimedia projectors and other office management support backed by the superior after- sales service and guarantee provided by ABC and backed by cutting edge technology from pioneer global brands. Within the Doclink division they sell consumable items which are compatible for photocopiers.

Doclink rental business maintains its position as the market leader in the industry, further enhancing its service levels year by year. E solutions offered by ABC company also allowed ABC to further expand into the corporate and financial markets, with the Central Bank being a significant customer for ABC sector continued to develop essential support services such as customer response times and customer care, in addition to development of infrastructure for service expansion.

For the business savvy entrepreneur, they offer an innovative new service option where all the office management needs can be outsourced to ABC Business Solutions so that customers pay for what they get and reduce capital expenditure outflows.

1.2   Motivation

Motivation of this research is to facilitate Doclink Rental Business to find out the exact highly sold consumer items against their rental business. By identifying these consumable items, Doclink can come up with better business strategies where Doclink may be able to increase productivity. This opens a new window to Doclink business where they can market their products based on the seasons, from that they can opt which product need to be marketed in which season and from that reduce the wastages.

1.3   Problem Statement

Doclink Rental Business is the rental business handles by Service Automation division of ABC Company which does photocopier rental business. One of the main problem identified in Doclink rental business is, not having proper idea about how the consumable items sold out monthly. Consumable items available in the Doclink division are, developer, drum, toner and compatible chip. The sales of consumable items are varying monthly, which means when consider about the usage of consumable items there are high amount of usage in certain months and some months do not have high amount of usage. For the division, it is difficult to trace the exact reason for this variance. The division has very large amount of data related to these facts. But with lots of variations, combinations and relations the data has become complexed. Because of that, there is no accurate way to identify a pattern or a reason for this variation.

Hence that there is no proper identification of fast moving items (highly sold out items) in the division which may help to increase the profits in the division and which may help to come up with new business strategies to the division. This may lead the company to lose some of the customers because there are situations which they incapable to satisfy customers' demand due to the reason of lack of identification of most usable/buying items by customers. Due to that some of the consumable items like toner get outdated and unusable and that leads a loss for the division. If the division has a clear idea about this, they could get better decision to promote consumable items in a better way.

1.4 Aims and Objectives

Following are the aims and objectives of this research.

1. The main objective of this project is to, analyze the photocopier usage and customer impact on them and identify fast moving consumable items in the data set. From the found-out data patterns provide justifications for the high amount of usage in photocopiers in certain months.

2. From the identified patterns find out highly sold consumable items (fast moving) and identify seasonal trends of consumable items and provide suggestions to which items to be sold in which season.

3. From identified season patterns provide a justification for demand forecasting of photo copiers.

4. Provide executable innovative business strategies where Doclink may able to increase sales and reduce wastage.

1.5 Scope and Limitations

Data from Doclink Business for the past 5 years (2012-2016) will be used as the dataset for the research. The data set consists of all the consumable items, machines and customers that involved with the rental business. This dataset has duplicated, unprocessed data. Hence that it requires cleansing, formatting before using for the analysis.

The analysis and research will be carried out only for the Doclink Rental Business of Service Automation division of ABC Company. By Analyzing the dataset discover different types of data patterns which can lead to identify seasonal trends of consumable items and from that identify fast moving items in the Doclink rental business. Suggest corrective actions to promote dependent items to make more profits to the division. This project mainly discusses about following issues/points:

i. Seasonal trend analysis of consumable sales

ii. Forecast demand for photo copiers

iii. Identification of fast moving items based on combinational attributes like photocopier usage, customer, customer type, sales month and item type

Cross Industry Standard Process for Data Mining (CRISP -DM) is used as the process model to follow the data mining process of this research. Followings are the steps of CRISP-DM [11].

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

At the primary level of the analysis carried out a basic analysis in-order to gather the knowledge about the business and the value of the business and understand the business processes.

Once the preliminary understanding gathered thoroughly analyze the data set in-order to get familiar with it. By getting familiar with the data set it helps to discover unseen patterns within the data which may help to carry out an effective analysis and to provide vivid justifications for the analyzed data. From this initial level data set analysis helps to identify unwanted data for the research.

For an effective analysis, these unprocessed/unwanted data must be removed. In-order to remove these unprocessed data need to use cleansing and formatting. From this retrieve, rich data which helps find out final output of this research.

Once above steps were completed the next stage of the research is to, perform high level data analysis use of data mining techniques which helps to drill down data and take out unseen hidden data patterns in the data set. The use of data mining techniques and explore data patterns and from the identified patterns come up with seasonal trends for this data patterns and provide justifications for demand forecasting and business strategies.

One of the main limitations of this data set is some of the unprocessed data. There are data which have not maintained/entered properly. Hence that it is needed to remove those data from the analysis. Due to large dataset, for this research has been used only three years of data. Based on these three years of data will be given results and justifications.

1.6 Structure of The Thesis

Chapter 2- Literature Review

This chapter consists of background reading for the research area and its techniques. One of the highlighted part of this is, similar system/research study. The work had done previously by different authors in the same research area helps to set the foundation for the project. Also, provides foundation to the research.

Chapter 3 – Methodology

Chapter 3 discusses about the approach that has been applied in the research.

Chapter 4 – Evaluation and Results

Consisted of the evaluation methods used to find out main objectives of the research.

Chapter 5 – Conclusion and Future work

This chapter provides the overall critical evaluation of the project and its justification, the problems encountered throughout the project, lessons learnt, and the future enhancements that can be done for this project.

References

List of references which are have been used in the document.

2.0   Chapter 2: Literature Review

In the previous chapter, has given a background knowledge about working environment for the derived dataset. Also, has given introduction to the thesis and its structure.

This chapters includes the research area and methods used, in-order to solve identified problem. Since the problem is involved with pattern recognition and forecasting, data mining has been chosen to solve the problem. In this chapter discussed about what is data mining and possible data mining techniques which can be applied to solve identified problems.

Data mining is a technique which is used to analyze complex data sets. The main reason to use data mining techniques is that it helps to discover unseen data patterns within a dataset. In the previous chapter, has explained the complexity of the data set and the problem domain which shows need of data mining for this research.

2.1   Background

Data play a major role in the business industry. All the critical decision making do base on the previous data that has been gathered throughout the years. In-order to take effective decisions or to extract useful information, data must be explored [4]. But one of the crucial situations most of the companies' face is, the lack of knowledge, techniques to explore/analyze data. Because when business get established and matured in the industry, the size of data gets increased. Once the data is increased exploration/analyze gets complex. In-order to analyze these data need high end techniques. If the organization is lack of expertise, valuable information will be hidden inside the data. From that organization, may lose great opportunities from them.

The main objective of this project is to explore such opportunity which has described in above paragraphs within the organization which may help to increase profits. There are mainly two research areas in this project. They are, seasonal trend analysis and forecasting.

Like earlier mentioned this research involves with large amount of data and relationships. Hence that, data mining techniques are the most suitable techniques for analysis. Seasonal trend analysis (pattern recognition) has been done using data mining techniques and forecasting has been done by providing business justifications for the identified patterns. The business justification of the forecasting has been carried out for seasonal trends as well as for the fast-moving items pattern.

Categorization of sales data help to recognize different market trends [4]. Such as dead stock (DS), slow moving (SM) or fast moving (FM) of the sale [4]. From identified patterns, can

make future predictions/forecasting. As per [4] forecasting can be long term or short term while long term forecasting is less accurate.

As depicted in [4] data mining techniques like clustering and associations can be used to find meaningful patterns for future predictions.

2.2 Existing Work

As per [2] data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable and predictive models from large scale data.

As depicted in [5], data mining refers to extracting or "mining" knowledge from large amounts of data. Above mentioned definitions have given clear understanding of what meant by the data mining. Day by day data has been saved in databases. But not all the data have been used to take critical decisions in organizations. The reason behind this was not having a proper technique to analyze these large amounts of data. Data mining is the solution to this problem. Data mining is a technique which use to discover/identify hidden opportunities in and organization. Data mining deals with large data sets. Data mining is a technique helps identify hidden opportunities in an organization or a sector. It is a process which helps to remove unwanted data, disturbance and get the best data set in-order to take certain decisions in an organization. There are different kinds of data mining techniques such as, clustering, market basket analysis, decision trees etc.

Rajeswari & Ramu [8] have used hybrid clustering association in-order to identify frequent patterns in stock data. In this research, Rajeswari & Ramu have used K-means and the Most Frequent Pattern (MFP) algorithms. First has used K-means in-order to identify clusters and after that has used MFP for pattern recognition purpose. Based on the analysis data have been classified as dead stock, fast moving and slow moving.

Cluster and portioning analysis have been carried out for detecting seasonal trends and cluster motion visualization for very high dimensional transactional data [9]. From this research, different trends and patterns has discovered. Also from the analysis, consistent purchases, correlated cluster motions and customer drift has found out.

has found out Research done by Joshi et al [4] to find out use of data mining techniques to increase sales and marketing K-means and MFP algorithms have been used to improve the effectiveness of sales and marketing. Furthermore, they have explained the importance of data and its role in the business industry as discussed in the 2.1 background section of this chapter. Per Joshi et al [4] since portioning techniques are suitable for analyze complex data sets they

are very famous in data mining. But among portioning techniques K-means at the very top in the list because of its simplicity and straightforwardness. Using cluster analysis stock has been divided into dead stock, fast moving and slow moving. Pattern recognition has done based on the identified clusters as in fast moving, slow moving and dead stock.

A case study has been conducted by Aponso et al [10] in-order analyze and predict fast moving consumer goods. This case study has two research areas as in sales forecasting and pattern recognition which has been used to identify fast moving consumable items. Aponso et al [10] have further described about the importance of fast moving consumer goods (FMCG) for the business. Per Aponso et al [10] in the research paper, has explained that in the analysis of finding FMCG may lead to identify patterns in unrelated data as well. Those found out patterns leads to explore hidden opportunities in the business. In above research, main objective had been to identify consumer buying patterns that would be produced in upcoming seasons. Linear regression model has been used to identify the pattern recognition in FMCG. Linear regression model, has used to identify dependent and independent variables. Main focused seasons of Aponso et al research was New Year and Christmas seasons in Sri Lanka. For that independent and dependent variables, have been categorized respectively customer product sales and product sales. By using linear regression model pattern recognition had been conducted for data of past five years separately and past three years' data. By differentiating these data has been found out patterns of FMCG. Above described research is much more like the author's research since author's research's final output is to identify fast moving items and for their demand forecasting.

Anderson & Kotsiopoulos [12] have done a research paper on use of data mining techniques to enhance decision making in retail business. They have addressed the crucial situations faced by retailers with the development of technology and high amount of data collected throughout the years when decision making. Furthermore, Anderson & Kotsiopoulos [12] have discussed how a new strategy can build up with efficient decision making. In this research, as well has been focused on pattern recognition of consumer buying patterns. From that improve the inventory control like reduce wastage. Furthermore Anderson & Kotsiopoulos research has addressed to different data mining techniques and application of it in decision making in retail industry. Following are the techniques that has been discussed in the paper, association rules, classification, clustering, artificial neural networks, decision trees, genetic algorithm, nearest neighbor method, rule induction and data visualization. Finally, it has been discussed how these techniques can be used for the prediction.

By using K-means and most frequent pattern (MFP) Walinijikar, Shah and Maskeen [13] have done a business support system. Like earlier mentioned this research paper also has discussed the importance of data mining to a business. This system has been built up using a hybrid model which has both K-means and MFP. This research has been conducted in a way, where at the beginning has used flat files and online transaction processing in-order to obtain customer data and combine it into dataset. After that K-means algorithm has been used to divide dataset into clusters. These clusters have been defined based on their attributes. After that each cluster has formed slow moving, fast moving and dead stock. Slow moving items are the ones have less sales and demand, fast moving items have high sales and demand and dead stock are the items which are dormant. From the identified three clusters, has identified most frequent item sets in the system.

Rajagopal [14] has used clustering in-order to identify most suitable customer segment for a business. This research has been carried out based on existing customer cluster/ segmentation into methodology-oriented and application-oriented approaches. Further it has been covered on different clustering methods and use of those. Such as partitioned clustering which is highly recommend for large datasets because of its relatively low computational requirement. Likewise, has discussed about behavioral clustering and demographic clustering. In this research, proposed model was a two-phase model. In the first phase involved with collect data and cleansing. Irrelevant, missing, incomplete data and noise are removed. In the second phase generate clusters and profile the cluster to identify the best cluster. As the outcome, has given rank based on the customer count, profit and revenue. After that has suggest business strategies to achieve above exploration.

Pattern recognition of fast moving items is a research area which has been discussed in the topic of data mining. Big data has become a big game changer in most of modern industries because of the competitiveness among industries. In-order to come up with new business strategies, effective decision making, explore new opportunities need to use available data in an effective manner. Data need to analyze to use for decision making or seek for new opportunities. From exploring complex data businesses can improve their productivity. Because of that this research area is very important in the business.

2.3 Photocopier Rental Business

Photocopier rental business evolves with renting photocopiers to different customers for a certain period. Especially small businesses attract to the option of renting photocopiers because it is cost saving for them [15]. But because of the flexibility and easy budgeting even large organizations do rent photocopiers. Because of that there is a good demand for photocopier rental business. There are different types of photocopiers in different brands which cater on different purposes. Such as, analog, digital, mono, color, multifunction and desktop.

2.3.1 Types of Photocopiers

➤ Analog Photocopiers

Analog photocopiers are old types of photocopiers. Currently very few number of machines are sold out from this type. This type of machines use lights, mirrors and lenses to produce the image of the document that has been insert to the platen onto a photo receptor for printing [15]. Now days it is a difficult part to find parts and support for this kind of photocopiers. Since this is a legacy photocopier it is prone to mechanical failures. Hence that these types of machines are not using in most of the businesses.

➤ Digital Photocopiers

This is one of the latest photocopier types. This is a multi-function machine. These types of machines are capable of make photocopies, fax, and scan and print. This has high demand because it fulfills several office functions [15].

Another key feature of this machine is, it has greater paper capacity to make daily works reliable and fast. These machines are user-friendly and fulfill most of the day to day work.

➤ Mono Photocopiers

These types of photocopiers have only one toner color which is black. This is the reason to call this as monochrome. These photocopiers have been designed to cater office day to day work. These types of machines have range of different capacities. Any business can be benefited from these types of machines [15].

➤ Color Photocopiers

These photocopiers cater purpose of color copies as well as black and white copies. These machines also multifunction. This has four drums and 4 cartridges or toner reservoirs. The four primary colors are CYMK or Cyan, Yellow, Magenta and Black [15]. The equipment mixes these colors to create other colors [15].

➢ **Multifunction Photocopiers**

Multifunction photocopies caters different types of purposes. This is not just a basic photocopier machine. These machines known as all in one. Basic functions include copying, scanning, printing and faxing. This machine has gone beyond these basic functions which means it can connect to internet [15]. This can be connected to office network and share documents via email. It also has the feature of scanning features as well. Laptops can be connected via USV or wireless connection. It is also capable of install customized document management software in some models [15].

➢ **Desktop Photocopiers**

These photocopies are small this can be placed on a desk. This serves basic as well as multi purposes. These photocopiers have designed to A4 size and not support for A3 papers. It can accommodate additional paper tray modules allowing the machine to be raised to a normal height usage.

2.3.2 Consumable Items
➢ **Toner**

A powder used in laser printers and photocopiers to form the printed text and images on the paper

Figure 2-1: Toner

Source: Lanka Sale (2016) [16]

➢ **Drum**

Like a rolling stamp, with toner sticking to some parts and not to others before being applied to a page.

Figure 2-2: DRUM

Source: Alphatec (2016) [19]

➢ **Developer**

A powder usually made up of iron fragments. In a copier or laser printer, developer is used in conjunction with toner to print an image.

Figure 2-3: Developer

Source: Olx (2016) [17]

➢ **Compatible Chip**

Compatible chip is a small spare part that used to check how much toner has been used and it gives a warning when toner is over in the photocopier machine.
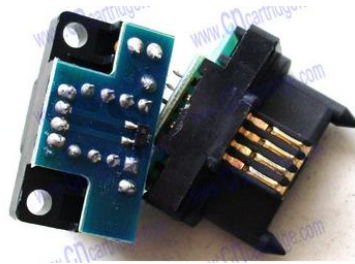
Figure 2-4: Compatible Chip

Source: Ali baba (2016) [18]

In the section 2.3.1 has discussed about different types of photocopier machines. Each photocopier machine has its own capabilities and serve different purposes. In the section of 2.3.2 discussed about consumable items use for photocopier machines. This section has been done to give a brief knowledge about photocopier machines and their consumable items to reader since the research has been carried out for a photocopier rental business and their consumable items.

Summary

Potential of knowledge discovery by fast moving consumable items of photocopier has been highlighted in existing literature. In most of the studies K-means, MFP algorithms have been used to identify fast moving items in a dataset. After the discussion of data mining techniques has done a brief discussion on photocopier rental business, different types of photocopiers and consumable items of photocopier machines.

The next chapter is Methodology. It includes comprehensive discussion on data extraction, data preprocessing, understanding data and data mining techniques use to analyze the dataset.

3.0 Chapter 3: Methodology

Previous chapter consisted of literature review of the research. In the chapter discussed about similar work had done by different researchers. This chapter consisted of four main sub components. They are data extraction, understanding data, design and model which use to analyze the dataset. Under the section of data extraction discuss about techniques use in data extraction and the problems face during the data extraction. The next component is understanding data which is one of the important components in this chapter. It is the foundation for the analysis. Data understanding has done use of tables and different graphical representations like bar charts, receiver operating characteristic curves (ROC). In the section of design, discuss about steps use to preprocess data, techniques (data mining) which can apply to analyze data. The last section of this chapter is modeling. Under the section of modeling the approach use to analyze data.

3.1 Data Extraction

Data extraction is one of the most difficult parts in data mining projects. Because the data replacement is not easy to understand because of its complexity. Sometimes data comes as text files, sometimes it is difficult build up relations between data. Due to these kind of reasons data extraction has been difficult in data mining.

In this research data, has been extracted from a large database. Since the size of the dataset is too large to maintain, only three years (2014-2016) has been extracted to carry on the research. Data extraction has been done using SQL queries.

Data extraction had mainly two stages such as, pattern recognition of usage volume in photocopiers and fast moving consumable items related to photocopiers. Data have been stored in SQL database. Data extraction had been done as followed,

1. Pattern recognition for usage volume in photocopiers

Table 3-1:Pattern Recognition for Usage Volume in Photocopiers

Table	Description
MSM_SvcCallTable	This table consisted of usage volume count, start date and end date for the period and customer details
MSM_ServiceObjectTable	This table directly link with the MSM_SvcCallTable with the machine id (photocopier machine). This table consisted of photocopier machine and the machine type, start date and end date of usage of photocopiers

```
SELECT
MSM_SvcCallTable.custAccount,MSM_SvcCallTable.custname,
MSM_SVCCALLTABLE.USAGEVOLUME,MSM_SVCCALLTABLE.SERVICEOBJECTID,MSM_SVCCALLTA
BLE.SVCCALLID,MONTH(MSM_SvcCallTable.CREATEDDATETIME)as
CallMonth,MSM_SERVICEOBJECTTABLE.STARTUSEDATE,MSM_SERVICEOBJECTTABLE.ENDUSE
DATE,MONTH(MSM_SERVICEOBJECTTABLE.STARTUSEDATE)as ServiceStartUse
FROM
MSM_SVcCallTable
INNER JOIN
MSM_SERVICEOBJECTTABLE
onMSM_SVCCALLTABLE.serviceobjectid=MSM_SERVICEOBJECTTABLE.SERVICEOBJECTID
and MSM_SVCCALLTABLE.DATAAREAID = '206' and
MSM_SERVICEOBJECTTABLE.DATAAREAID = '206'
ORDER BY MSM_SvcCallTable.custAccount
```

2. Pattern recognition of fast moving items

Table 3-2: Pattern Recognition of Fast Moving Items

Table	Description
InventTable	This is the table consisted of all item related data. There are certain subdivisions which belongs only for consumable items for photocopiers. Those divisions are the ones showed below as dimension (206DOC and 206SER) and those consumable items belong to certain product groups which are showed as dimension2 (206CNM).

- Data from–Sales related data had been captured in SalesLine Table which connects with InventTable with item id.

```
SELECT
InventTable.Dimension,InventTable.DIMENSION2_,InventTable.DATAAREAID,SALESL
INE.SALESID,* from InventTable
INNER JOIN SalesLine
ON (InventTable.Dimension = '206DOC' or InventTable.Dimension = '206SER')
and InventTable.DIMENSION2_ = '206CNM' and InventTable.DATAAREAID = '206'
and
SALESLINE.ITEMID = InventTable.ITEMID and SALESLINE.DATAAREAID = '206'
```

- Connect with MSM_SvcCallTable and MSM_TaskTable to get the relation to sales with SalesLine Table

```
SELECT
MSM_SvcCallTable.custAccount,MSM_SvcCallTable.custname,InventTable.Dimensio
n,InventTable.DIMENSION2_,InventTable.DATAAREAID,SALESLINE.SALESID,SALESLIN
```

15

```
E.ITEMID,MSM_SVCCALLTABLE.USAGEVOLUME,MSM_SVCCALLTABLE.SERVICEOBJECTID,MSM_
SVCCALLTABLE.SVCCALLID,SALESLINE.SALESQTY
FROM InventTable
INNER JOIN SalesLine
on (InventTable.Dimension = '206DOC' or InventTable.Dimension = '206SER')
and InventTable.DIMENSION2_ = '206CNM' and InventTable.DATAAREAID = '206'
and
SALESLINE.ITEMID = InventTable.ITEMID and SALESLINE.DATAAREAID = '206'
inner join MSM_TaskTable
on SALESLINE.SALESID =  MSM_TaskTable.SALESID and SALESLINE.DATAAREAID =
'206' and MSM_TaskTable.DATAAREAID = '206'
inner join MSM_SVCCALLTABLE
on MSM_TaskTable.SVCCALLID =MSM_SVCCALLTABLE.SVCCALLID and
MSM_TASKTABLE.DATAAREAID = '206' and MSM_SVCCALLTABLE.DATAAREAID = '206'
order by MSM_SvcCallTable.custAccount
```

- Connect with MSM_ServiceObjectTable to get the relation with photocopier machine

```
SELECT
MSM_SvcCallTable.custAccount,MSM_SvcCallTable.custname,InventTable.Dimensio
n,InventTable.DIMENSION2_,InventTable.DATAAREAID,SALESLINE.SALESID,SALESLIN
E.ITEMID,MSM_SVCCALLTABLE.USAGEVOLUME,MSM_SVCCALLTABLE.SERVICEOBJECTID,MSM_
SVCCALLTABLE.SVCCALLID,SALESLINE.SALESQTY,MONTH(SALESLINE.CREATEDDATETIME)
AS
salesMonth,SALESLINE.CREATEDDATETIME,MSM_SvcCallTable.CREATEDDATETIME,
MONTH(MSM_SvcCallTable.CREATEDDATETIME)
AS
CallMonth,MSM_SERVICEOBJECTTABLE.STARTUSEDATE,MSM_SERVICEOBJECTTABLE.ENDUSE
DATE,MONTH(MSM_SERVICEOBJECTTABLE.STARTUSEDATE)as
ServiceStartUse,MONTH(MSM_SERVICEOBJECTTABLE.ENDUSEDATE) as ServiceEndUse
FROM
InventTable
INNER JOIN SalesLine
on (InventTable.Dimension = '206DOC' or InventTable.Dimension = '206SER')
and InventTable.DIMENSION2_ = '206CNM' and InventTable.DATAAREAID = '206'
and
SALESLINE.ITEMID = InventTable.ITEMID and SALESLINE.DATAAREAID = '206'
inner join MSM_TaskTable
on SALESLINE.SALESID = MSM_TaskTable.SALESID and SALESLINE.DATAAREAID =
'206' and MSM_TaskTable.DATAAREAID = '206'
inner join MSM_SVCCALLTABLE
```

```
on MSM_TaskTable.SVCCALLID = MSM_SVCCALLTABLE.SVCCALLID and
MSM_TASKTABLE.DATAAREAID = '206' and MSM_SVCCALLTABLE.DATAAREAID = '206'
inner join MSM_SERVICEOBJECTTABLE
on MSM_SVCCALLTABLE.serviceobjectid =
MSM_SERVICEOBJECTTABLE.SERVICEOBJECTID and MSM_SVCCALLTABLE.DATAAREAID =
'206' and MSM_SERVICEOBJECTTABLE.DATAAREAID = '206'
order by MSM_SvcCallTable.custAccount
```

One of the problem faced in data extraction was human errors in data entering. There were data with poor data validations. The invalid/incorrect had removed from the dataset. There are certain data with different formats (ex: date format) which showed invalid data. Such there many other unwanted data were in the data set. When analyzing data needed to take necessary actions in-order to have a clear dataset for an effective research output.

Another problem faced during data extraction the size of the dataset was very big. Since that had to follow compression techniques to get the dataset out from the database.

3.2 Data Understanding

Excel sheets used to analyze data in the initial level. By analyzing excel sheets able to explore hidden patterns and variables that can be used to identify pattern recognitions which needs for this research. Further analysis of data had been carried out as followed.

Below Table 3-3 consisted of consumable items available in Doclink division. There are mainly four types of consumable items for photocopier machines. They are, chips, toners, developers and drums. In the section of 2.3.2, has given a brief explanation about the consumable items available for photocopier machines.

Table 3-3: Consumable Item Types Available in Doclink Division

CONSUMABLE ITEMID	ITEM NAME	ITEM DESCRIPTION
AR310ST-CHIP	COMPATIBLE CHIP	Spare part (there are different types of compatible chips which differs based on photocopier machine, size etc.)
OAT-MX312AT	TONER	Spare part
DOC T4AL	DOC Toner for Sharp Copier	Spare part
MX235AT-CHIP	COMPATIBLE CHIP	Spare part
OAT-MX235AT	COMPATIBLE CHIP	Spare part
OAT-MX237AT	TONER	A powder used in laser printers and photocopiers to form the printed text and images on the paper

AR020ST-CHIP	COMPATIBLE CHIP	Spare part
MX237AT-CHIP	COMPATIBAL CHIP	Spare part
OAT-AR020ST	TONER	A powder used in laser printers and photocopiers to form the printed text and images on the paper
MX312AT-CHIP	COMPATIBLE CHIP	Spare part
OAT-MX235AV	DEVELOPER	A powder usually made up of iron fragments. In a copier or laser printer, developer is used in conjunction with toner to print an image.
OAT-MX312AV	DEVELOPER	A powder usually made up of iron fragments. In a copier or laser printer, developer is used in conjunction with toner to print an image.
OAT-MX312AR	DRUM	Like a rolling stamp, with toner sticking to some parts and not to others before being applied to a page.
OAT-AR202DR	DRUM	Like a rolling stamp, with toner sticking to some parts and not to others before being applied to a page.

Below Table 3-4 consisted of different toner types available in Doclink division. There are mainly five types of toners. These toners vary based on the number of photocopies that can be printed.

Table 3-4: Different Toner Types and Their Usage

Toner Type	Description
OAT-MX312AT	For sharp photocopiers (max 100,000 photocopies)
DOC T4AL	For sharp photocopiers (more than 100,000 photocopies)
OAT-MX237AT	For shar photocopiers (max 10,000 photocopies)
OAT-AR020ST	For shar photocopiers (500 photocopies per month)
OAT-MX235AT	For shar photocopiers (500 photocopies per month)

Table 3-5 comprised of different types of compatible chips available in the Doclink division. Chips are compatible for different types of photocopiers. There are mainly five types of chips which available for photocopier machines.

Table 3-5: Different Chips Types and Their Usage

Chip Type	Description
AR310ST-CHIP	Compatible for R256L/316L/M258/M318//M237/M277/5625/5127 photocopier machines
MX235AT-CHIP	Compatible for AR5623/5620/5618 photocopier machines
AR020ST-CHIP	Compatible for AR5617/ 5620 /5623 photocopier machines
MX237AT-CHIP	Compatible for AR5516/5520 photocopier machines
MX312AT-CHIP	Compatible for MX237 /MX-237/FT MX-237/GT M photocopier machines

Different types of developers which are available for photocopier machines are shown in Table 3-6. Mainly two developer types available in the division. These developers are only compatible for Sharp photocopier machines. Compatibility of the developer based on the type of the photocopier machine.

Table 3-6: Different Developer Types and Their Usage

Developer Type	Description
OAT-MX312AV	ARM550 / ARM620 / M700 (Sharp Developer)
OAT-MX235AV	AR1808/AR2008/AR2308/ AR2035 (Sharp Developer)

Table 3-7 contained with different types of drums available for photocopier machines. Like in developers there are mainly two types of drums which are compatible only for Sharp photocopier machines.

Table 3-7: Different Drum Types and Their Usage

Drum Type	Description
OAT-MX312AR	Suitable for photocopier machines AR5617/ 5620 /5623
OAT-AR202DR	Suitable for photocopier machines AR-162/ AR-163/AR-201/ AR-202

Table 3-8: Item Sales (2012-2016)

Item Id	Sum of SALESQTY
AR020ST-CHIP	112
AR310ST-CHIP	12
DOC T4AL	1775
MX235AT-CHIP	876
MX237AT-CHIP	69
MX312AT-CHIP	380
OAT-AR020ST	54
OAT-AR202DR	2
OAT-MX235AT	677
OAT-MX235AV	8
OAT-MX237AT	106
OAT-MX312AR	2
OAT-MX312AT	96
OAT-MX312AV	3

Above Table 3-8 illustrates sales of consumable items 2012-2016. Highest sales recorded for item DOC T4AL which is a toner. The reason for this is, toner level gets reduced with the number of photocopies that get printed out. The lowest shows for drum. Drum does not get worn out as toners because when the number of photocopies taken by photocopier machines is high then the toner level get reduced in those photocopier machines but drums not like that it just a roll that helps to print images/texts in a printout. Figure 3-1 illustrates the graphical view of Table 3-8.
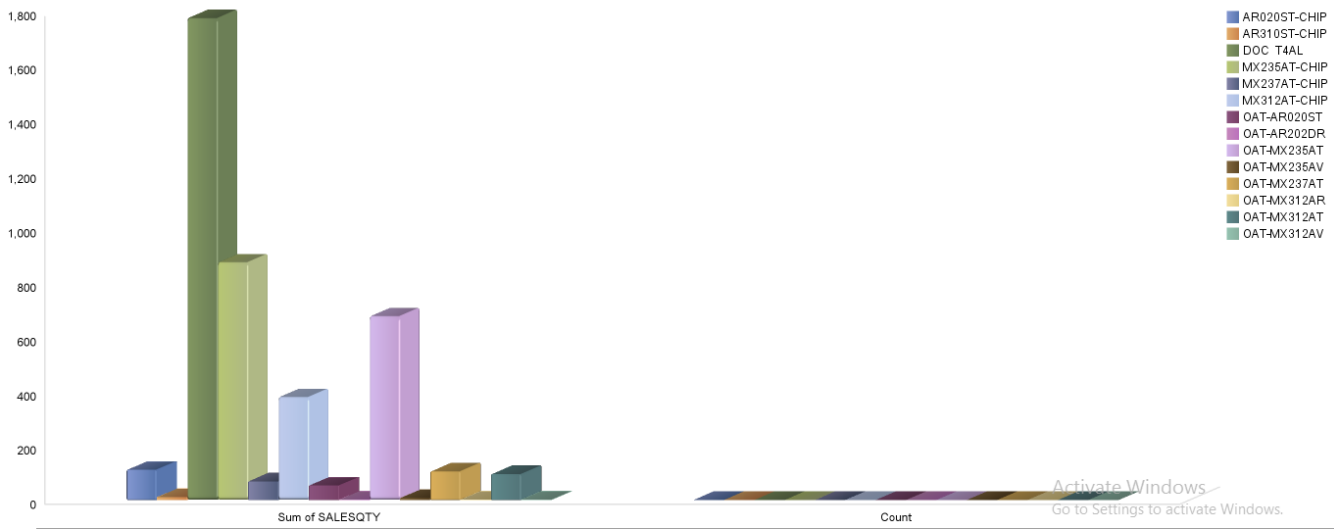
Figure 3-1: Consumable Items Sales (2012-2016)

Above figure 3-1 illustrates how the items have been moved for past three years. From this can have a basic idea of how the item has moved. There are 13 different unique consumable item types available. From above figure it is easy to get a clear picture of the way items have been moved within three years.

In this section, has taken one item which is DOC T4AL for further analysis in-order to get an idea about the item sales behavior.
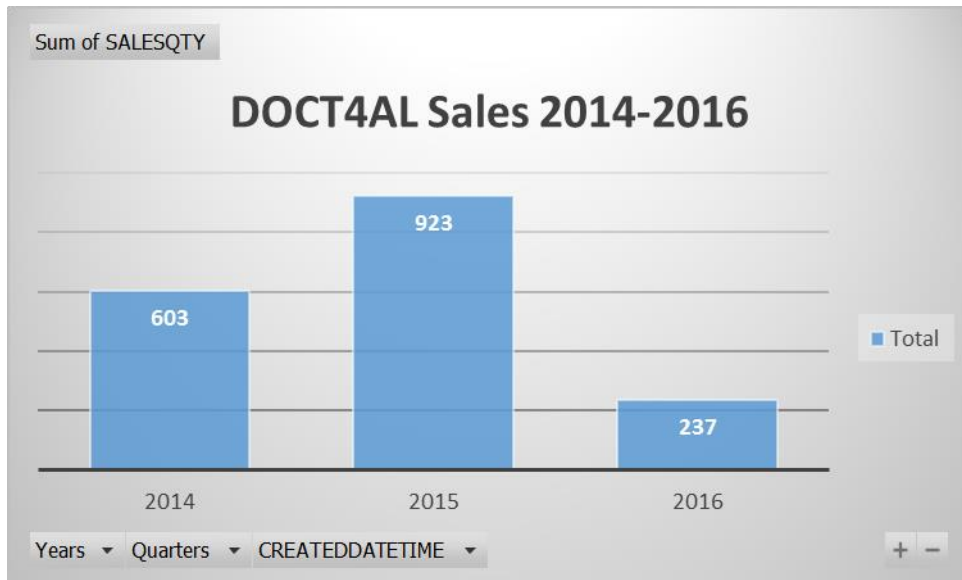

Figure 3-2 DocT4AL Sales 2014-2016 - Annually

Above figure 3-2 emphasizes the item DOC T4AL which recorded highest sales within 3 years. In the year 2015 has recorded the highest sales for DOCT4AL toner type which is a 65% increase when compared to previous year sales.
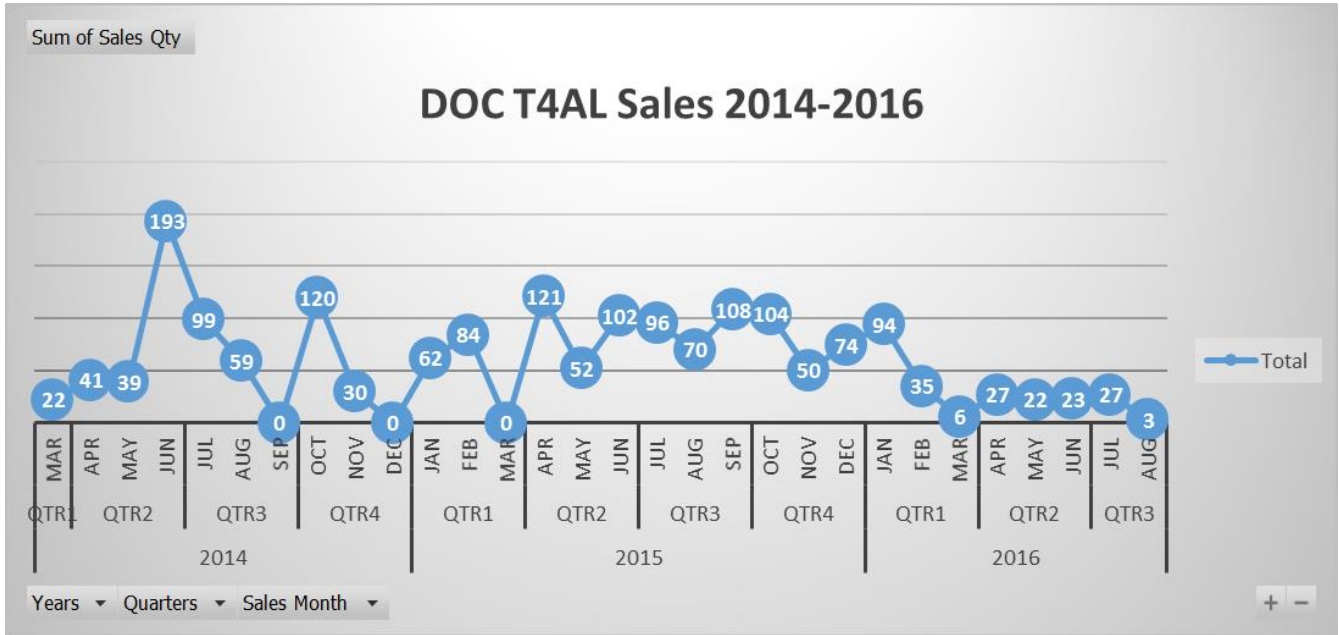
Figure 3-3 DocT4AL Sales 2014-2016 - Monthly

Above Figure 3-3 illustrates the monthly wise behavior of DOC T4AL item. There are certain months where sales have not been. Ex: 2014 September, 2014 December and 2015 March. In 2016 sales, has gone down compared to previous two years.

This illustration has been done for understanding data. The descriptive analysis of this will be included in the exploration part of this research. Analysis should be carried out to each item and with the relation of photocopier machines.

Table 3-9: Photocopier Machines

PHOTOCOPIER MACHINE	NAME
AR5618	COPIER-AR5618N
AR5618N	COPIER-AR5618N
MX-M354N	COPIER-MX-M354N
AR5520D	COPIER-AR5520D
AR5516	COPIER-AR5516
AR5620V	COPIER-AR5620V
AR5316E	COPIER-AR5316E
AR5623N	COPIER-AR5623N
MX-M 354N	COPIER-MX-M 354N
ARM206	COPIER-ARM206
AR5618D	COPIER-AR5618D
AR5520	COPIER-AR5520
AR5516D	COPIER-AR5516D
MX M2010U	COPIER-MX M2010U
AR5731	COPIER-AR5731
MX-M310N	COPIER-MX-M310N
MX-310N	COPIER-MX-M310N
AR5631	COPIER-AR5631
MX-M1810U	COPIER-MX-M1810U
AR5316	COPIER-AR5316
OCEVP3165-SERVICE	COPIER-OCE3165E
ARM420U	COPIER-ARM420U
OCEVP2110-SERVICE	COPIER-OCE VP2110
AR M 206	COPIER-AR M 206
MX-1810U	COPIER-MX-M1810U
OCEVP1075-SERVICE	COPIER-OCE VP1075
MX-M453U	COPIER-MX-M453U
MX M 2010	COPIER-MX M 2010
OCEVP2065-SERVICE	COPIER-OCE VP2065
AR5561	COPIER-AR5561
AR5320	COPIER-AR5320
AR6020N	COPIER AR6020N
MX-2010U	COPIER-MX M 2010
AR6020	COPIER - AR6020
AR5623NV	COPIER AR5623NV
AR6031N	AR6031N COPIER
AR-6023N	AR6023N SHARP COPIER
ARM623N	COPIER-AR5623N
MX-M364N	COPIER MX-M364N

Above Table 3-9 consisted of different types of photocopier machines available in the dataset.
Different customers use different types of photocopier machines. Even though these photocopy
machines are different all the photocopier machines belong to SHARP brand.

Table 3-10: Photocopier Usage Volume based on Machine Type (2012-2016)

	Sum of USAGEVOLUME
⊟ Total of Machine Type	104,814,303
AR5316	452,949
AR5316E	89,359
AR5516	3,980,281
AR5516D	44,934
AR5520	3,022
AR5520D	623,791
AR5618	4,334,295
AR5618D	161,889
AR5618N	2,865,282
AR5620V	1,406,937
AR5623N	417,370
AR5631	960,113
AR5731	969,549
ARM206	820,533
MX M2010U	1,342
MX-310N	573,287
MX-M 354N	5,606,652
MX-M1810U	208
MX-M310N	250,284
OCEVP3165-SERVICE	81,252,226

Table 3-10 consisted of all the available machine types in the dataset. This is to show the amount of photocopies have been recorded for each machine type. Also for the analysis of photocopier usage has only considered black and white photocopier machines and the photocopies which have been done for A4 papers and one side printouts only.

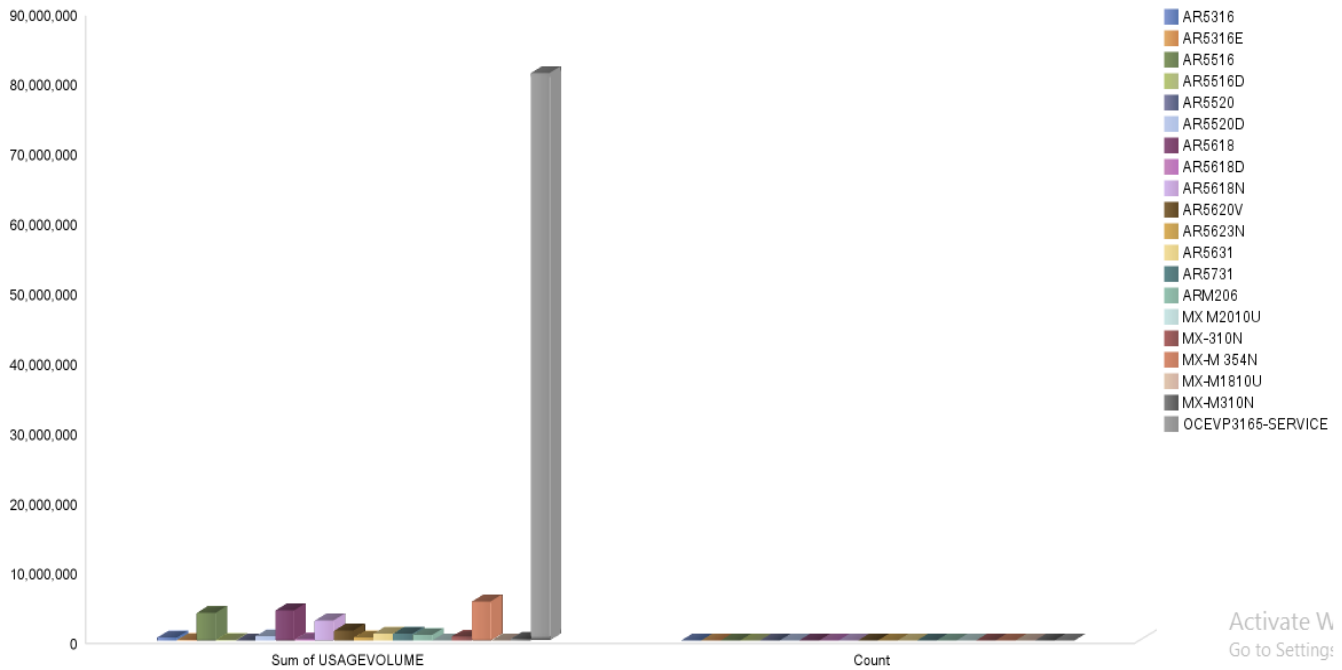The graphical view of table 3-10 shows in figure 3-4.

Figure 3-4: Machine Type Wise Photocopier Usage Volume Analysis

Figure 3-4 illustrates machine type wise usage volume for the year 2012-2016. The highest usage volume has been reported from OCEPV3165-Service item. These machine types breakdown into different photocopier machines. Same machine type can be shared by different customers but photocopier machine (service object id) is unique for a customer.

In-order to get a general understanding of behavior of photocopier usage volume different variables such as customer, consumable item type, customer type, consumable item, month of sales have been used to get the best relation to get the output of pattern recognition of usage volumes.

The next analysis has been carried out for the service object (photocopier machine) vs usage volume to analyze the number of photocopies had took from each photocopier machine.
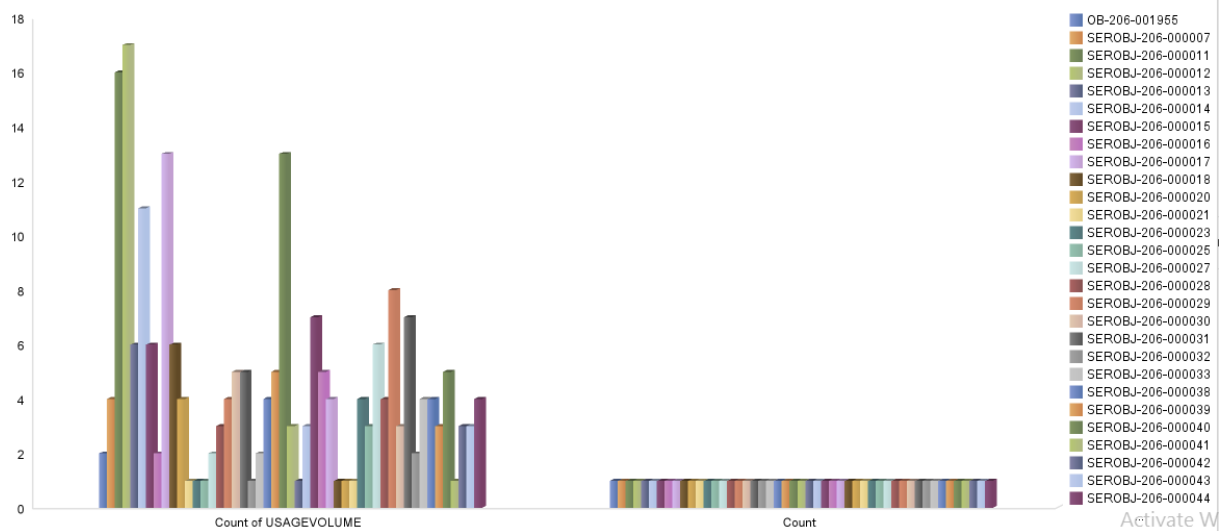
25

Figure 3-5: Photocopier Machine Wise Photocopier Usage Volume

Figure 3-5 illustrates photocopier usage volume behavior against photocopier machines. This is one of the key variable that use to do pattern recognition of usage volume. Service object id 12 recorded the highest usage volume.
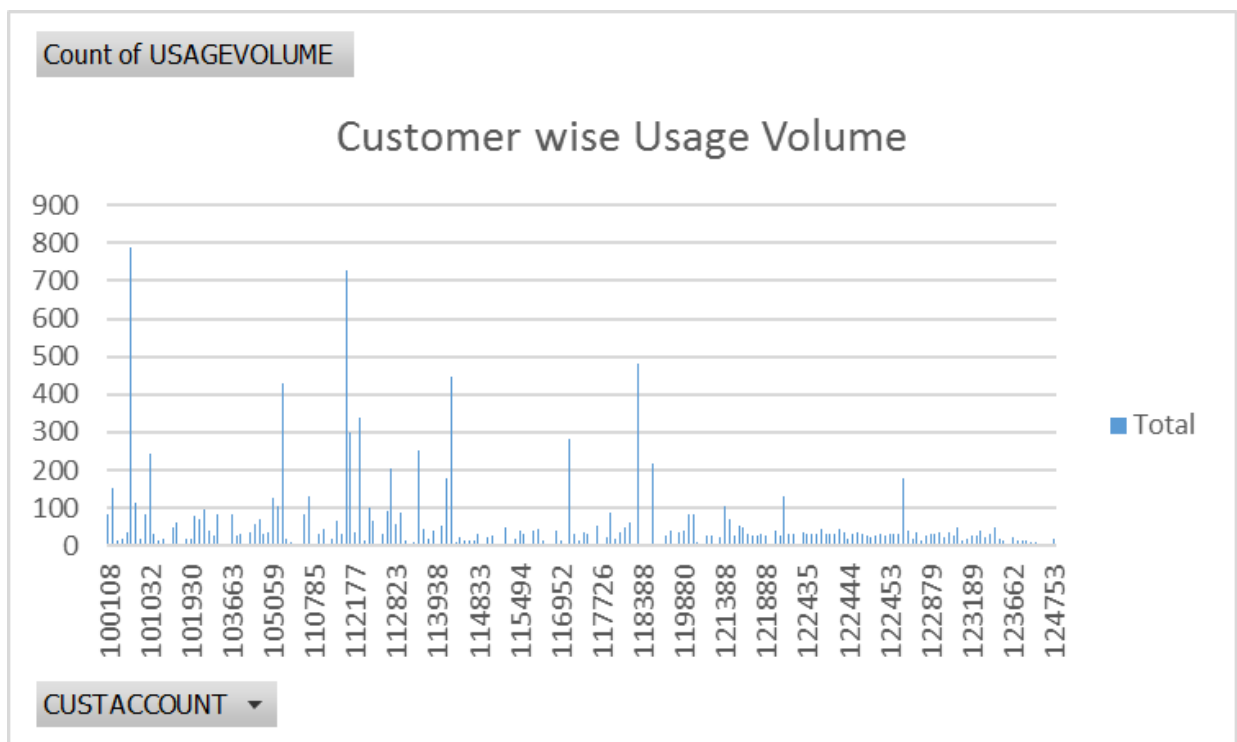

Figure 3-6: Customer Wise Usage Volume

Figure 3-6 illustrates of customer wise usage volume behavior. This variable helps to suggest new business strategies based on the fast-moving consumable items and usage copy count of photocopier machines.

Figure 3-6 has given an overview of customers against their photocopier usage volume. After that few customers has selected and has done an individual analysis for each customer.

Table 3-11: ACBT Usage Volume

Photocopier Machine Used Date	Photocopier Usage Volume
4/28/2014	957561
5/12/2014	963088
5/21/2014	966628
5/30/2014	966628
6/3/2014	829184
6/30/2014	975138
7/23/2014	976466
8/20/2014	979886
8/29/2014	980855
9/2/2014	981010
10/16/2014	985652
12/10/2015	986523

Table 3-11 consisted of the photocopier usage for the 2014-2016 of ACBT. The graphical view of the table 3-11 shows in figure 3-7.
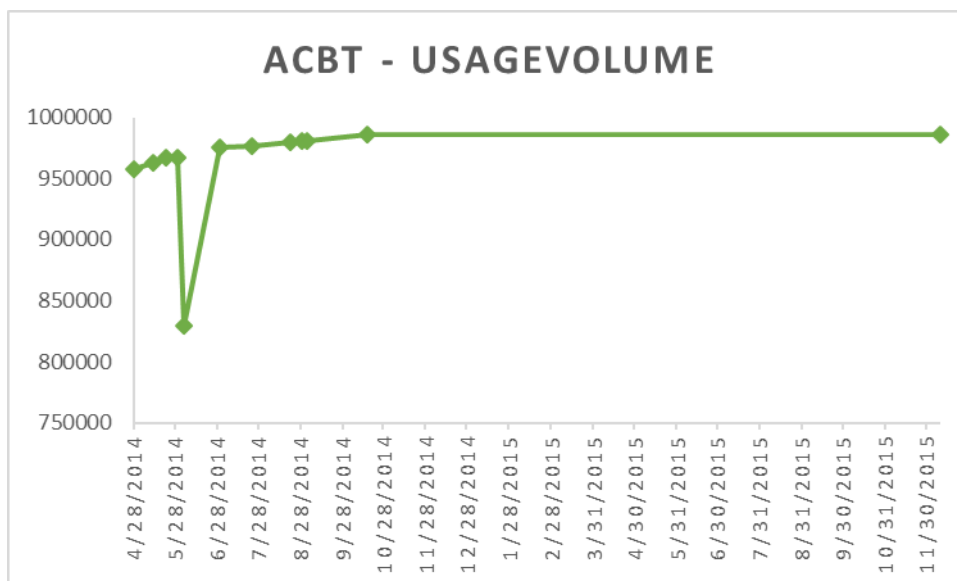

Figure 3-7: ACBT – Photocopier Usage Volume Analysis

Figure 3-7 represents the graphical view of usage volume behavior of ACBT. As shown in the graph between the month of May – July it shows the lowest copy count. After October 2014, there is no much of explicit in the graph.

Table 3-12: 3M Lanka Usage Volume

Photocopier Machine Used Date	Photocopier Usage Volume
3/21/2014	2611
4/28/2014	104922
4/30/2014	104922
5/21/2014	108102
5/27/2014	108102
6/28/2014	112157
7/21/2014	115470
7/23/2014	115470
8/20/2014	118675
8/26/2014	118675
9/22/2014	122224
10/16/2014	125840
10/22/2014	122224
11/26/2014	131663
12/15/2014	134204
12/16/2014	134204
1/28/2015	138318
2/20/2015	140954
2/25/2015	140954
3/20/2015	144499
4/22/2015	147510
5/23/2015	150493
5/26/2015	150493
6/18/2015	153231
7/17/2015	156362
8/21/2015	160374
9/17/2015	163767
10/19/2015	167144
11/21/2015	171302
12/23/2015	174307
1/25/2016	176797
2/19/2016	179781
3/2/2016	180671
2/25/2016	179781
2/26/2016	179781
3/17/2016	181983
3/28/2016	183124
4/28/2016	185435

Table 3-12 consisted of 3M Lanka phtocopier usage volume variations. The graphical view of Table 3-12 as followed:
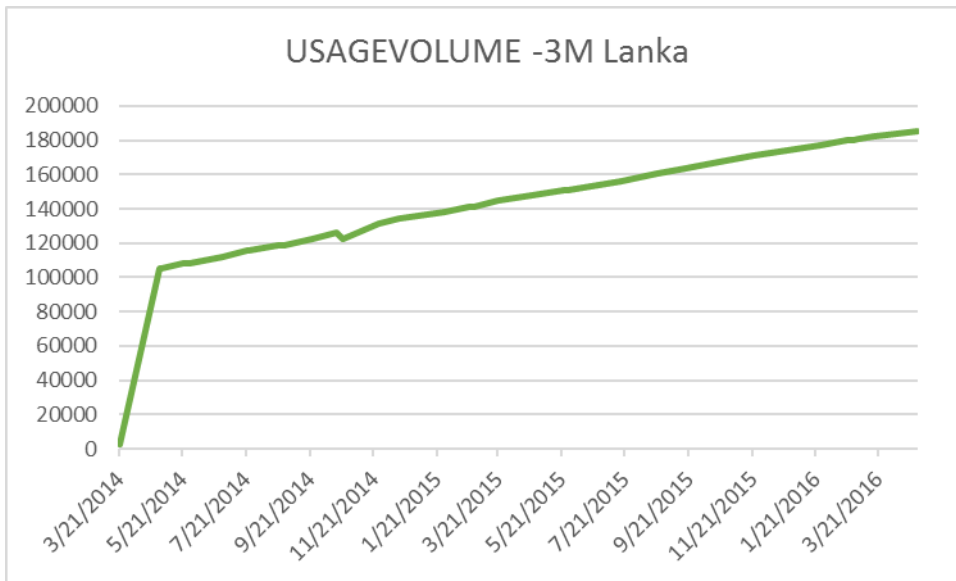

Figure 3-8: Usage Volume - 3M Lanka

Figure3-8 illustrates 3M Lanka behavior of copy counts. Like in Aitken Spence in this also it is difficult to give straight forward conclusion or explanation for this. There is no much of variation in the graph.

Table 3-13: Aitken Spence Usage Volume

Photocopier Machine Used Date	Photocopier Usage Volume
3/28/2014	114165
5/28/2014	19315
4/30/2014	119165
3/27/2015	7542
5/13/2015	236246
6/9/2015	245216
7/2/2015	249742
9/24/2015	265905
8/25/2015	80584
9/24/2015	279926
12/3/2015	295918
1/13/2016	321520
2/24/2016	335974
3/18/2016	346866
4/27/2016	356490
5/5/2016	384864

Table 3-13 consisted of photocopier usage volume of "Aitken Spence". The graphical representation of table 3-13 show in figure 3-9.
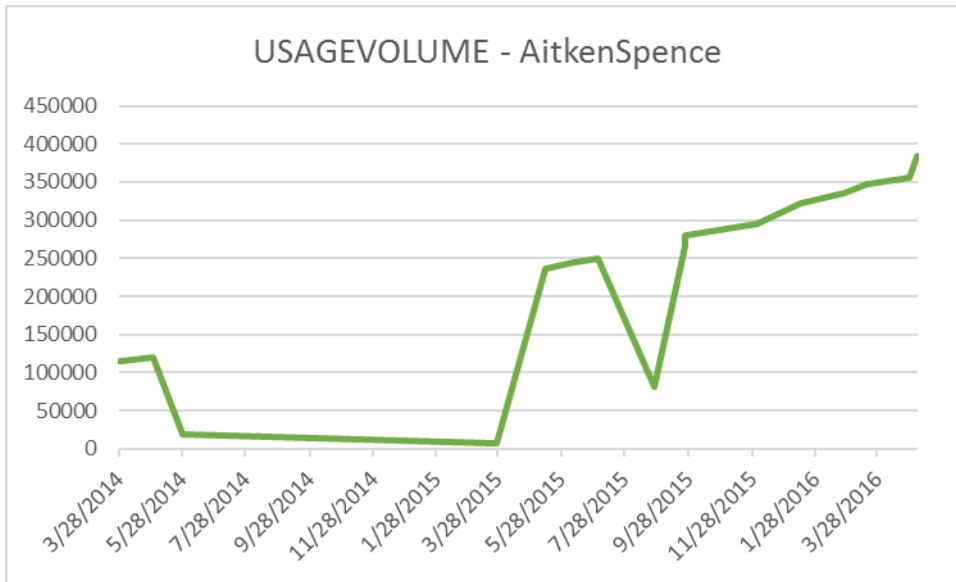

Figure 3-9: Usage Volume Aitken Spence

Aitken Spence showed a variation rather than 3M and ACBT graphs. Between, May 2014 to March 2015 shows very poor copy count and it has begun to increase in the month of May 2015. After that it show gradual increment and again in September 2015 copy count has been dropped and in November 2015 it has again has increased and maintained a gradual increment.

Table 3-14: Lex Duco Usage Volume

Photocopier Machine Used Date	Photocopier Usage Volume
7/23/2014	572
8/20/2014	2791
9/23/2014	10296
9/24/2014	10296
10/17/2014	17189
12/17/2014	102690
1/2/2015	28503
1/19/2015	39579
1/27/2015	42103
2/26/2015	48068
3/19/2015	54166
3/18/2015	54166
4/22/2015	60890
5/23/2015	67364
6/16/2015	78264
7/16/2015	88993
8/21/2015	105375
9/17/2015	113618

10/17/2015	115307
11/21/2015	117486
12/28/2015	119374
1/25/2016	121927
2/25/2016	125515
3/17/2016	128822
4/27/2016	132050

Table 3-14 shows copy count of LEX DUCO. The photocopier machine has been used from July 2014 to April 2016. Figure 3-10 shows the graphical representation of "Lex Duco" data.
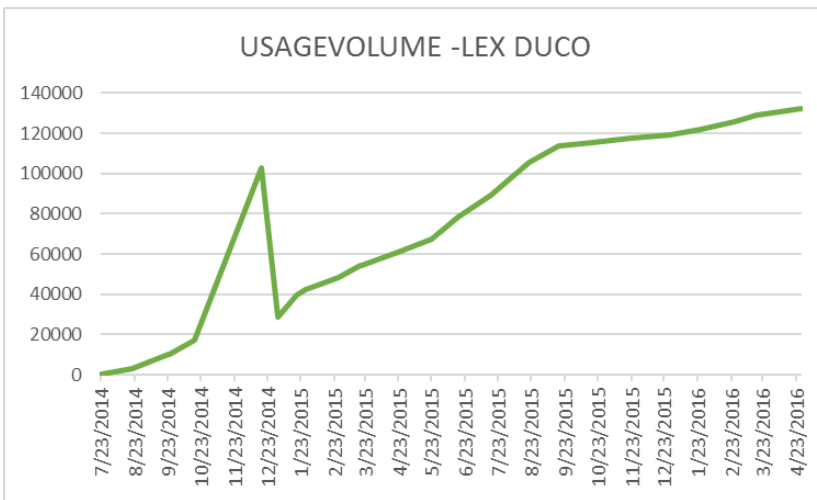

Figure 3-10: Usage Volume LEX DUCO

Figure 3-10 illustrates LEX DUCO photocopier machine's copy count. From July 2014 October 2014, it shows slight increment in copy count and it has dropped in between October 2014 to November 2014. Again, it has started to grow after November 2014. Between, November 2014 to December 2014 shows the highest copy count usage. Again, in January 2015 it has dropped and after January 2015 it has again grown and there were slight up downs shows till April 2016.

Likewise, basic analysis can be carried out for the dataset. Though this information give idea about the data set it does not fulfill the requirement of pattern recognition of fast moving items. Above analysis has done by separating attributes. As example at the very beginning of the is analysis data has been analyzed only taking consideration of sales components. Based on monthly and yearly has given DOC T4AL is the item which has highest sales within five years. But in that case photocopier usage volume or type of customer has not been taken to consideration. By having combinational attributes, analysis give a better picture to concluded. But using above techniques such as charts and tables does not give a clear picture to identify fast moving items.

Hence that it needs to carry on advance analysis for this dataset. This where data mining techniques come to play their role. Data mining techniques help to drill down dataset and identify hidden patterns within the data. Following sections will discuss about the approach that use to analyze dataset and techniques use for those approaches.

3.3 Design

Processing unprocessed data to useful information is the main task of data mining. Explore dataset, understand them and remove unwanted data, identify hidden patterns and relations which help to discover hidden opportunities are the main activities done by data mining techniques in a complex data set.
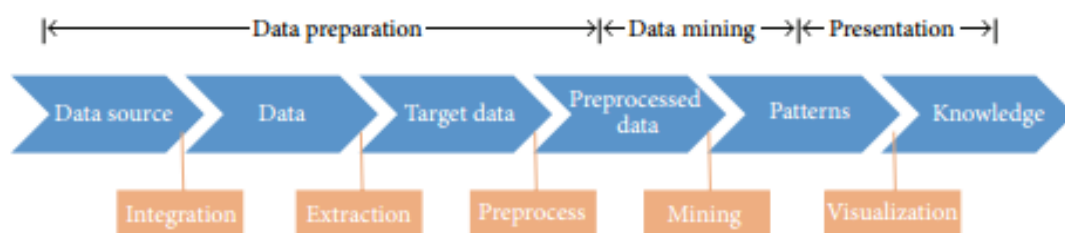
Figure 3-11: Data Mining Process

Above Figure 3-11 [21] illustrates the data mining process. In the section of design above mentioned process will be discussing descriptively with the application for the research area. Data source, data and target data which are sub components include in data preparation section has been discussed in first two chapters. The next two main steps in data mining process which are data mining and presentation discuss under third and fourth chapters. The key role which will be discussed in this section is data preprocessing and data mining techniques and patterns and finally the knowledge which will be discussed in chapter 4 will include evaluation and results of the research.

3.3.1 Data Pre-Processing

Data at its original presentation does not suitable for data mining. In-order to apply data mining unwanted data like noise should be removed from the data set. In the process of data mining all these disturbances are removed in form of transformation, integration and aggregation which finally provide effective output. This process called as knowledge discovery process [6].

Data preprocessing is one of the most important steps in data mining process. If there is irrelevant, noisy, unreliable data then it is difficult for knowledge discovery in training phase [22]. Data preprocessing consisted of data cleaning, data integration, data transformation and data reduction. The final output of data preprocessing is the final training dataset [22].

The first stage of the data preprocessing is the data extraction. The data has been extracted using SQL queries and stored in excel files. For the analysis, the excel sheet has been converted to CSV files. The descriptive discussion on data extraction has been done in section 3.1.

As per Rahm and Hong [23] data cleansing can be referred to removing errors and inconsistence from data in-order to increase the richness of the data. The problems with extraction data could be misspelling during data entry, poor validations, missing information or invalid data [23]. Data cleansing should be performed to detect and remove major errors in both individual and integrated multiple sources.

custAc	custname	Dimen	DIMEN	DATAAREA	SALESID	ITEMID	USAGEVOLUME	SERVICEOBJECTID	SVCCALLID	S
122457	UB FINANCE COMPANY LIMITED	206DOC	206CNM	206	206_071375	MX235AT-CHIP	F2-64 ERROR	SEROBJ-206-000399	SRQ206-018273	
122457	UB FINANCE COMPANY LIMITED	206DOC	206CNM	206	206_068388	OAT-MX235AT	F2-64 ERROR	SEROBJ-206-000399	SRQ206-017436	
122457	UB FINANCE COMPANY LIMITED	206DOC	206CNM	206	206_072570	OAT-MX235AT	F2-64 ERROR	SEROBJ-206-000591	SRQ206-018881	
122457	UB FINANCE COMPANY LIMITED	206DOC	206CNM	206	206_077059	OAT-MX235AT	F2-64 ERROR	SEROBJ-206-001153	SRQ206-021078	
124850	LOLC GENERAL INSURANCE LIMITED	206DOC	206CNM	206	206_077229	OAT-MX235AT	F2-64 ERROR	SEROBJ-206-000580	SRQ206-021263	

Figure 3-12: Invalid Data

Above Figure 3-12 emphasizes one of the errors found in the dataset. Usage volume is the field where record usage counts of photocopiers. The input data should only allow numeric values. Hence that there are two problems with this error. They are poor validation in the system and data entry error which led to an invalid information. These data have been removed from the dataset in-order to maintain the quality of the data.

custAc	custname	Dimen	DIMEN	DATAAREA	SALESID	ITEMID	USAGEVOLUME	SERVICEOBJECTID	SVCCALLID	
100108	ACBT	206DOC	206CNM	206	206_039281	AR310ST-CHIP		SEROBJ-206-000008	SRQ206-006737	
100108	ACBT	206DOC	206CNM	206	206_039618	AR310ST-CHIP		SEROBJ-206-000008	SRQ206-006915	
100108	ACBT	206DOC	206CNM	206	206_042274	AR310ST-CHIP		SEROBJ-206-000008	SRQ206-008094	

Figure 3-13: Null Data

Figure 3-13 illustrates another error found in the data set, which is null data. These values do not add any value to the dataset. This may lead to bad decision making since this field (usage volume) is one of the key attributes in the research. Null values lead to have outliers in the analysis which does not interpret any information of the data. In-order to avoid misleading these kinds of errors have been removed in the dataset.

After cleansing the dataset, the next main task was data integration. Data integration evolves with combine data from multiple sources into a coherent dataset with having meaningful relation to each data [22]. In this research, data has been gathered from different tables and integrated using relationships.

33

Figure 3-14: Data Files

Above Figure 3-14 illustrates some of the data files involved to retrieve the final dataset for the analysis. In the integration, some of the data has been removed because those data did not fit for the best relationships existed in the dataset. By using these removals and integration final data set has been retrieved for the analysis.

Another task in this process is, selecting suitable attributes for the research. Table 3-15 shows selected attributes for the analysis.

Table 3-15: Attributes Used for Analysis

Attribute	Description
CustAccount	Customer Account
CustName	Customer Name
ItemID	Consumable Item Id
Usage Volume	No of Copies
CreatedDateTime	Usage time of photocopier machine
Item Type	Type of the Item
Sales Month	Sales Month

Some of the attributes such as, sales quantity, call id, customer address etc. have been dropped out from the dataset because those attributes did not show value for the analysis. Customer account and name has been selected because these two attributes help to identify customer impact on high usage for photocopiers. Usage volume, start use date, end use date are the attributes related photocopiers. Item id and created date time involve with item related attributes.

Above steps have been followed in-order to retrieve the dataset with data which gives meaning to the dataset and its analysis.

K-means algorithm and drill down technique have been used as data mining techniques for analyze the dataset.

3.3.2 K- Means Clustering

Cluster analysis separate data into groups for understand or summarization [3]. These groups (clusters) become the goal of analysis. Clusters grouped items with similar attributes for the gave scenario. Furthermore, cluster algorithm groups objects based on the information provided in the context of relationships and objects [3]. The main goal of cluster analysis is that the objects in a cluster should have similarities and objects in different clusters should have high percentage of dissimilarities.

Clustering refers to act of partitioning an unlabeled dataset into groups of similar objects [24]. In different research, there are different application of cluster analysis such as clustering as a data processing technique in artificial intelligence, image analysis, pattern recognition, marketing, object recognition, data mining, data compression and many more [24]. Any cluster should show two main properties as in low inter class similarity and high intra class similarity. Inter cluster similarity means similarity between two different clusters and intra cluster similarity means similarity between objects in the same cluster. Clustering happen based on observation rather than examples. Hence that clustering known as unsupervised learning [24]. Clustering helps in gaining, overall distribution of patterns and correlation among data objects.

In clustering pattern recognition data analysis evolves with predicative model, for the given training dataset predict behavior in unseen data set [24].

One of the popular cluster algorithm is K-means clustering. K-means highly used to identify clusters where numerical values involved. Joshi, Pandey and Chawla have described K means algorithm as simple, straightforward and is based on the firm foundation of analysis of variances [4].

The mechanism of K- means is, it has two inputs such as number of cluster to be formed k and the data to be clustered [4]. Clustering starts with the center chosen randomly or follow certain procedures. This followed as iterations. The distance calculates using Manhattan distance or Euclidean distance between cluster point and the center. The center of the cluster recalculated and center of each clusters calculated as mean of all belongings of the same cluster [4]. Joshi, Pandey and Chawla further explained this algorithm aimed at minimizing an objective function, in this case a squared error function. Objective function of K-means clustering as followed:

$$ J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 $$

Figure 3-15: K- Means Equation

35

$$\left\| x_i^{(j)} - c_j \right\|^2$$

= Distance between cluster point and the center

Figure 3-15 illustrates K-Means function used in K-Means algorithm. K is the number of clusters and differentiation of X and C gives the distance between cluster points for the centroid.

Furthermore, K- means algorithm can be defined as followed [6]:

1. Place K points into the space represented by the objects that are being clustered; these points represent initial group centroids.
2. After that each of the object need to assign to a group which has closest centroid.
3. After assigning all the objects to respective groups recalculate the positions of the K centroids.

   Likewise step 2 and 3 need to be repeated until the centroid no longer change. As [6] described, this process helps to produces a separation of the objects into groups from which the metric to be minimized can be calculated.

In this research, K-means clustering has been used to divide items into clusters in-order to identify photocopier consumable items. In-order to identify fast moving items it needs to use different K values to identify most suitable K value for the analysis. Further K-means algorithm also has been used to identify seasonal trends of volume usage in photocopiers based on customers.

Most of the time transactional data shows some seasonal trend. It is difficult to identify patterns in the dataset because of high amount of data availability [9].

Fast moving item pattern recognition measure based on sales quantities or frequent sale dates. These two factors are the dependencies for clustering. In this analysis, there are thirteen different consumable items and there was large amount of sales involved with these items.

Pattern recognition of usage volume in photocopiers based on the volume and usage period of the photocopier machine. This helps to identify which consumable items will be fast moving based on the usage volume when the usage volume increases gradually consumable items also increases. The main task of this is to identify which consumable items has been sold based on these usage volumes. Hence that seasonal pattern recognition of usage volume and fast moving items pattern recognition is the most critical point in this analysis. Since this analysis evolves with clusters K-means algorithm is the most suitable clustering algorithm for this dataset.

Also, this analysis leads to suggest business strategies for the business based on the seasonal trend analysis.

3.3.3   Most Frequent Pattern [MFP] Algorithm

As depicted in [7] Association rule mining is one of the most important and well defined technique that use for extract correlations, frequent patterns, associations or causal structures among sets of items in the transaction databases or other repositories. There are many areas that can be applied association rules such as, market analysis, risk management, health care, stock data and many more. Association rule is one of the widely-used techniques in different applications. Among them Apriori algorithm strongly support in deriving association rules [7].

MFP is an algorithm which efficiently support to generate frequent patterns and identify strong associations among them [6].

MFP algorithm as followed: Set X of N items in a dataset having set Y of attributes. This algorithm counts maximum of each attribute

In this research, MFP algorithm used to identify the fast-moving item set. From the K-means algorithm initially dataset has been divided to clusters based on the suitable K value. For the identified clusters apply MFP to identify fast moving items.

3.3.4   Modeling

In-order to analyze a dataset it is necessary to have a proper plan/model. Model helps to carry out an analysis in an effective manner. Model is an approach use to analyze data. Because for an analysis there are different stages. Analysis conducts to solve a problem or identify an opportunity. In this research analysis, has been done to solve a problem by identify potential attributes related to the problem.

Figure 3-16 shows approach used in – order to solve the identified problem.
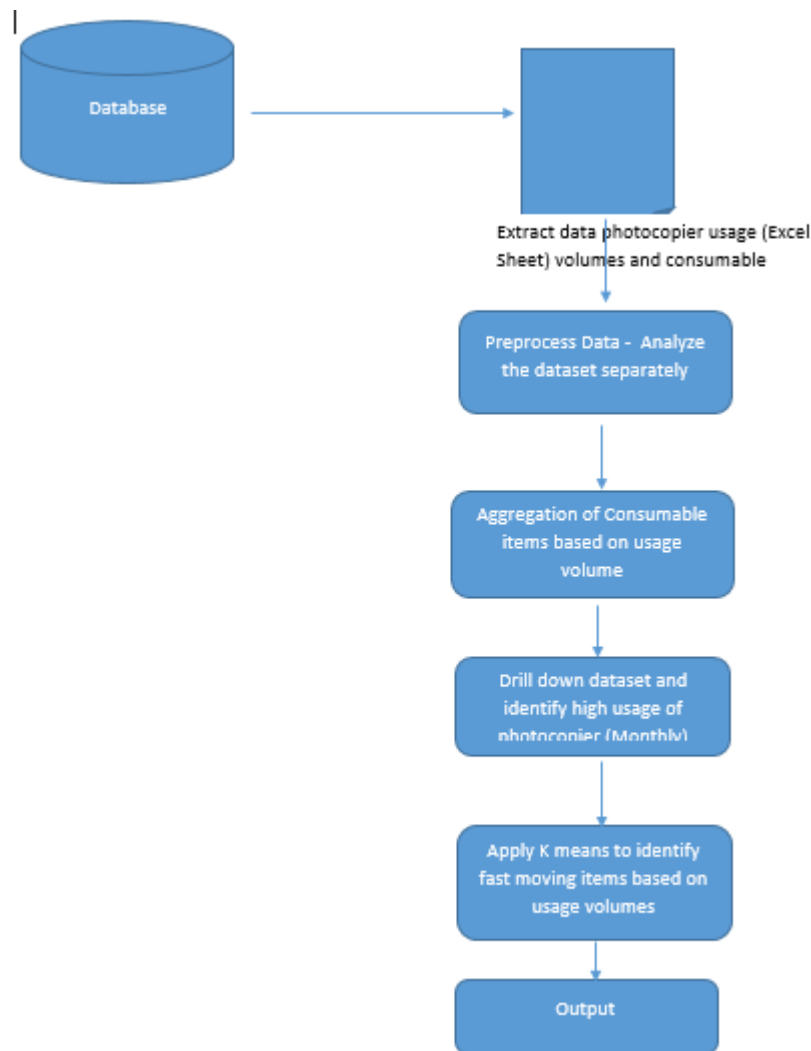
Figure 3-16: Methodology

.

As showed in figure 3-16 necessary data sets have been retrieved from the database by using different queries.

The captured data has been exported to an Excel sheet. From the Excel sheet carried out a study of the data set. From studying the data set got clear understanding of the business and the data business dealing with. After the study and analysis removed unnecessary data and preprocessed the data set. There were mainly two separate data sets. They are,

1. The dataset with copy count of the copiers and related data
2. The dataset with consumable items and related data

After that aggregate these two data sets with related fields. From that identified association between these two datasets.

After aggregating the dataset once again carried out an analysis to get a brief idea about the flow of items have been moved. From the analysis get a clear understanding how the seasonal wise consumable items have been moved against the copy count of photocopiers.

After aggregating identify fast moving, slow moving and dead stock. This can be found out from the sales of an item. Finally, most frequent set will be given as the output of the research.

Summary

From the point of data extraction to modeling has been included in this chapter. Data extraction methods and problems faced during the data extraction has been included in this. Understanding data is one of the important point in a research project. In-order to understand data it is required analyze data in different ways using different approaches. The process of pre-processing evolves with cleansing, integrating and transforming data. The use of data mining in complex data and the importance of K-means clustering algorithm in pattern recognition photocopier usage volume and clustering consumable items based on their sales has been descriptively discussed in this. From the drill down identify customers who are using photocopier machines for high number of times. Based on results apply K-means to identify fast moving item set.

The next chapter discuss about the ways use to evaluate data and the results derive from the analysis.

4.0　Chapter 4: Evaluation and Results

Previous chapter is the foundation of chapter 4. In chapter 3 has discussed about techniques and models used to prepare data which suitable for analysis. As discussed in previous chapter K-means and MFP selected as data mining techniques to analyze the dataset. The results of the research will be descriptively discussed in this chapter.

This chapter is the key chapter of this research. Identification of patterns in photocopier usage and identification of fast moving items in the dataset perform in this chapter.

Evaluation conduct in two main steps. First, drill down the dataset into quarters. From that identify the customers with high usage and provide business justification reasons for high volume usage of photocopies during the period. The second and the most important step is the identification of fast moving consumable items use K-Means analysis.

4.1　Key Evaluation and Findings – Drill-down Method
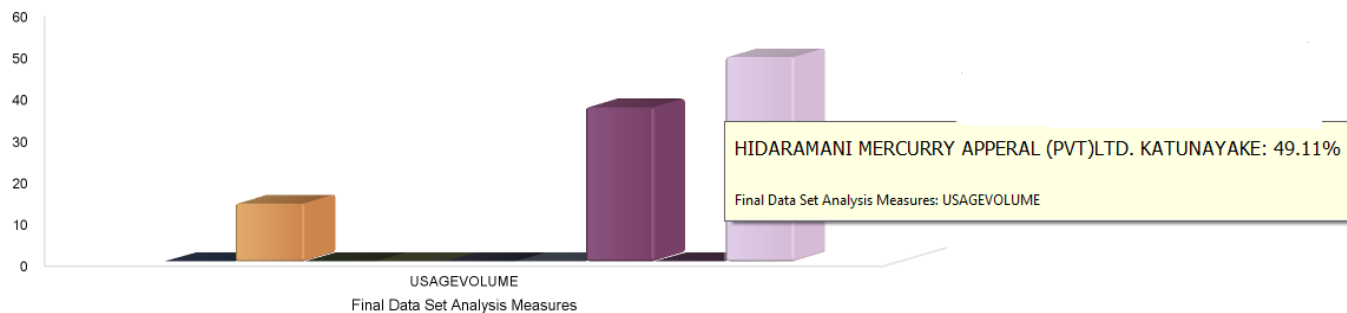
Figure 4-1: 2012 - Q1 Photocopy Usage Volume Variation Against Customers

Figure 4-1 illustrates 2012 quarter 1photcopy usage volume against customers. "Hidramani" Mercury Apparel had the highest usage volume which is 49.11 % from the total usage. In the first quarter of 2012 month February has not recorded any significant usage volume from any of the customers.
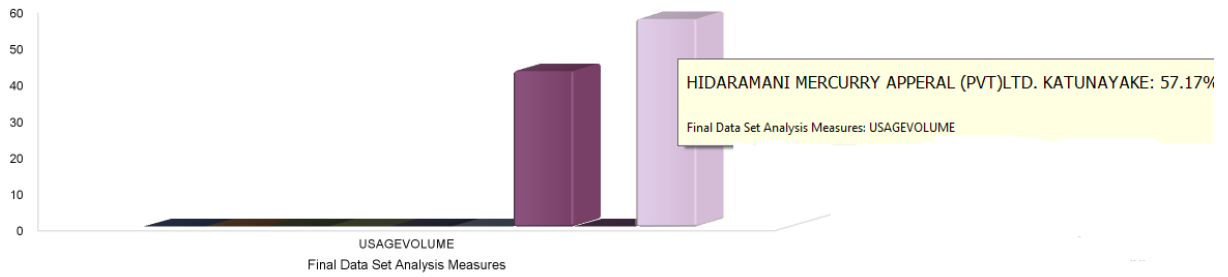
Figure 4-2: 2012 January Photocopy Usage Volume Variation Against Customers

Figure 4-2 shows photocopy usage volume for month of January. "Hidramani" had the highest usage volume which is 57.17% from the total usage. From total usage, more than half of the usage recorded from "Hidramani."
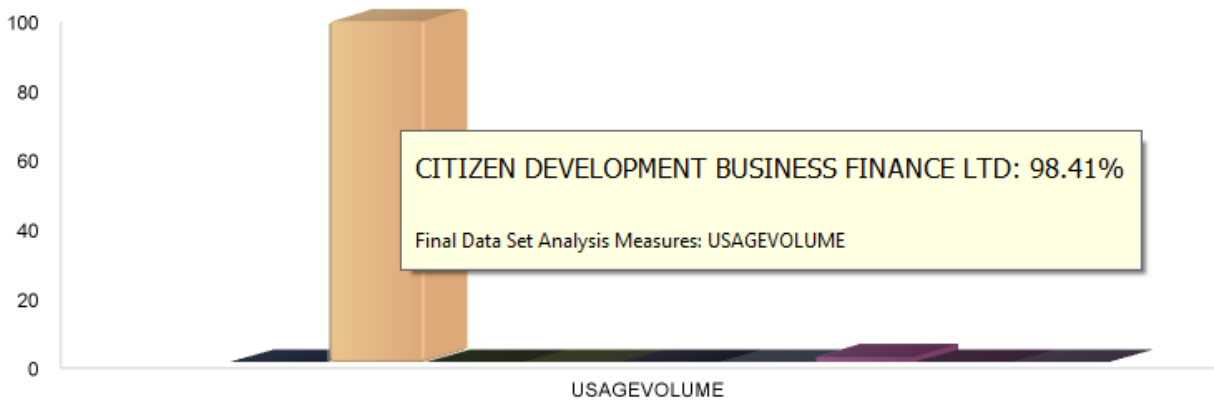
Figure 4-3: 2012 March Photocopy Usage Volume Variation Against Customers

March 2012 photocopy usage volume illustrates in figure 4-3. Citizen Development recorded the highest usage volume. The reason for this high usage could be because of the end of financial year. Since Citizen Development attached to financial sector lot of paper work involve with the business.
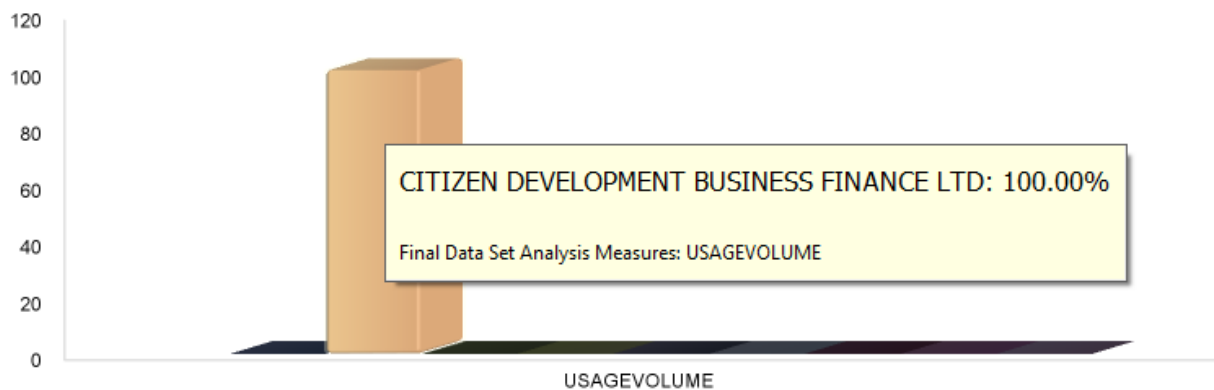
Figure 4-4: 2012-Q2 Photocopy Usage Volume Variation Against Customers

41

Photocopy usage volume of 2012 quarter to shows in figure 4-4. One of the highlights facts in quarter two is all three months in quarter 3 Citizen development bank recorded the highest usage volume.
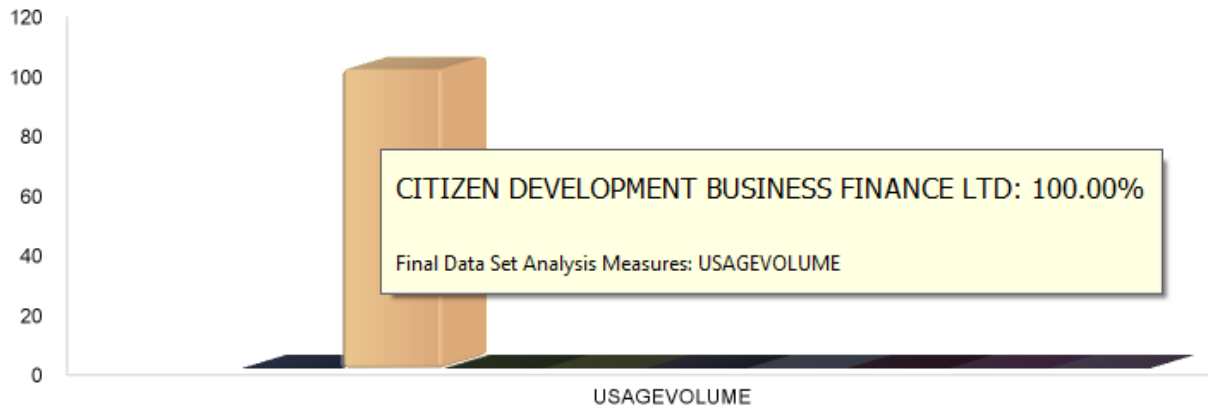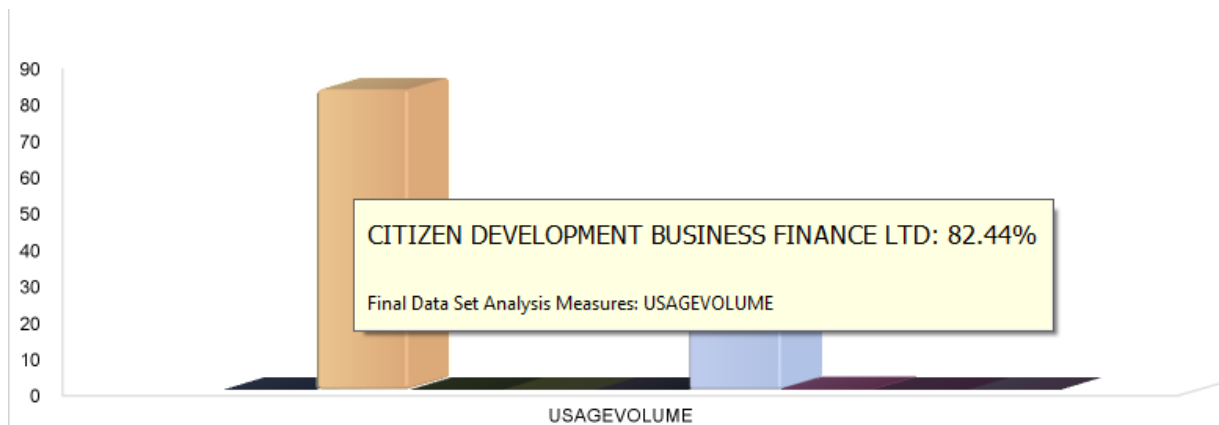

Figure 4-5: 2012 April Photocopy Usage Volume Variation Against Customers

Figure 4-5 illustrates the usage volume against customers for the month of April 2012. The highest photocopy usage volume recorded from Citizen Development Business Finance Ltd.
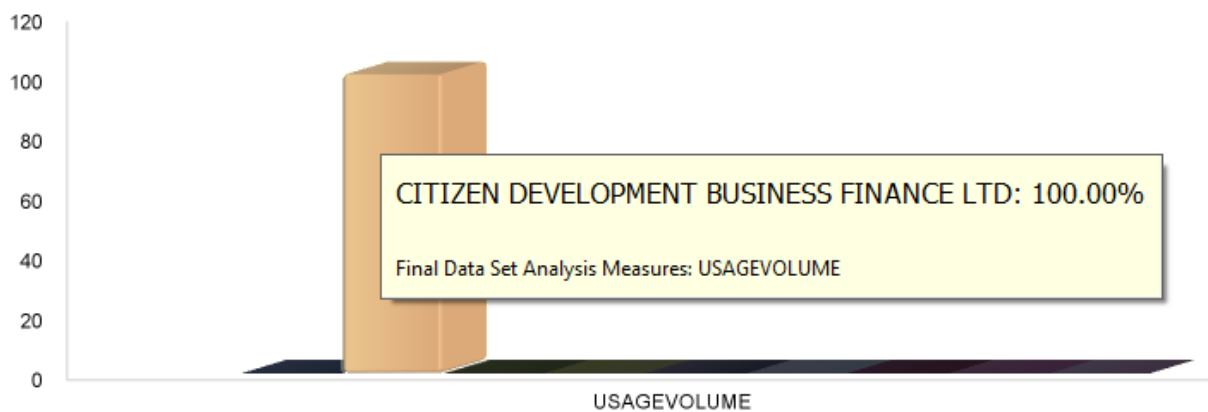
Figure 4-6: 2012 May Photocopy Usage Volume Variation Against Customers

Figure 4-6 demonstrates the usage volume against customers for the month of May 2012. The highest photocopy usage volume recorded from Citizen Development Business Finance Ltd.
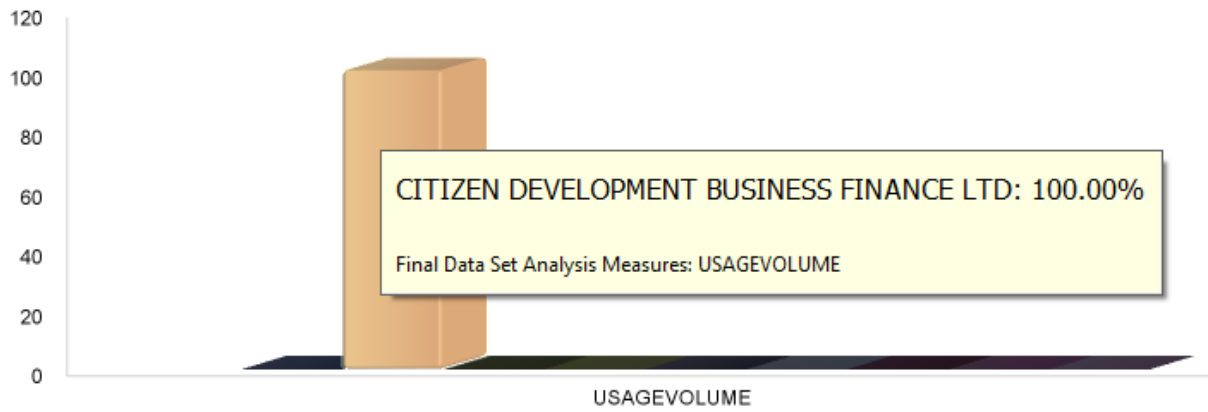
Figure 4-7: 2012 June Photocopy Usage Volume Variation Against Customers

As shown in figure 4-7 like in April and May in the month of June highest usage volume has recorded from the Citizen Development Business Finance Ltd.
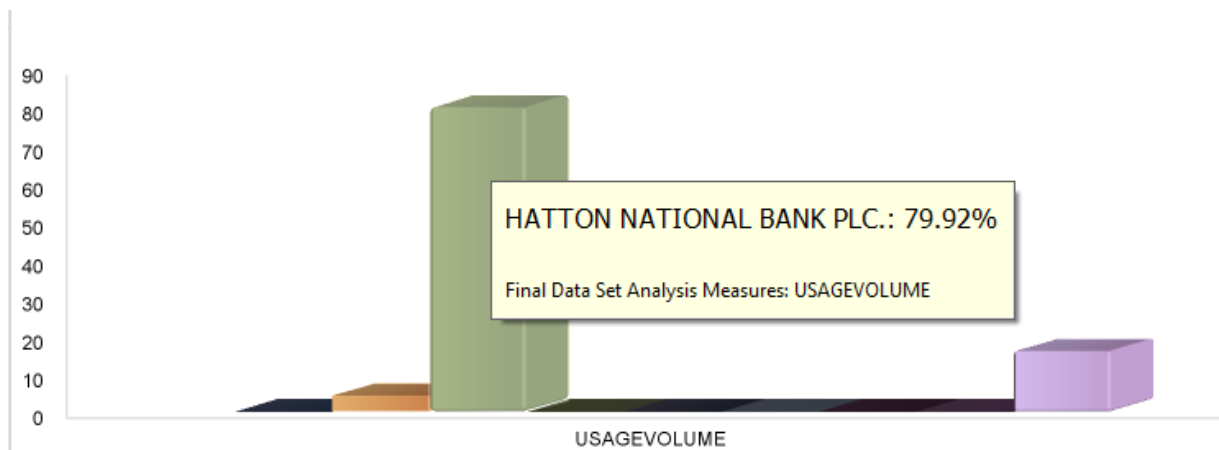
Figure 4-8: 2012 Q3 Photocopy Usage Volume Variation Against Customers

In the third quarter of 2012 highest usage volume has recorded from Hatton National Bank which showed in figure 4-8. It has recorded as 79.92 % from the total usage which is more than half of the total usage.
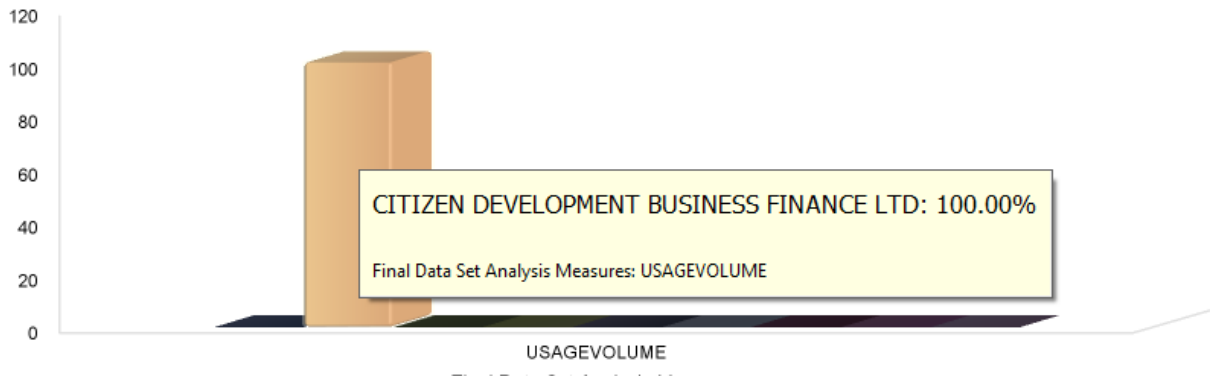
Figure 4-9: 2012 July Usage Volume of Photocopies Against Customers

In the month of Jul 2012 highest photocopy usage volume records from Citizen Development which shows in figure 4-9.
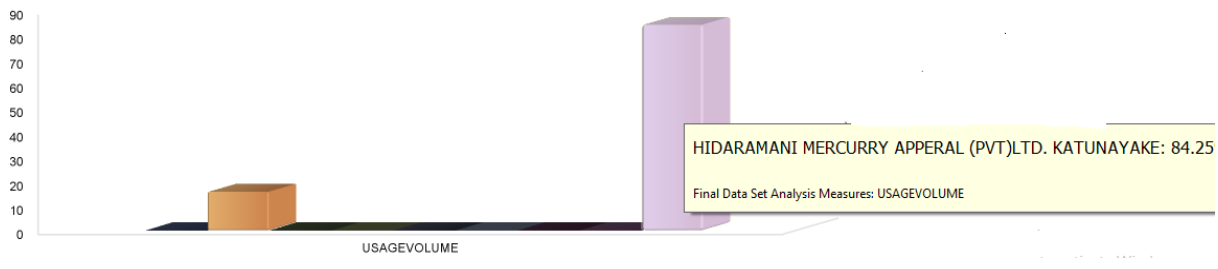
Figure 4-10: 2012 August Photocopy Usage Volume Variation Against Customers

Figure 4-10 illustrates photocopy usage volume in the month of August 2012 against the customers. "Hidramani" Mercurry Apparel shows the highest usage volume of photocopies which recorded as 84.25% from the total usage in month of August.
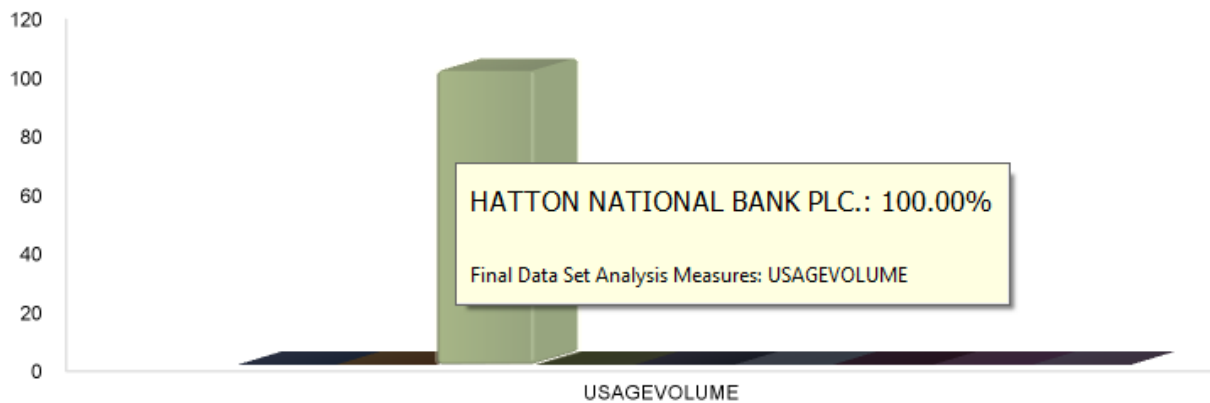
Figure 4-11: September Photocopy Usage Volume Variation Against Customers

Figure 4-11 demonstrates photocopy usage volume against customer for the month of September. Hatton National Bank recorded the highest usage volume with significant percentage which is 100% from the total photocopy usage volume.

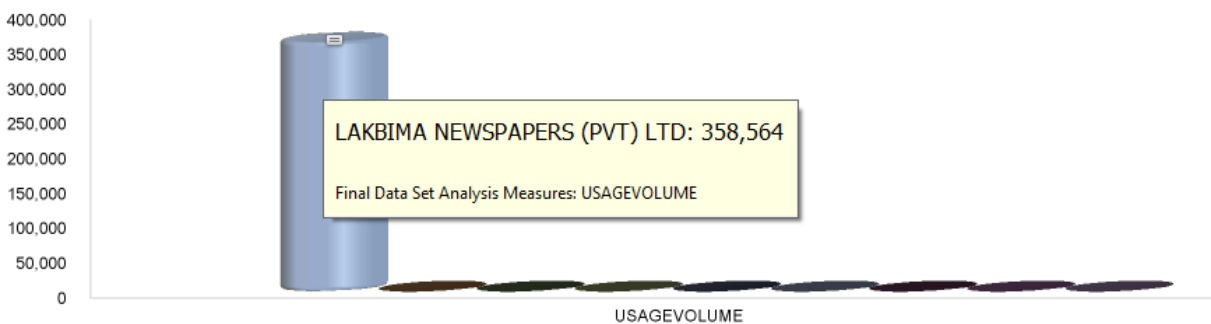
Figure 4-12: 2012- Q4 Photocopy Usage Volume Variation Against Customers

Figure 4-12 illustrates the usage volume variation against customers. Highest usage volume recorded from the Citizen Business Finance Ltd which is 82.44 %. In the fourth quarter, none of the customers shows significant amount of usage volume.
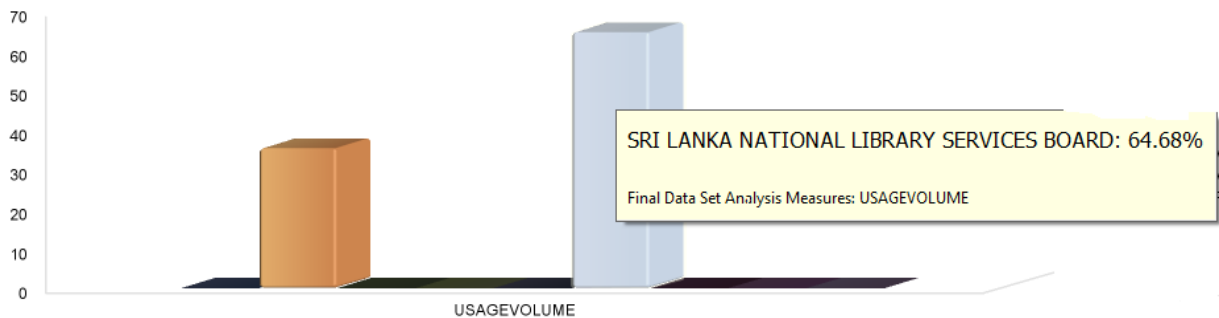
Figure 4-13: 2012 November Photocopy Usage Volume Variation Against Customers

November 2012 usage volume variation shows in figure 4-13. Sri Lanka National Library Services shows the highest usage volume which is 64.68% from the total usage from the month of November.
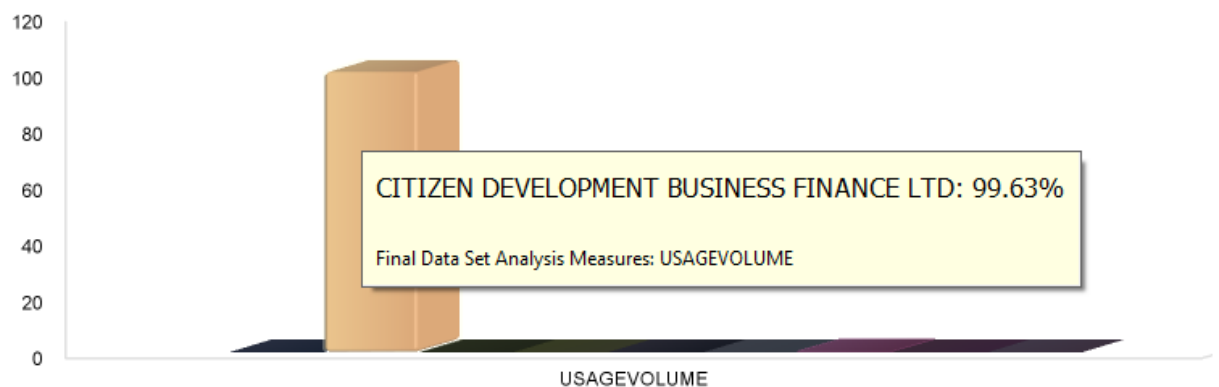
In the month of December 2012 Citizen development shows the highest usage volume which is nearly 100%.
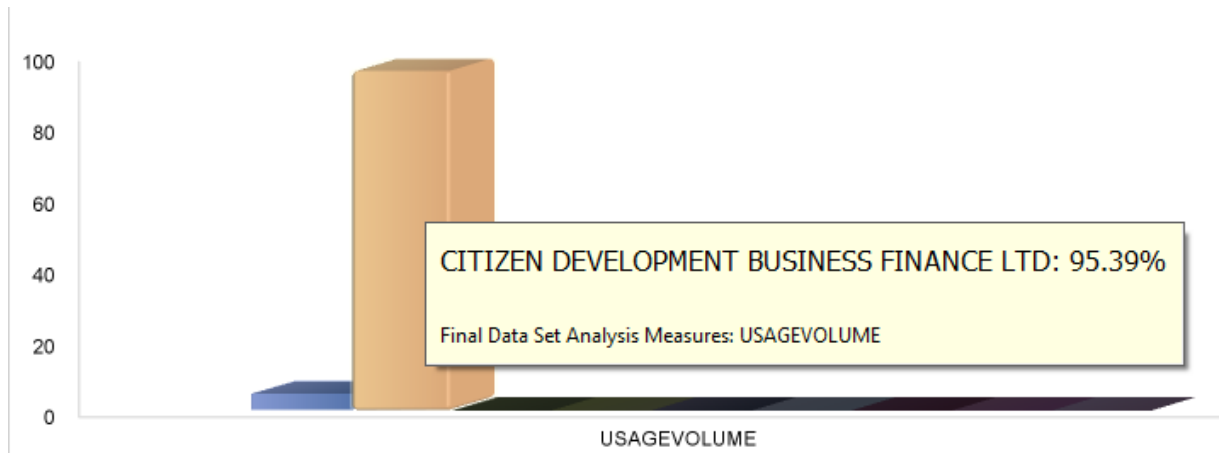
Figure 4-15: 2013 – Q1 Photocopy Usage Volume Variation Against Customers

Figure 4-15 illustrates 2014 quarter one photocopy usage against different customers. In the 2014 quarter one Citizen Development has recorded highest usage volume.

Figure 4-16: 2013 January Photocopy Usage Volume Variation Against Customers

Figure 4-16 illustrates usage volume of photocopies during the month of January 2013. In January Citizen Development recorded 6,270,930 photocopy usage volume.
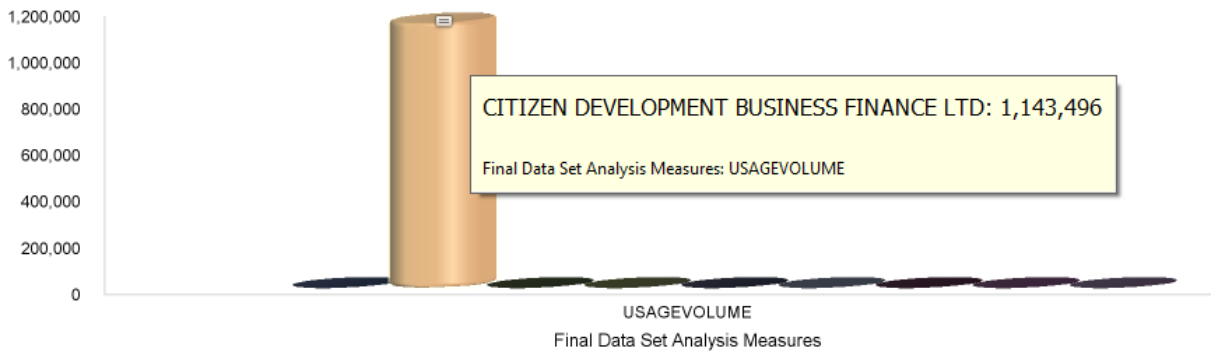
Figure 4-17: 2013 February Photocopy Usage Volume Variation Against Customers

Figure 4-17 illustrates usage volume of photocopies during the month of February 2013. In January Citizen Development recorded 1,143,496 photocopy usage volume. Even though Citizen Development shows highest usage volume in February the usage volume has decreased by 18% compared to month of January.

Figure 4-18: 2013 March Photocopy Usage Volume Variation Against Customers

Figure 4-18 shows the usage volume visualization against customer for the month of March 2013. "Lakbima" newspapers recorded the highest usage volume. But it is very less compared to January and February.

Figure 4-19: 2013- Q2 Photocopy Usage Volume Variation Against Customers

Figure 4-19 illustrates usage volume of photocopies for second quarter 2013. In the second quarter Citizen Development has recorded highest usage volume. In the first quarter also Citizen Development had highest usage volume but in the second quarter it shows slight increment in their usage of photocopy machines. Also one of the higlighted facts in the second quarter of 2013 none of the customers showed significant usage volume in month of April.
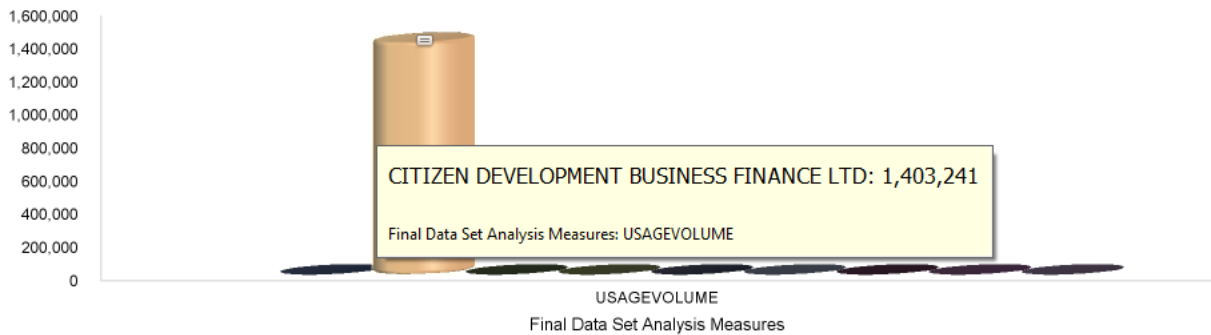
Figure 4-20: 2013 May Photocopy Usage Volume Variation Against Customers

Figure 4-20 shows photocopy usage for the month of May. Once again Citizen Development showed the highest usage volume of 1,403,241.
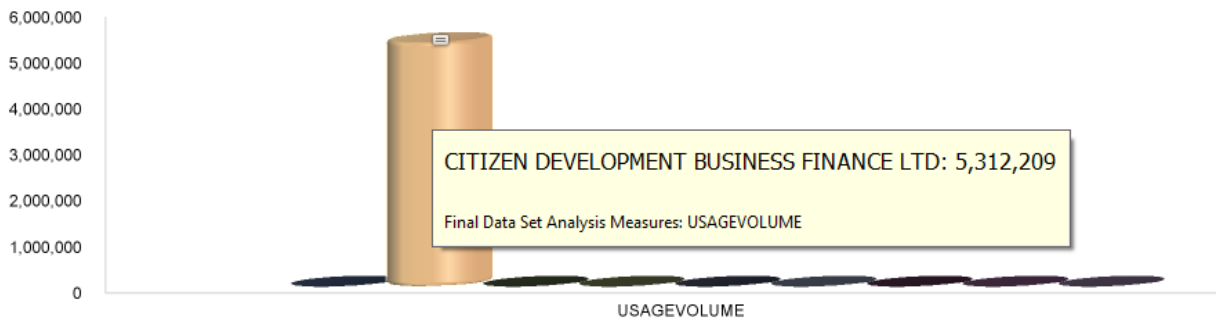
Figure 4-21: 2013 June Photocopy Usage Volume Variation Against Customers

In the month of June Citizen Development shows the highest usage volume of photocopies which illustrated in figure 4-21. It shows 26% of increase compared to month of May.
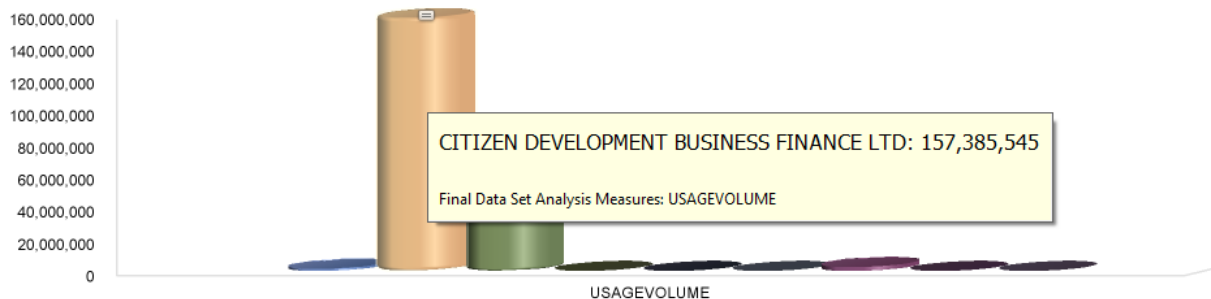
Figure 4-22: 2013-Q3 Photocopy Usage Volume Variation Against Customers

In the third quarter of 2013 like in first and second quarters Citizen Development recorded highest usage volume which showed in figure 4-22. In the third quarter recorded the highest usage volume compared to first and second quarters.
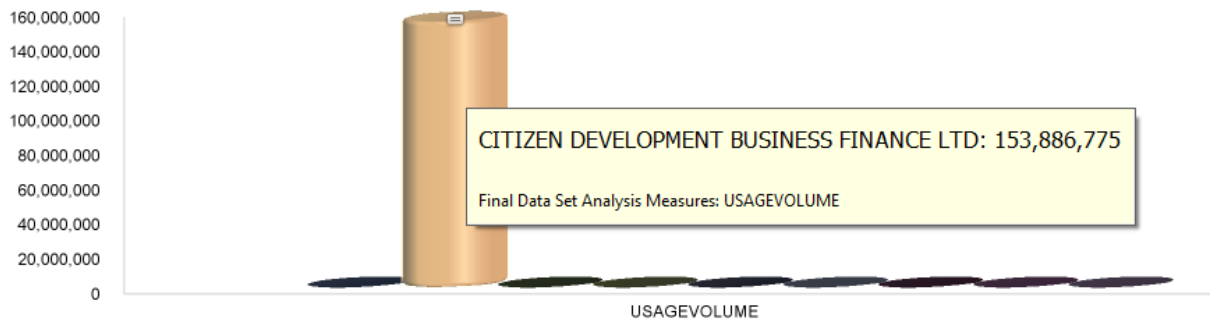
Figure 4-23: 2013 July Photocopy Usage Volume Variation Against Customers

Figure 4-23 illustrates photocopy usage volume variation against customers for the month of July 2013. Citizen Development recorded highest photocopy usage volume which is 253,886,775.

Figure 4-24: 2013 August Photocopy Usage Volume Variation Against Customers

In month of August 2013 Hatton National Bank shows highest usage volume which demonstrated in figure 4-24. Even though Hatton National Bank shows highest usage volume the photocopy usage of Hatton National Bank less than Citizen Development.
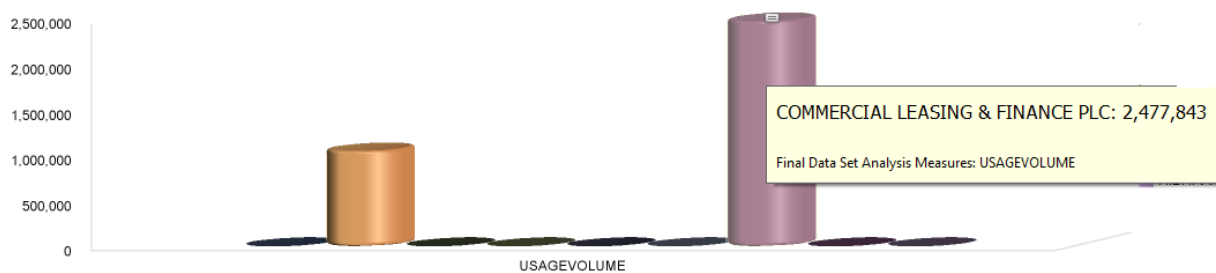
Figure 4-25: 2013 September Photocopy Usage Volume Variation Against Customers

49

Figure 4-25 illustrates photocopy usage for September 2013. Commercial Leasing and Finance recorded highest usage volume. But this the lowest usage volume recorded in third quarter of 2013.
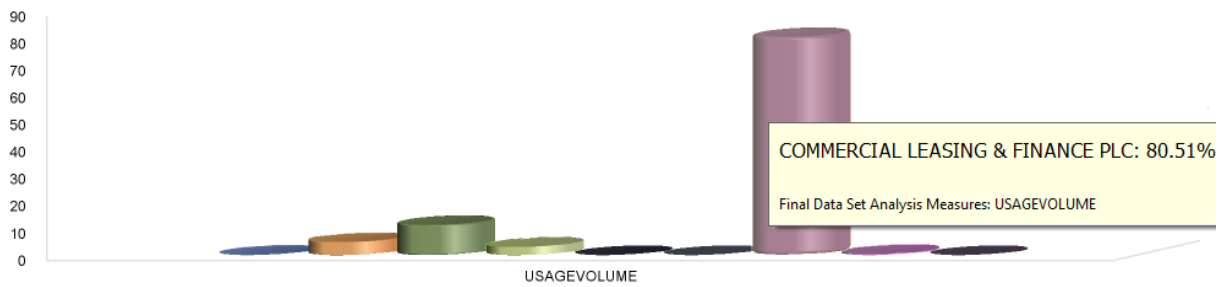
Figure 4-26: 2013- Q4 Photocopy Usage Volume Variation Against Customers

In the final or the fourth quarter of 2013 highest usage volume shows in Commercial Leasing and Finance. It has recorded 80.51% usage volume.
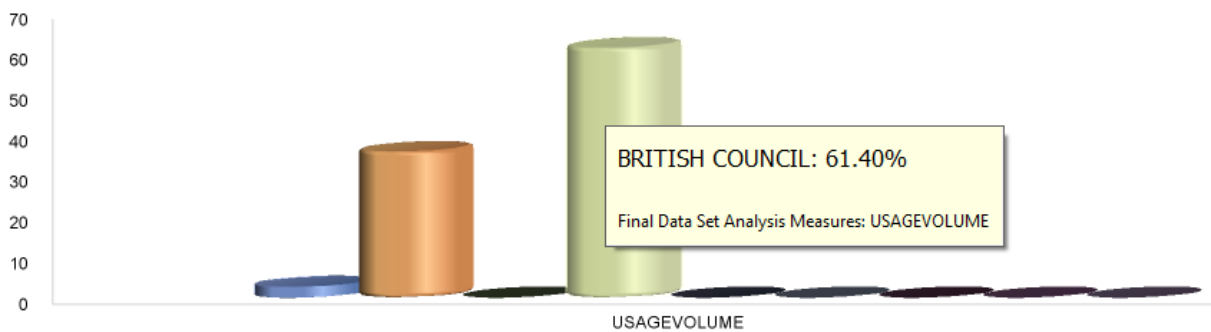

Figure 4-27: 2013 October Photocopy Usage Volume Variation Against Customers

A new customer highlighted in the month of October having highest usage volume of 61.5%. British Council recorded with highest usage volume in the month of October whish showed in figure 4-27.
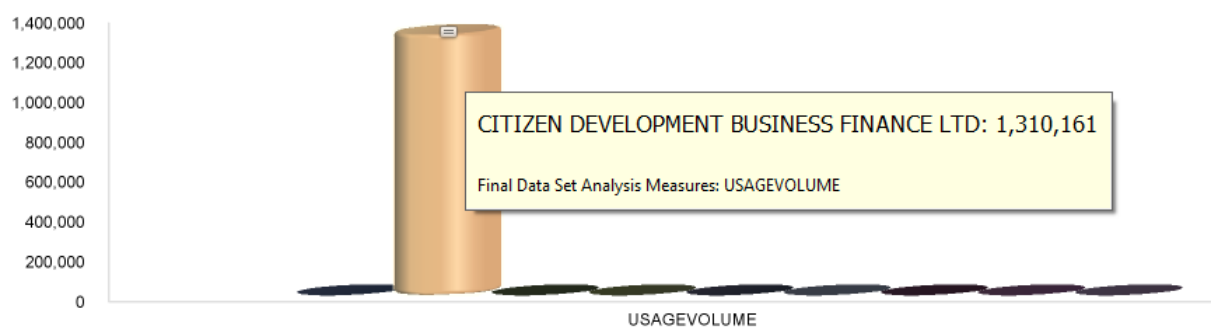
Figure 4-28: 2013 November Photocopy Usage Volume Variation Against Customers

Figure 4-28 illustrated usage volume variation in the month of November. Once again Citizen Development highlighted with the highest usage volume of 1,310,161.
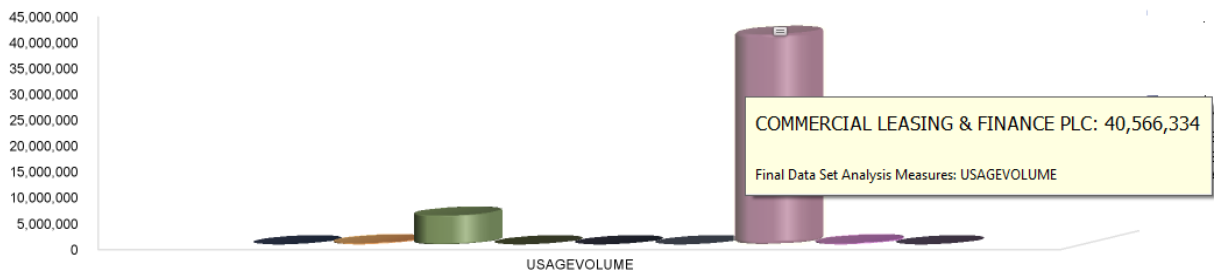

Figure 4-29: 2013 December Photocopy Usage Volume Variation Against Customers

The last month of 2013 which is December had Commercial Leasing as the customer who had highest usage volume which showed in figure 4-29. The photocopy usage volume has recorded as 40,566,334 which is the highest photocopy usage volume in 2013 fourth quarter.
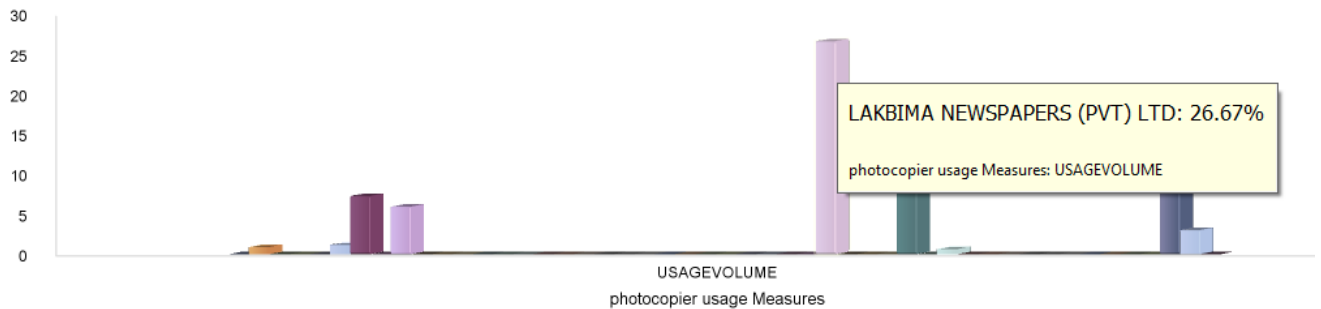

Figure 4-30: Compare 2014 - Q1 Photocopy Usage Volume Variation Against Customers

Above Figure 4 -30 illustrates the photocopy usage in the first quarter in 2014. As shown in the figure the highest usage volume recorded in "Lakbima" newspapers (Pvt) Ltd. It records 26.67% (452,893) from the total usage volume during the first quarter. Furthermore, these highest usage volumes have recorded in month of March.
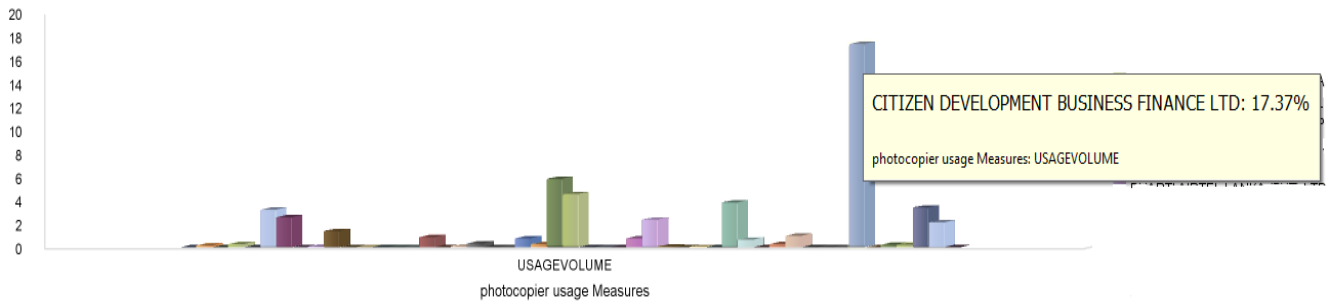
51

Figure 4-31: 2014 - Q2 Photocopy Usage Volume Variation Against Customers

Figure 4-31 illustrates photocopy usage volume in quarter 2 in the year of 2014. As shown above Citizens Development Business Finance has highest usage in the second quarter which is 17.37% from total usage. Unlike in first quarter in the second quarter considerably high usages have recorded in few months such as April, May and June.
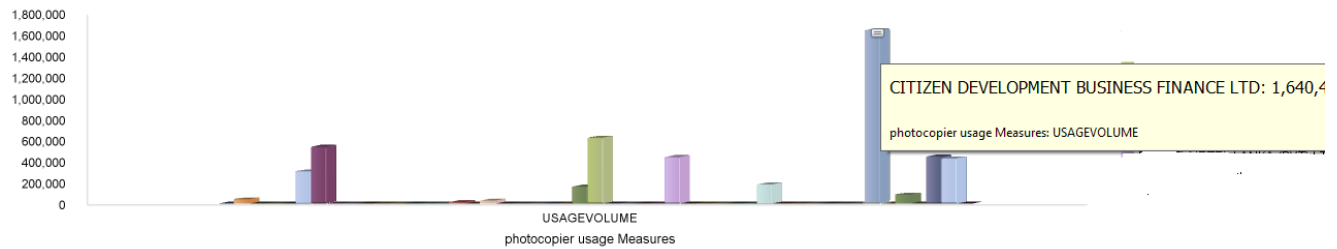

Figure 4-32: 2014 April Photocopy Usage Volume Variation Against Customers

In 2014 May, Highest photocopy usage volume recorded by Citizens Development and usage volume recorded as 1640469 which showed in figure 4-32.
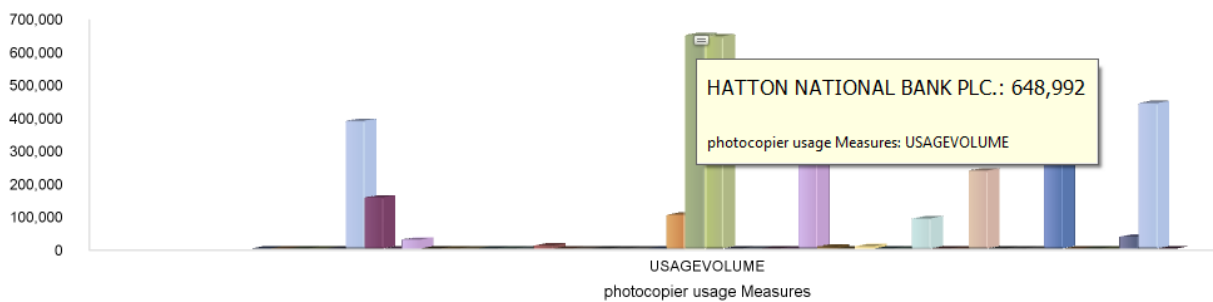

Figure 4-33: 2014 May Photocopy Usage Volume Variation Against Customers

Even though in the second quarter the highest usage volume records by Citizens development in 2014 May Hatton National Bank has had the highest usage volume which shows in Figure 4-33 recorded as 648,992 copies.
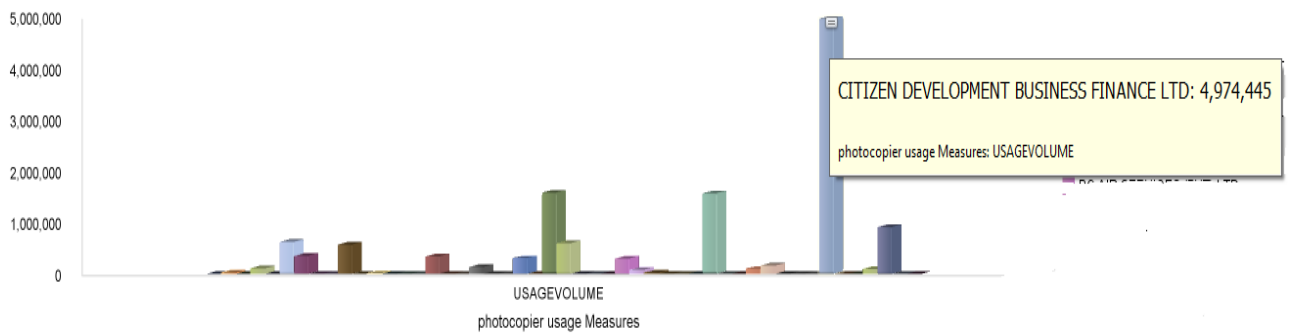

Figure 4-34: 2014 June Photocopy Usage Volume Variation Against Customers

Figure 4-34 illustrates the usage volume of 2014 June. Once again Citizen development has had the highest usage volume which is 4,974,445.
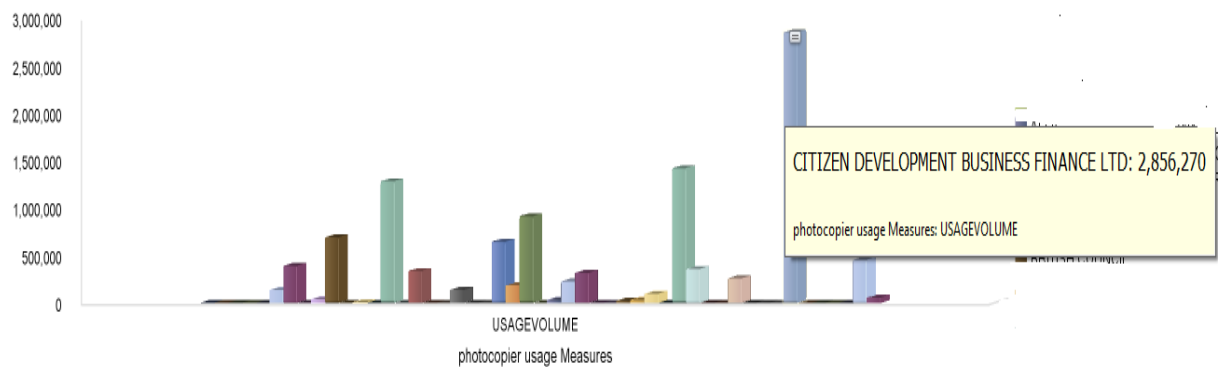

Figure 4-35: 2014 - Q3 Photocopy Usage Volume Variation Against Customers

Above Figure 4-35 consists of usage volume distribution of 2014 quarter 3. Like in quarter 2 in this quarter also highest volume usage recorded from Citizens Development Company. These usage volumes have been distributed in month of July and August.
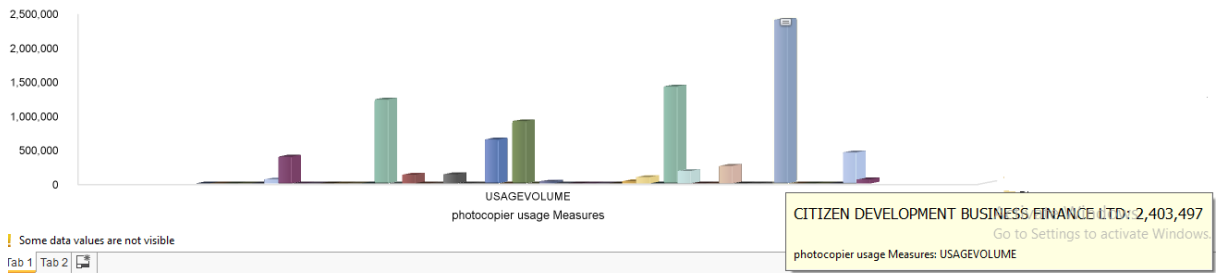
Figure 4-36: 2014 July Photocopy Usage Volume Variation Against Customers

Figure 4-26 illustrates photocopy usage volume in month of July 2014. In the first month of third quarter (July) as expected Citizens Development has recorded the highest usage volume which cis 2,403,497.
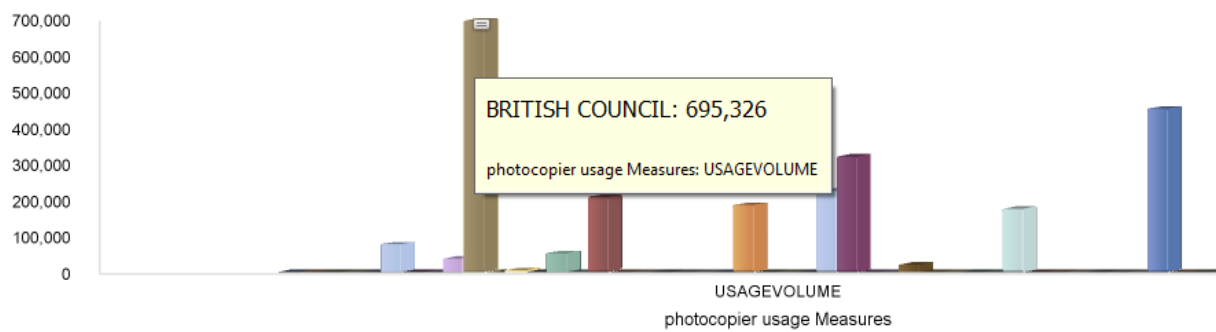

Figure 4-37: 2014 August Photocopy Usage Volume Variation Against Customers

Figure 4-27 illustrates photocopy usage volume in month of August 2014. Even though Citizens Development has highest usage volume in quarter 3, in the month of August British Council has recorded highest usage volume.
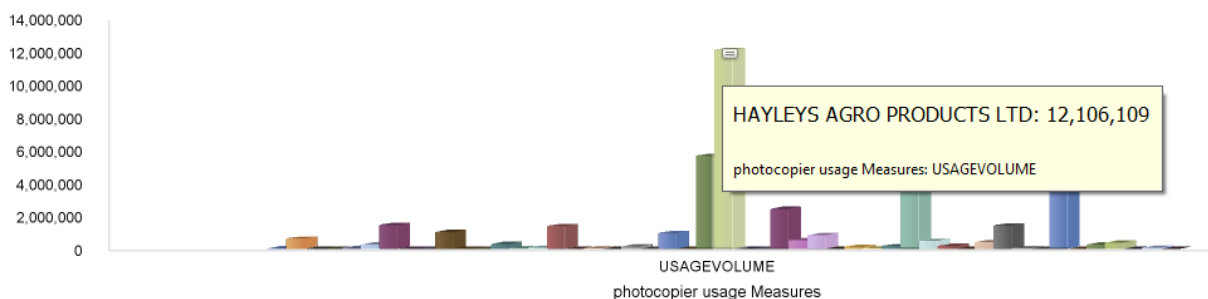

Figure 4-38: 2014 - Q4 Photocopy Usage Volume Variation Against Customers

54

Figure 4-38 illustrates quarter 4 usage volume in different customers. Hayleys Agro Products records highest usage volume which is 12,106,109. In the fourth quarter, considerable usage volumes have recorded in all three months.
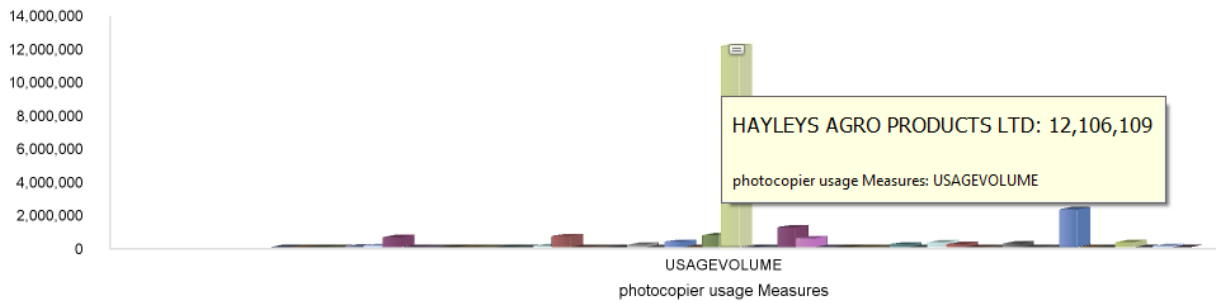

Figure 4-39: 2014 October Photocopy Usage Volume Variation Against Customers

Figure 4-39 illustrates photocopy usage volume in month of October 2014. In the month of October as recorded in quarter 4 Hayleys Agro has the highest usage volume.
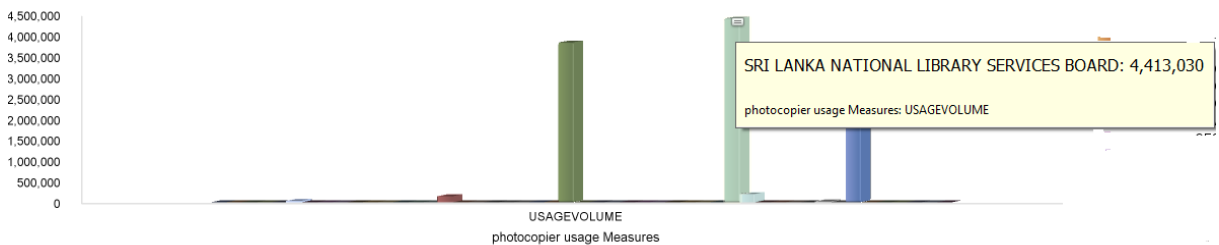

Figure 4-40: 2014 November Photocopy Usage Volume Variation Against Customers

But in November 2014 (Figure 4-40) Sri Lanka National Library Board shows highest usage volume.
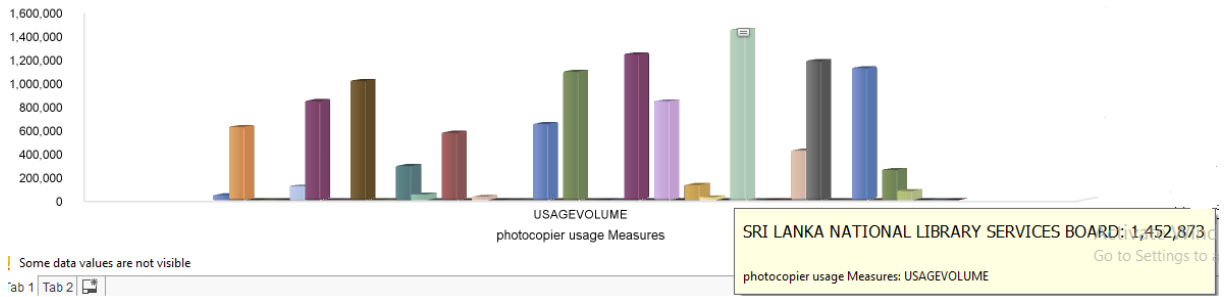
Figure 4-41: 2014 December Photocopy Usage Volume Variation Against Customers

Figure 4-41 illustrates photocopy usage volume in month of December 2014.The month of December shows significance difference compared to other months. There are several companies which have considerable amount of usages which is close to highest values in December 2014.

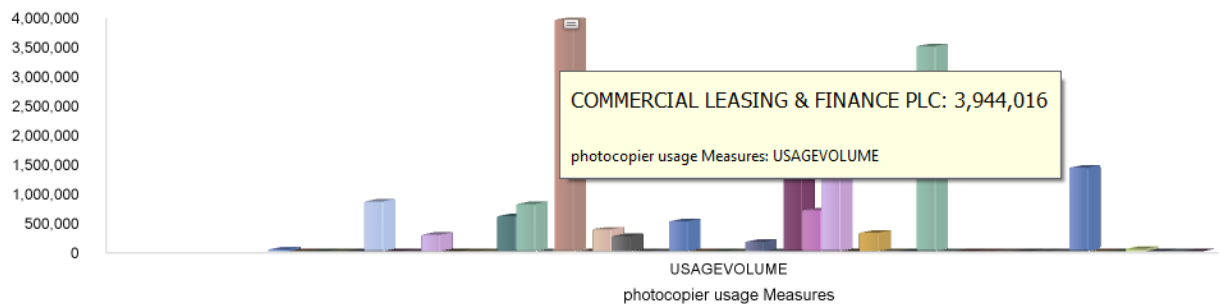Figure 4-42: 2015 - Q1 Photocopy Usage Volume Variation Against Customers

Figure 4-42 illustrates photocopy usage volume in first quart of 2015. In 2015 Commercial Leasing and Finance has the highest usage volume which is 3,944,016. Unlike in 2014 in 2015 all three months in first quarter shows considerable usage volumes.
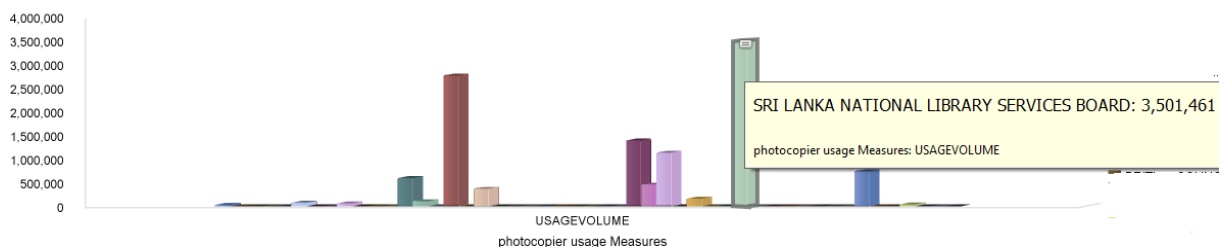

Figure 4-43: 2015 January Photocopy Usage Volume Variation Against Customers

Figure 4-43 shows the highest usage volume in January 2015 which is the first month of first quarter. Even though Commercial Leasing has highest usage volume in the month of January highest usage volume shows in Sri Lanka National Library Services.

Figure 4-44: 2015 February Photocopy Usage Volume Variation Against Customers

Figure 4-44 illustrates photocopy usage volume in month of February 2015. In the month of February as expected Commercial Leasing has recorded the highest usage volume which is 1,165,978.
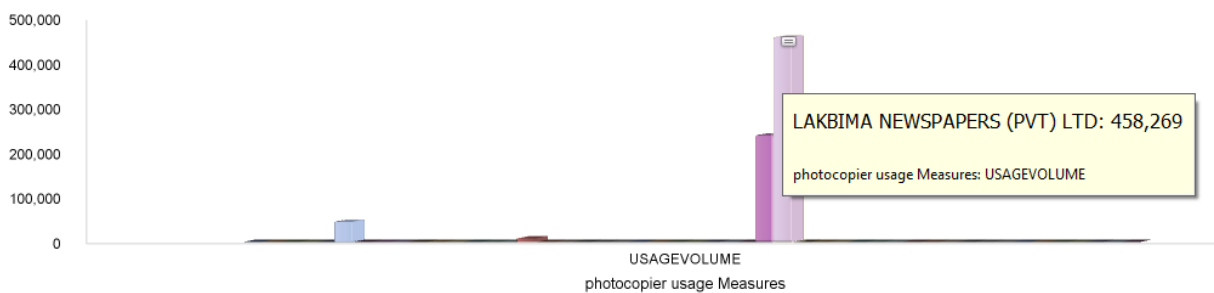
Figure 4-45: 2015 March Photocopy Usage Volume Variation Against Customers

Figure 4-45 illustrates the usage volume in March 2015. Lakbima Newspapers has highest usage volume. This quarter has three different highest usage volumes in different three months which is not shown in the year 2014.
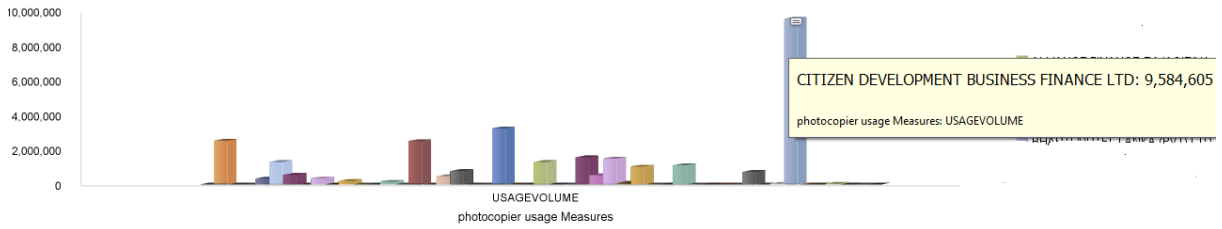
Figure 4-46: 2015 - Q2 Photocopy Usage Volume Variation Against Customers

Figure 4-46 illustrates the usage volume in the second quarter. As in 2014 second quarter in 2015 second quarter Citizens Development has recorded the highest usage volume and all three months in quarter 3 shows significance usage volumes.
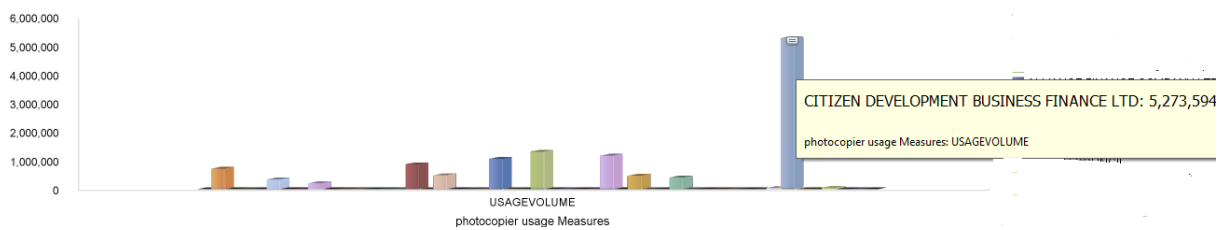
Figure 4-47: 2015 April Photocopy Usage Volume Variation Against Customers

Figure 4-47 illustrates the usage volume of April 2015. As recorded in quarter 2 in the month pf April as well Citizens development records highest usage volume which is 5273594.

Figure 4-48: 2015 May Photocopy Usage Volume Variation Against Customers

Usage volumes in May 2015 illustrates in Figure 4-48. In this month, new company has recorded a highest usage volume which is not recorded earlier quarters. Hidramani Mercury Apparel in Katunayake has recorded highest usage volume of 1,569390 in May 2015.

CITIZEN DEVELOPMENT BUSINESS FINANCE LTD: 3,244,166
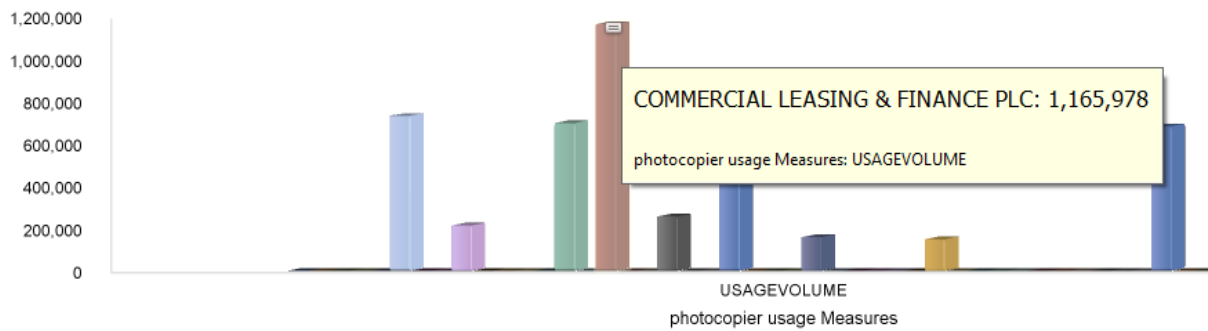
photocopier usage Measures: USAGEVOLUME

Figure 4-49: 2015 June Photocopy Usage Volume Variation Against Customers

Figure 4-49 illustrates photocopy usage volume in month of June 2015. In the last month of quarter 2 June 2015 again Citizens Development has had the highest usage volume.

CITIZEN DEVELOPMENT BUSINESS FINANCE LTD: 8,197,325

photocopier usage Measures: USAGEVOLUME
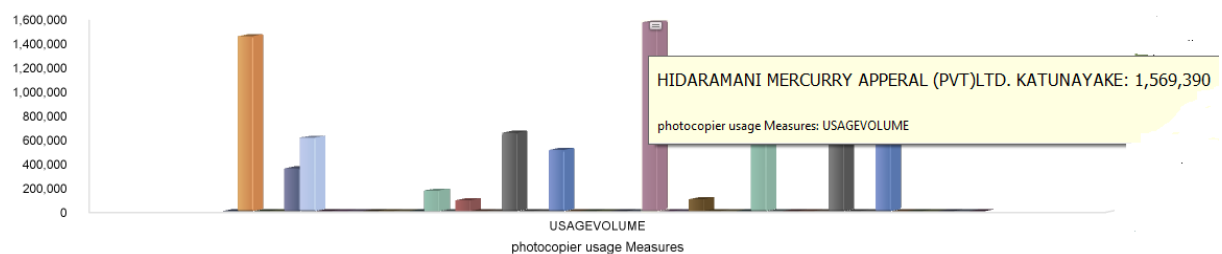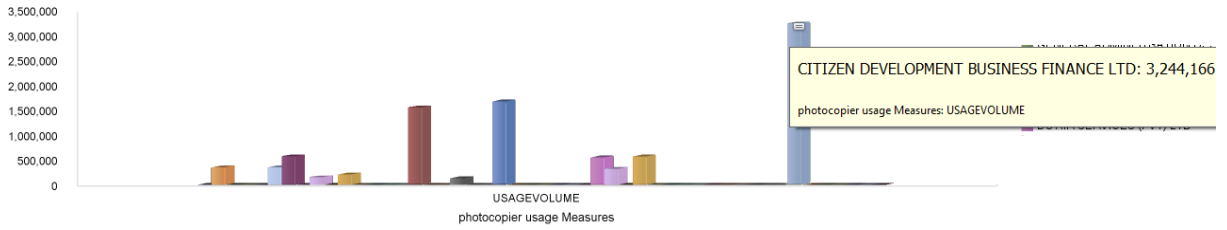
Figure 4-50: 2015 - Q3 Photocopy Usage Volume Variation Against Customers

Figure 4-50 consists of usage volumes of different customers in 2015 3rd quarter. As in 2014 third quarter in 2015 Citizens development shows highest usage volume of 8197325.

CITIZEN DEVELOPMENT BUSINESS FINANCE LTD: 5,515,241

photocopier usage Measures: USAGEVOLUME

Figure 4-51: 2015 July Photocopy Usage Volume Variation Against Customers

Figure 4-51 illustrates photocopy usage volume in month of July 2015. In the first month of third quarter once again Citizens development has the highest usage volume.

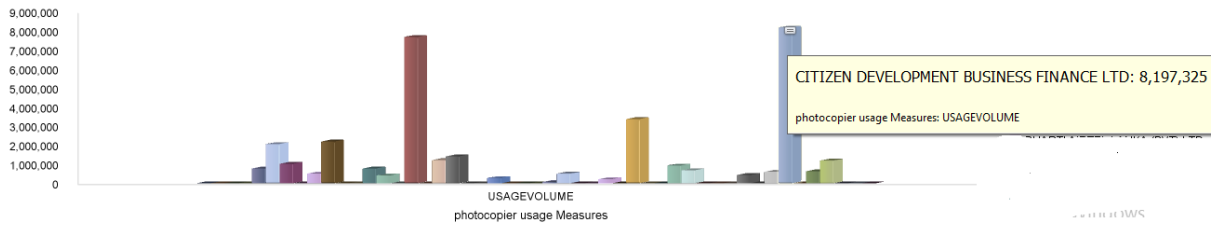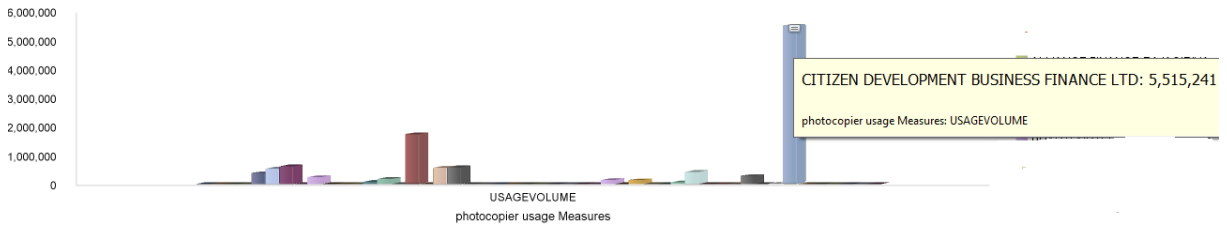COMMERCIAL LEASING & FINANCE PLC: 3,374,682
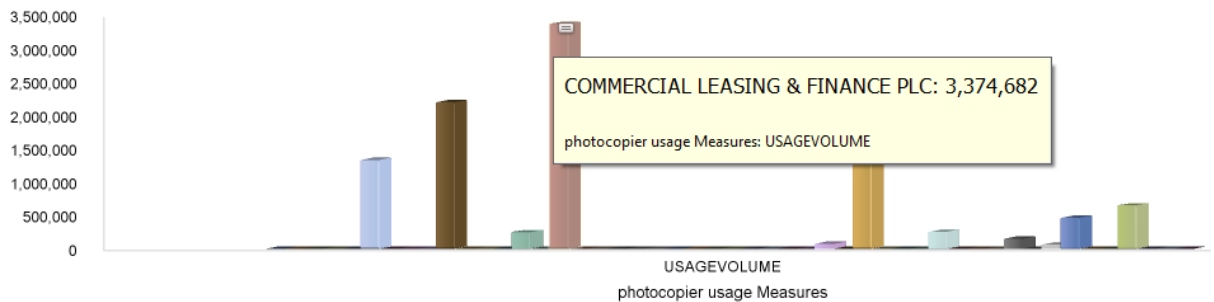
photocopier usage Measures: USAGEVOLUME

Figure 4-52: 2015 August Photocopy Usage Volume Variation Against Customers

As shown in Figure4-52 and Figure 44 Commercial Leasing has shown highest usage volume in months of August and September. But which is comparatively low to month of July.

Figure 4-53: 2015 September Photocopy Usage Volume Variation Against Customers

Figure 4-53 illustrates photocopy usage volume in month of September 2015. Commercial Leasing has recorded the highest usage volume which is 2,583,231.

Figure 4-54: 2015 - Q4 Photocopy Usage Volume Variation Against Customers

Above Figure 4-54 shows fourth quarter of usage volume in 2015. As in second and third quarter in this year in as well as in fourth quarter citizen development has the highest usage volume which is much higher value 11572348.

Figure 4-55: 2015 October Photocopy Usage Volume Variation Against Customers

Figure 4-55 illustrates photocopy usage volume in month of October 2015. In the month of October there are several more companies show considerable usage volumes but Citizen Development has the highest usage volume.
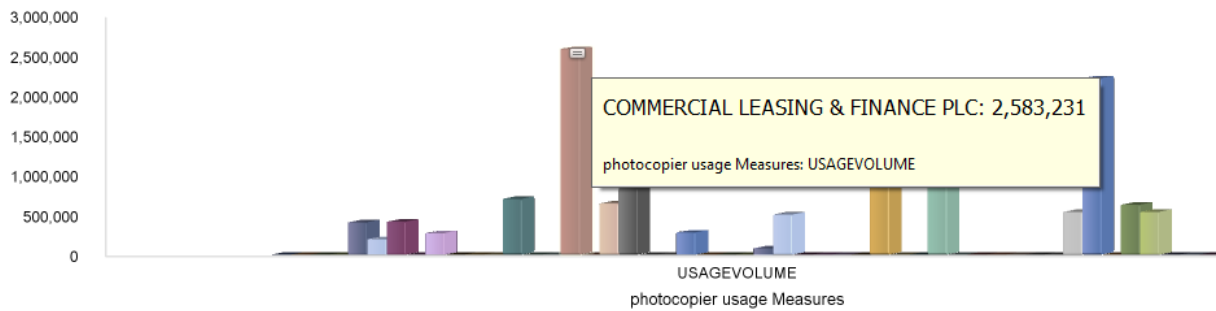
Figure 4-56: 2015 November Photocopy Usage Volume Variation Against Customers

Figure 4-56 illustrates photocopy usage volume in month of November 2015. November and December 2015 once again Citizens Development has the highest usage volume. Since Citizen Development is a company which falls under fiancé sector evolves with lots of paper work. This would be the reason for highest usage volumes highlights from this customer.
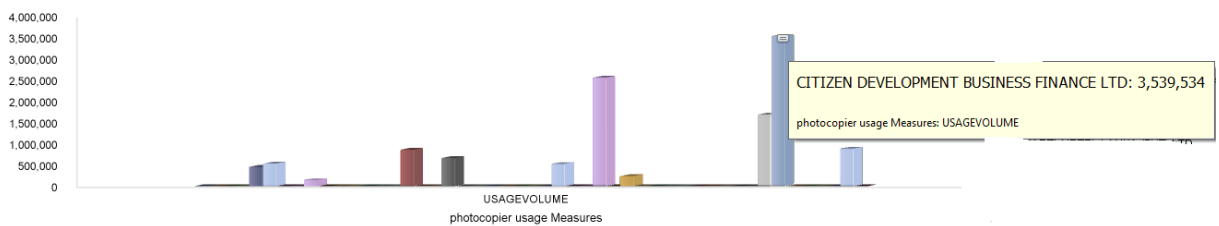
Figure 4-57: 2015 December Photocopy Usage Volume Variation Against Customers

As shown in Figure 4-57 as usual there are several more customers which has usage volumes much closer to highest usage volume since December is the seasonal month lots of marketing works carrying out which lead to more paper work.
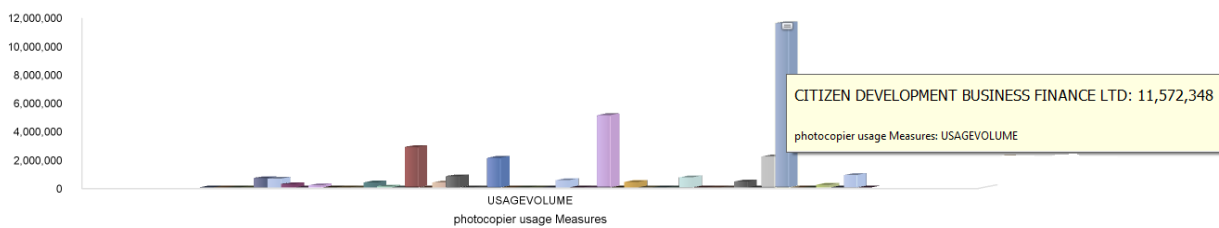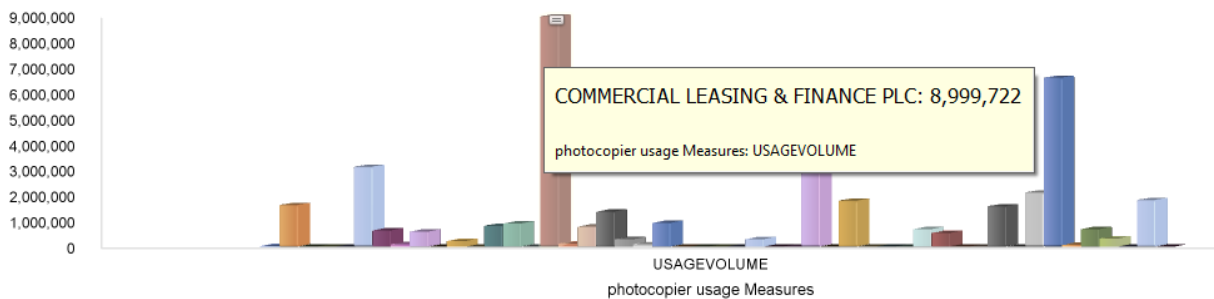

Figure 4-58: 2016 - Q1 Photocopy Usage Volume Variation Against Customers

Above Figure 4-58 shows the 2016 quarter 1 usage volumes. Commercial Leasing has been shown the highest usage volume while citizen Development has recorded second highest which records highest usage volumes in previous years.
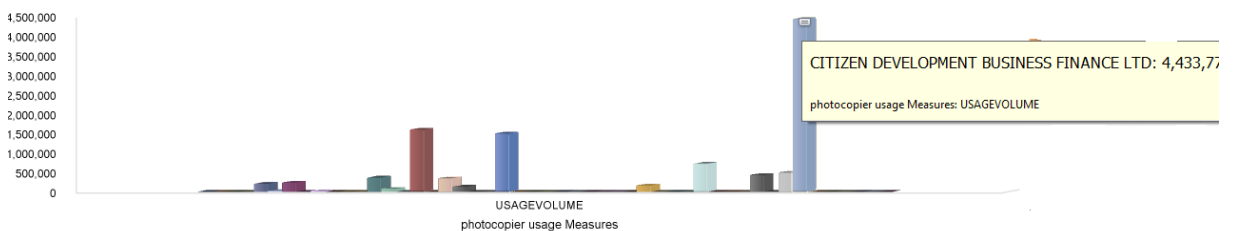
Figure 4-59: 2016 January Photocopy Usage Volume Variation Against Customers

Figure 4-59 illustrates photocopy usage volume in month of January 2016. In the month, January, commercial Leasing has recorded the highest usage volume which is 2611219.

Figure 4-60: 2016 February Photocopy Usage Volume Variation Against Customers

As shown in Figure 4-60 like in January in February also commercial Leasing has the highest usage volume.

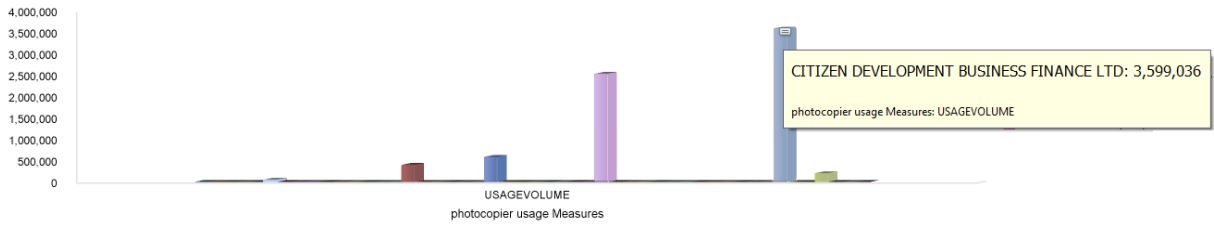Figure 4-61: 2016 March Photocopy Usage Volume Variation Against Customers

Figure 4-61 illustrates photocopy usage volume in month of March 2016. In March citizen Development, has recorded highest usage volume which is 2396202 but it is comparative low to month of January and February.
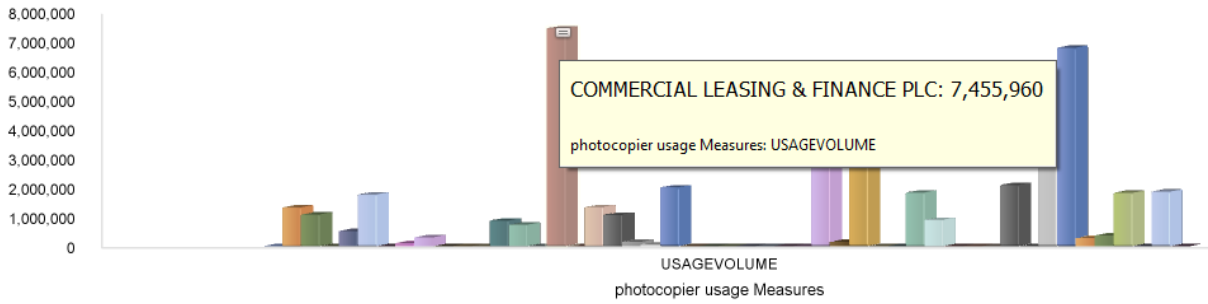
Figure 4-62: 2016 - Q2 Photocopy Usage Volume Variation Against Customers

Figure 4-62 illustrates 2015 second quarter usage volumes. Likewise, in first quarter in second quarter also commercial leasing has had the highest usage volume which is 7455966 and Citizen development has recorded much closer value to Commercial Leasing.

Figure 4-63: 2016 April Photocopy Usage Volume Variation Against Customers

Once again as shown in Figure 4-63 April 2015 Commercial Leasing has highest usage volume and Citizens development usage volume has dropped in April 2015.

Figure 4-64: 2016 May Photocopy Usage Volume Variation Against Customers

Figure 4-64 illustrates photocopy usage volume in month of May 2016. In month of May Commercial leasing has highest usage volume and Citizen Development has the second highest usage volume.
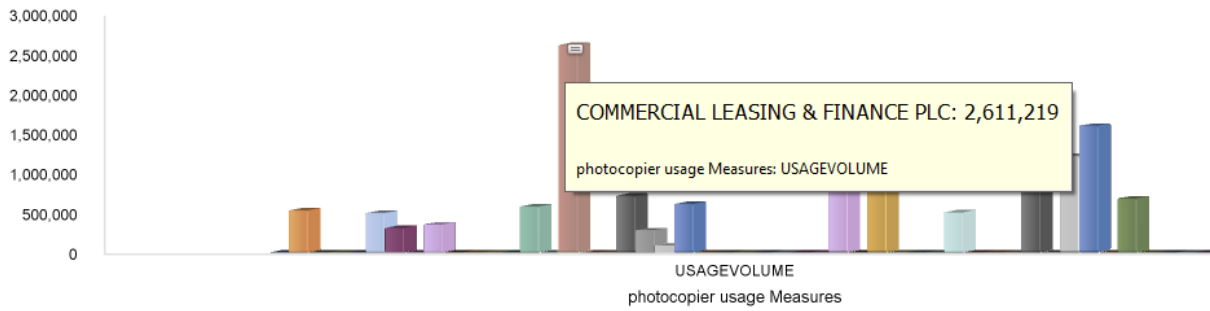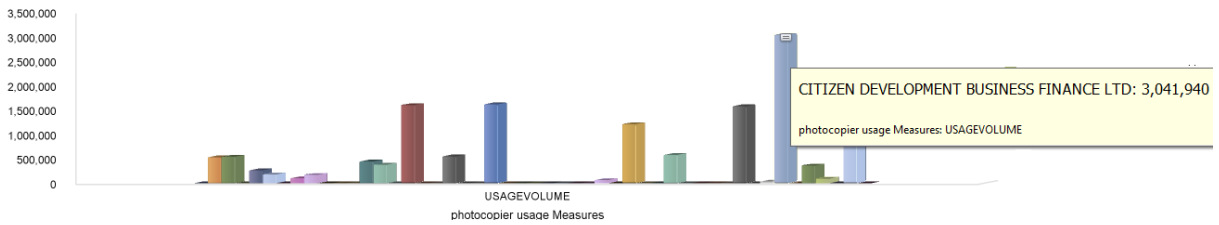
Figure 4-65: 2016 June Photocopy Usage Volume Variation Against Customers

Figure 4-65 illustrates photocopy usage volume in month of June 2016. Not like in April and May in June Citizen Development has the highest usage volume and Commercial Leasing has very low usage volume. But when consider about the whole quarter as one still Commercial Leasing has the highest usage volume.
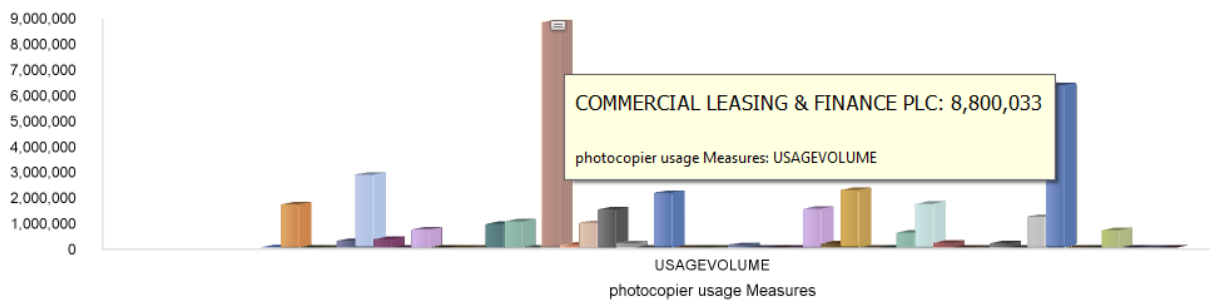
Figure 4-66: 2016 - Q3 Photocopy Usage Volume Variation Against Customers

Figure 4-66 illustrates the usage volumes of different customers in third quarter 2016. Like in first and second quarters in 2016 in the third quarter also Commercial Leasing has used high number of photocopies and citizen development is the second highest like in first and second quarters.
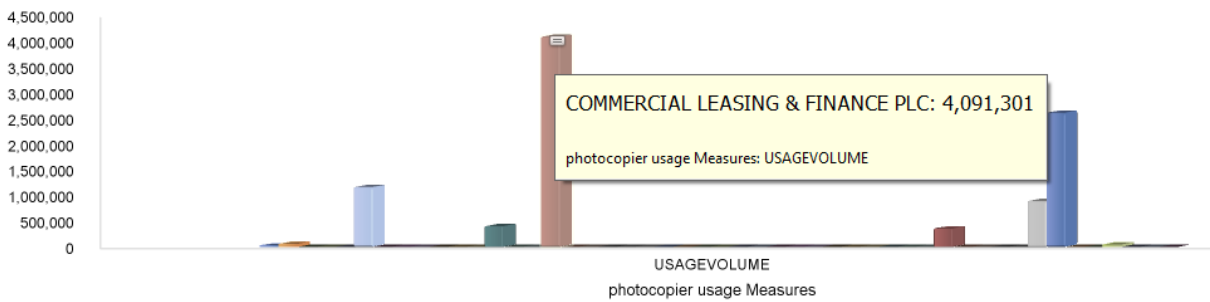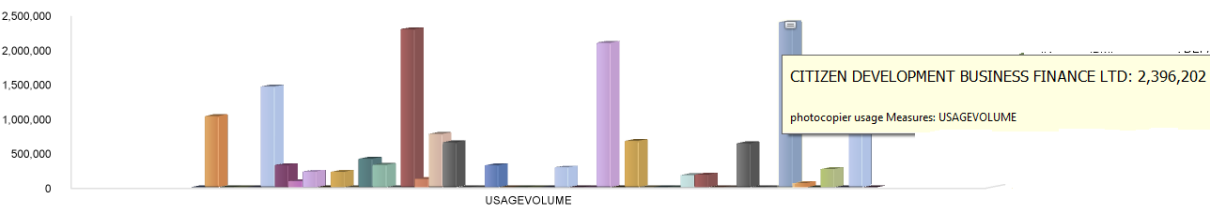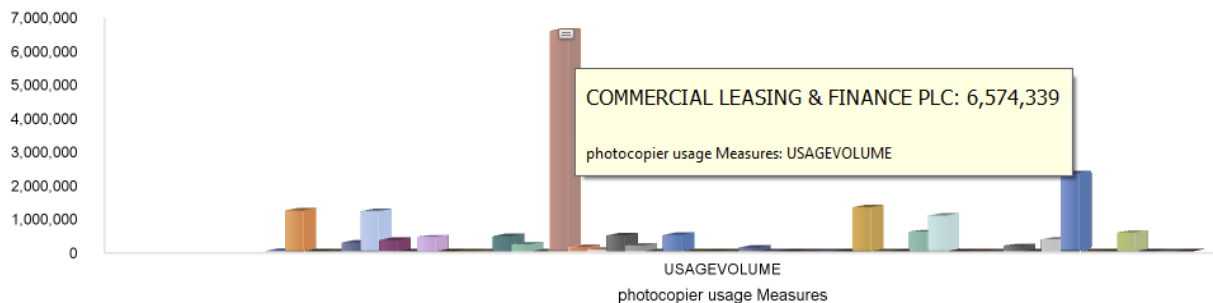

Figure 4-67: 2016 July Photocopy Usage Volume Variation Against Customers

As shown in Figure 4-67 in July Commercial Leasing has highest usage volume mean while citizen Development has second highest volume but it is very low compared to other months in 2016.
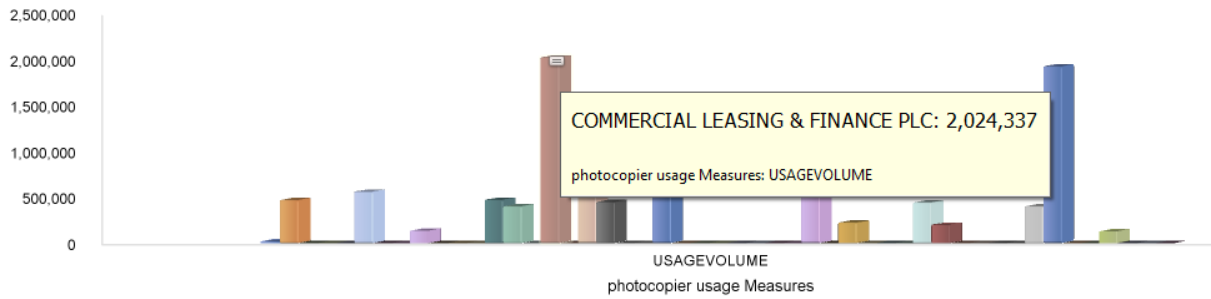
Figure 4-68: 2016 August Photocopy Usage Volume Variation Against Customers

Figure 4-68 illustrates photocopy usage volume for the month of August 2016. In August, also Commercial Leasing has highest usage volume and citizen Development has second highest as usual in 2016 usage volume behavior which is much close to Commercial Leasing.
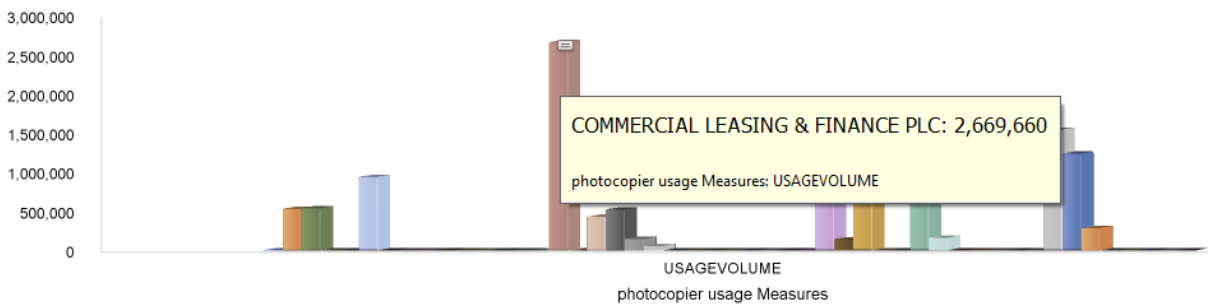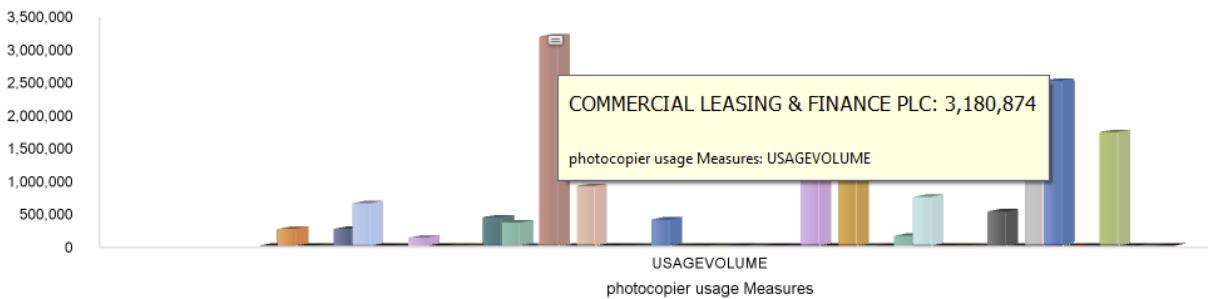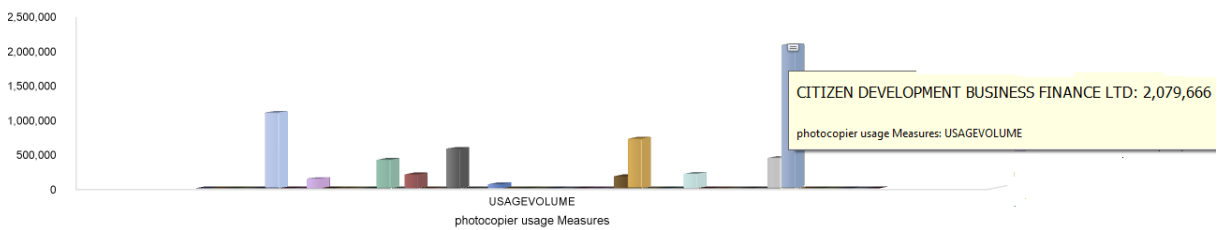

Figure 4-69: 2016 September Photocopy Usage Volume Variation Against Customers

Figure 4-69 illustrates photocopy usage volume for the month of September 2016. In September 2016 Citizen Development has highest photocopy usage.

Photocopier usage can vary from customer to customer based on their business. But having a clear picture of behavior photocopy usage volume in different customers helps to increase company profits. The evaluation of photocopy usage against has done from 2012-2016. The summarization of above drill down analysis shows below.

From the above drill -down analysis highlighted three customers for their high photocopy usage. Next step is to analyze those three customers in a detail view.
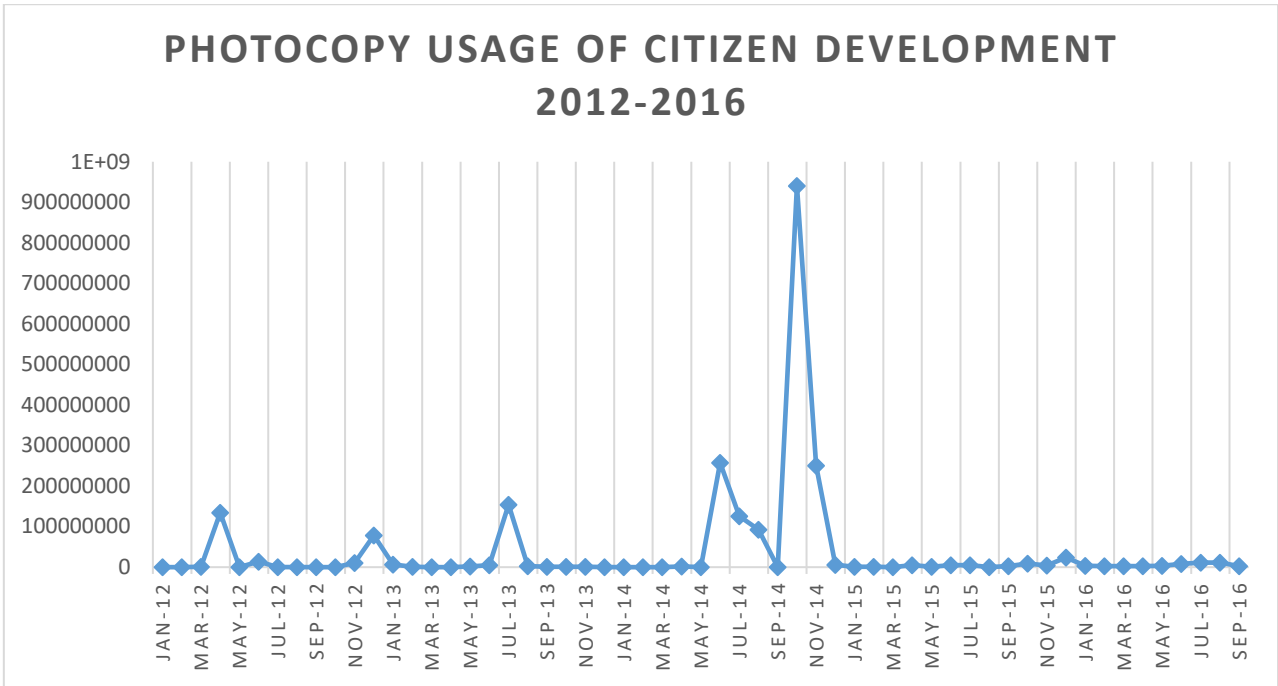
Figure 4-70: Photocopy Usage Volume of Citizen Development 2012-2016

Figure 4-70 illustrates the photocopy usage of Citizen Bank from 2012 -2016 September. Per figure 4-70 Citizen Development Bank has the highest photocopy usage in October 2014.
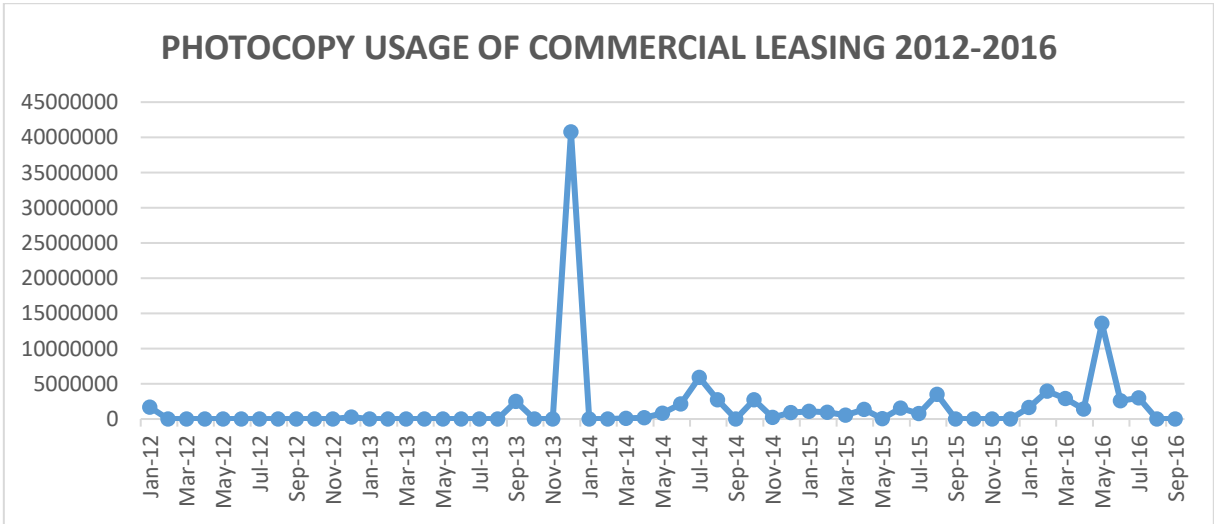
Figure 4-71: Photocopy Usage Volume of Commercial Leasing 2012-2016

Figure 4-71 shows the photocopy usage volume of commercial Leasing from 2012 – 2016 September. The highest usage shows in December 2013. Month of December is highlighted in every year.
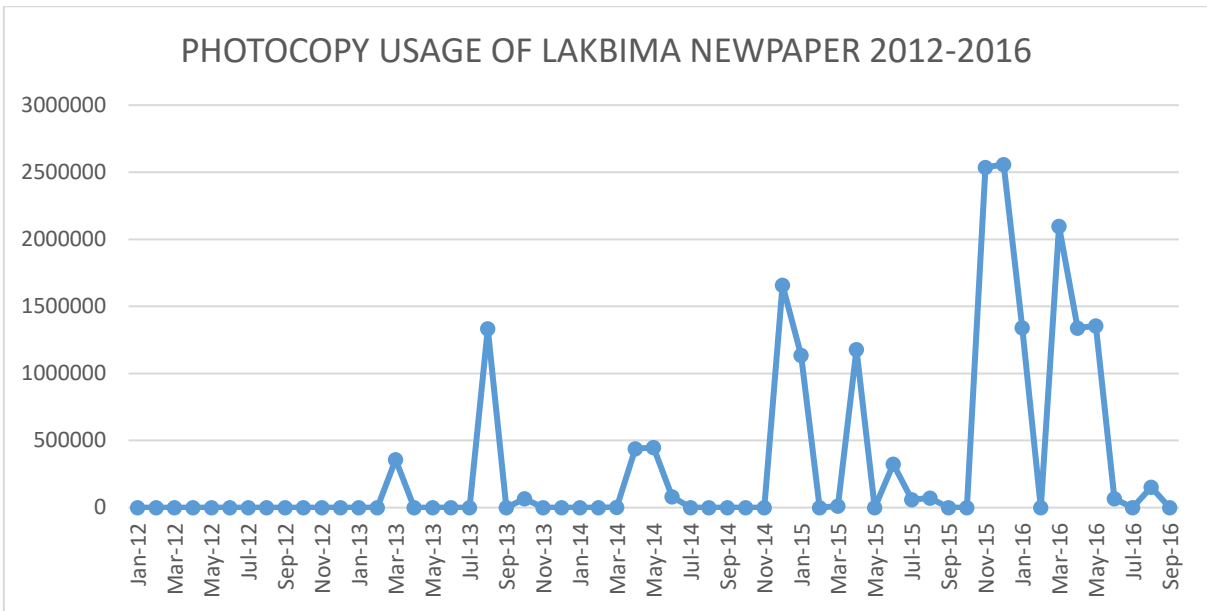
Figure 4-72: Photocopy Usage Volume of Lakbima Newspapers 2012-2016

Figure 4-72 shows usage volume of Lakbima newspapers from 2012-2016 September. Lakbima Newspapers has low profile in 2012. But in 2015-2016 has variations both up and down.

These three customers Commercial Leasing, Citizen Development and Lakbima Newspapers are customers who involve with lots of paper work. Citizen Development and commercial Leasing are in financial sector while Lakbima Newspaper involve with newspapers industry. When comparing figure 4-70 and 4-71 both are in same industry and both figures shows significant positive variation specially in second and fourth quarters. In the period of March, April and November, December they have high usage of photocopies.

Lakbima Newspapers usage is comparatively low to Commercial Leasing and Citizen Development. But unless in financial sector customers Lakbima Newspapers do not have seasonal trends. They have significant amount of usage in every month.

Table 4-1: Summary of Photocopy Usage Volumes and Customers

Year	Quarter	Month	Customer
2012	Q1	January	Hidaramani Mercurry Apperal (Pvt)Ltd. Katunayake
		February	-
		March	Citizen Development Business Finance Ltd.
2012	Q2	April	Citizen Development Business Finance Ltd.
		May	Citizen Development Business Finance Ltd.
		June	Citizen Development Business Finance Ltd.
2012	Q3	July	Citizen Development Business Finance Ltd.
		August	Hidaramani Mercurry Apperal (Pvt)Ltd. Katunayake
		September	Hatton National Bank PLC
2012	Q4	October	-
		November	Sri Lanka National Library Services Board
		December	Citizen Development Business Finance Ltd.
2013	Q1	January	Citizen Development Business Finance Ltd.
		February	Citizen Development Business Finance Ltd.
		March	Lakbima Newspapers (Pvt) Ltd
2013	Q2	April	-
		May	Citizen Development Business Finance Ltd.
		June	Citizen Development Business Finance Ltd.
2013	Q3	July	Citizen Development Business Finance Ltd.
		August	Hatton National Bank PLC
		September	Commercial Leasing and Finance PLC
2013	Q4	October	British Council
		November	Citizen Development Business Finance Ltd.
		December	Commercial Leasing and Finance PLC
2014	Q1	January	-
		February	-
		March	Lakbima Newspapers (Pvt) Ltd
2014	Q2	April	Citizen Development Business Finance Ltd.
		May	Hatton National Bank PLC

		June	Citizen Development Business Finance Ltd.
2014	Q3	July	Citizen Development Business Finance Ltd.
		August	British Council
		September	-
2014	Q4	October	Hayleys Agro Products Ltd
		November	Sri Lanka National Library Services Board
		December	Sri Lanka National Library Services Board
2015	Q1	January	Sri Lanka National Library Services Board
		February	Commercial Leasing and Finance PLC
		March	Lakbima Newspapers (Pvt) Ltd
2015	Q2	April	Citizen Development Business Finance Ltd.
		May	Hidaramani Mercurry Apperal (Pvt)Ltd. Katunayake
		June	Citizen Development Business Finance Ltd.
2015	Q3	July	Citizen Development Business Finance Ltd.
		August	Commercial Leasing and Finance PLC
		September	Commercial Leasing and Finance PLC
2015	Q4	October	Citizen Development Business Finance Ltd.
		November	Citizen Development Business Finance Ltd.
		December	Citizen Development Business Finance Ltd.
2016	Q1	January	Commercial Leasing and Finance PLC
		February	Commercial Leasing and Finance PLC
		March	Citizen Development Business Finance Ltd.
2016	Q2	April	Commercial Leasing and Finance PLC
		May	Commercial Leasing and Finance PLC
		June	Citizen Development Business Finance Ltd.
2016	Q3	July	Commercial Leasing and Finance PLC
		August	Commercial Leasing and Finance PLC
		September	Citizen Development Business Finance Ltd.

Table 4-1 illustrates customers who have the highest number of photocopy usage. When consider about the customers there are two main customers who have highlighted throughout the years. They are Commercial Leasing and Finance PLC and Citizen Development Business Finance Ltd. These two customers are in financial sector these customers invloce with lots of paper work. This would be the reason to have high usage volumes from these two customers.

As shown in figures 4-70 and 4-71 April and December months highlighted in the analysis. Since the both the customers in fiancial sector April is the month of end of financial clousre so most of the paper work handle within this month. December is the year end because of seasonal month most of the work handle during this period.

When promoting the photocopiers should focus on customers who are work on financial sector and media because these two categroies mostly invlove with paper work with their day to day work.

Month of March, April and December would be good months to marketing for these type of customers.

In the drill down analysis, mid year (June and July) had high usage of photocopier machines. It was highlighted through out the period of 2012-2016. From the analysis found out customers available in the dataset are from banks, finance sector, media, apperal, education and many other industries. By focussing on customers who invlove with lots of paper work in their day to day activities company could increase their profits.

4.2 Key Evaluation and Findings – K-Means Analysis

The main objective of this project is to identify fast moving consumable items. In-order to identify fast moving items it is necessary to apply data mining technique because of the complexity of data. K-means data mining technique has applied to identify fast moving items. The reason to select K-Means data mining technique was K-Means is a technique which clustered items with similar characteristics. If a set of items to be fast moving there should be similar characteristics in each item in one cluster. Furthermore, justification for selecting K-Means data mining technique has discussed in the literature review chapter.

In-order to apply K-means technique it is necessary to identify most suitable attributes to find out similarities in clusters. Figure 4-73 shows the attributes list which selected to perform K-Means.
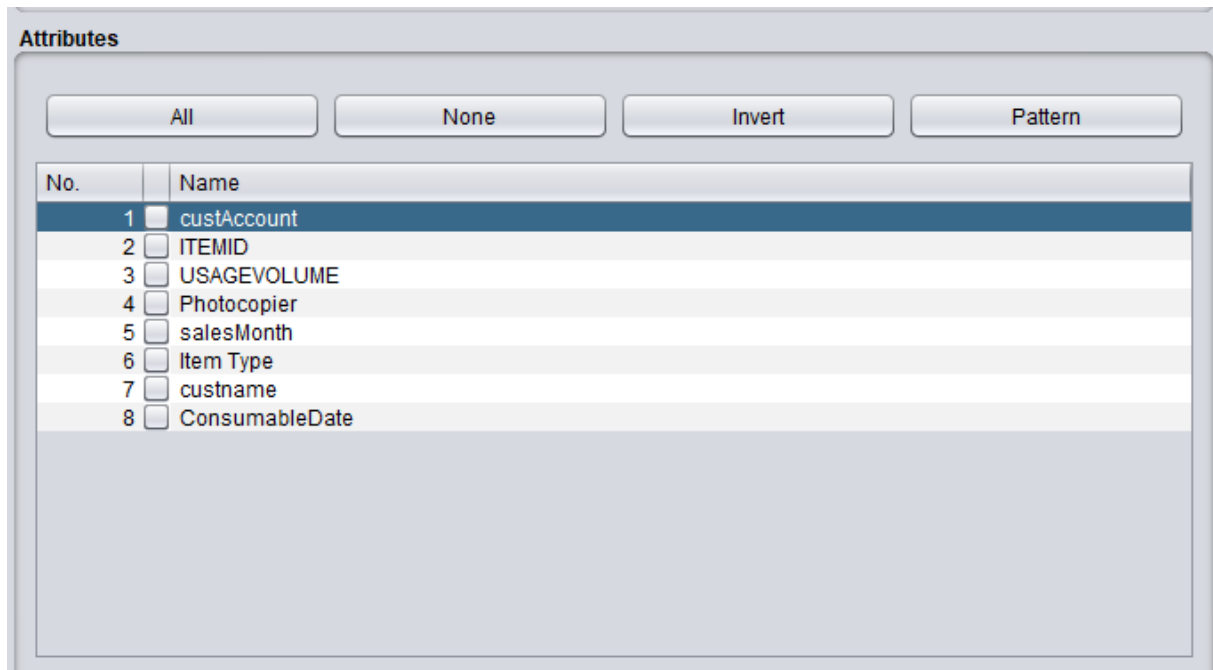
Figure 4-73: K-Means Attributes Selection

The detail description of attributes has done under the section 3.3.1

After selecting suitable attributes for the analysis, needs to preprocess data where suitable for the analysis. After preprocessing, the data is ready to analyze and next step in the analysis is, apply K-means technique to the dataset.

canopyMaxNumCanopiesToHoldInMemory	100
canopyMinimumCanopyDensity	2.0
canopyPeriodicPruningRate	10000
canopyT1	-1.25
canopyT2	-1.0
debug	False
displayStdDevs	False
distanceFunction	Choose **EuclideanDistance** -R fir:
doNotCheckCapabilities	False
dontReplaceMissingValues	False
fastDistanceCalc	False
initializationMethod	Random
maxIterations	500
numClusters	3
numExecutionSlots	1
preserveInstancesOrder	False
reduceNumberOfDistanceCalcsViaCanopies	False
seed	10

Figure 4-74: K-Means Properties

Figure 4-74 shows the property setup for K-Means analysis. Euclidean distance method has applied to analyze the dataset. The K value has been set to three. K value is the number of clusters that set to perform K -Means analysis for the dataset. Before finalized the K value to three different K values has used such as two, three, four up to ten to find out most significant clusters. But most significant K value was three when K value is three, it has high intra cluster similarity and low inter cluster similarities.
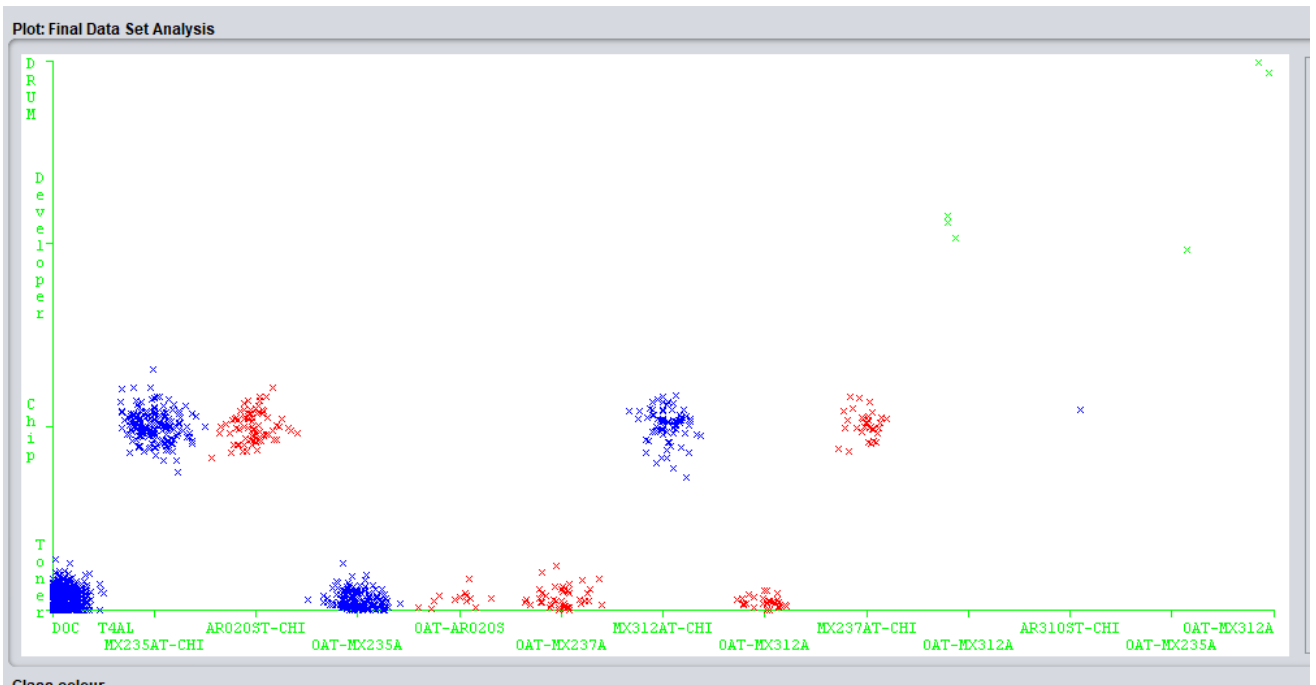
Figure 4-75: K Means Analysis - Item Types Against Items

Figure 4-75 shows K-means analysis which has applied for the dataset against item type and item id. For the Y axis, item type has set and for X axis item id has set. Cluster indicates by three different colors such as blue, red and green. As shown in figure 4-75 toners and chips has fallen into blue and red clusters while drums and developers has fallen one common cluster which is green. DOC T4AL, OAT-MX235AT, MX235AT-CHIP, MX312AT-CHIP and AR310ST-CHIP chips are in one cluster which is blue and other chips and toners are in red clusters.
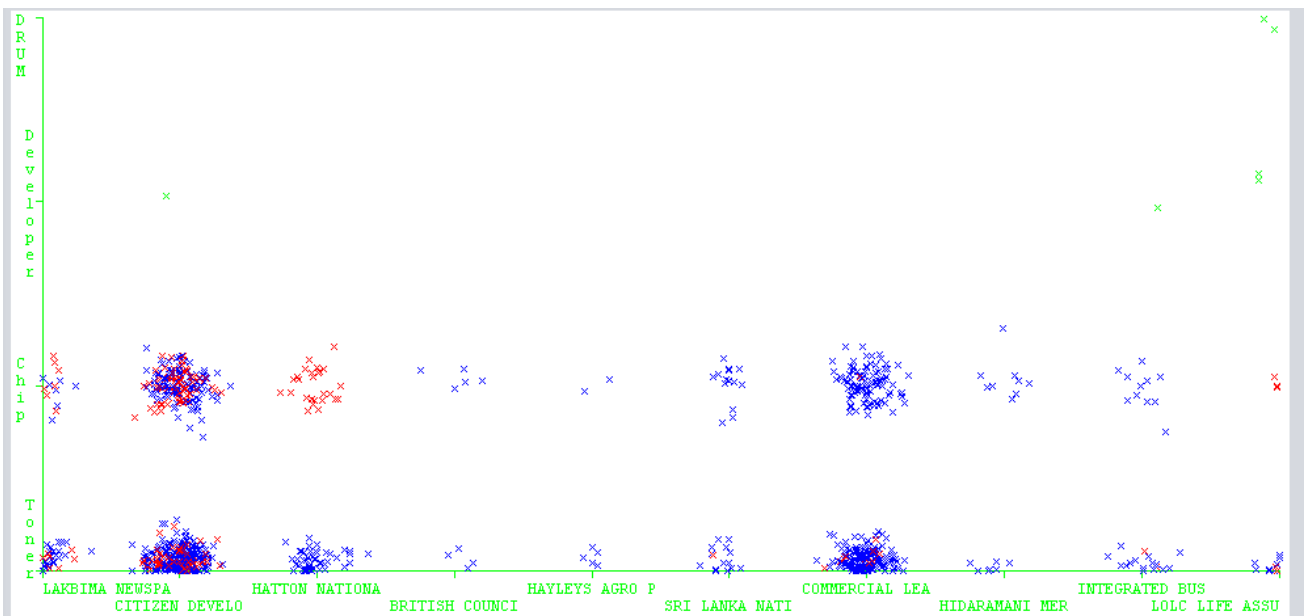
Figure 4-76: K Means Analysis - Item Types Against Customers

73

Figure 4-76 shows K- Means cluster analysis for item type against customer. Cluster types indicate by different three colors. Like in Figure 4-75, chips and toners has clustered into blue color and red color clusters. Citizen development bank and Commercial Leasing bank clusters have more points compared to other customers. From this emphasizes these two customers has had more consumable items for their photocopier machines compares to other customers. Developers and drums have not much involved with any of the customers. Both developers and drums consisted in same cluster which is the green cluster.
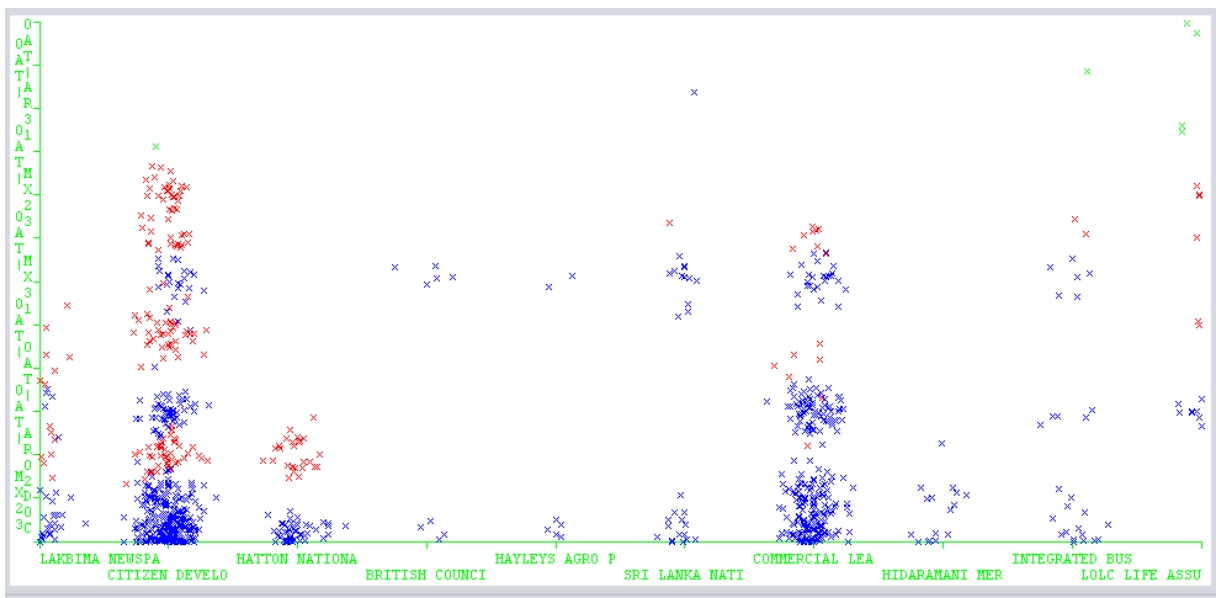
Figure 4-77: K Means Analysis - Items Against Customers

Figure 4-77 illustrates K-Means analysis of items against customers. Cluster types indicate by different three colors such as blue, red and green. As shown in figure 4-76 Citizen Development bank and Commercial Leasing have more cluster points compared to other customers. Both Citizen Development and Commercial Leasing have red color and blue color clusters. There is significant amount of cluster points are in both the clusters. Commercial Leasing has more blue color clusters and all the cluster points from toners and chips. Citizen Development has both red color cluster and blue color cluster. More cluster points belong to DOC T4AL toner type.

74
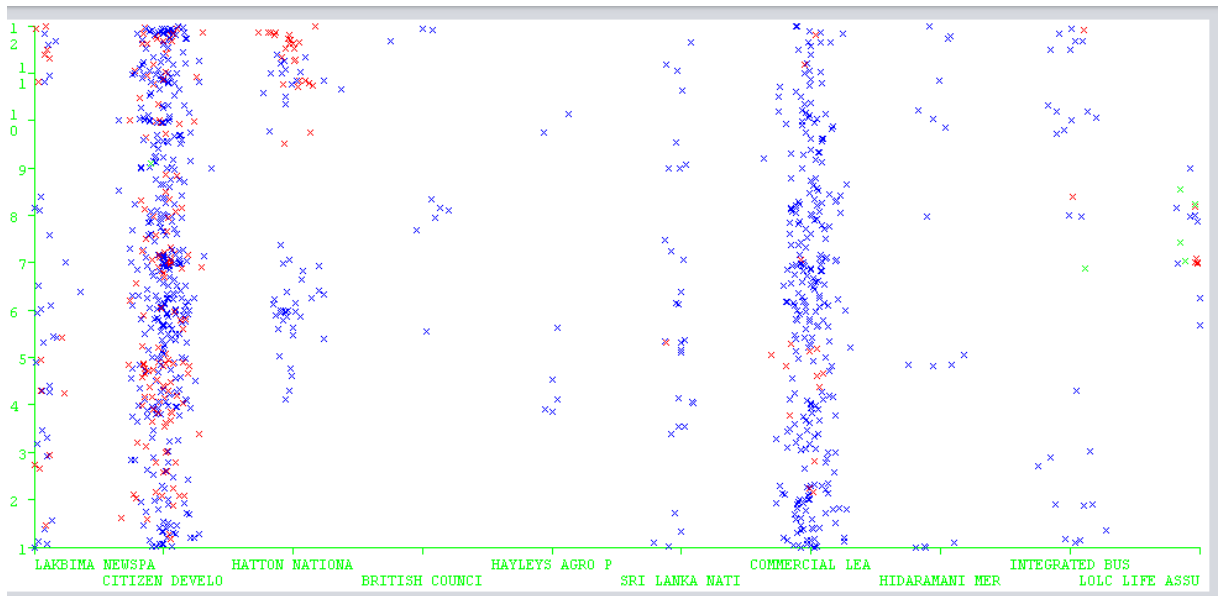
Figure 4-78: K Means Analysis – Sales Months Against Customers

Figure 4-78 illustrates the K-Means analysis of customers against sales month. Like in figure 4-77 In figure 4-78 Commercial Leasing and Citizen development bank highlight with more cluster points. Blue color cluster more oriented towards Commercial Leasing. Citizen Development has mixed cluster points of red and blue. Green color cluster points show very minimal impact towards customers and sales month. Cluster types indicate from three different colors. Month of December, June and April show significant cluster point movement.

Figure 4-79: K Means Analysis – Sales Months Against Items

75

Figure 4-79 demonstrates K-means visualization of items against items sales month. Cluster type indicate by different colors. Chips and toners have oriented towards blue and red color clusters while drums and developers oriented towards green color cluster. Significant of cluster points can see in months of April, June and December. DOC T4AL has more cluster points. Toner and chips movement can see throughout all the months.
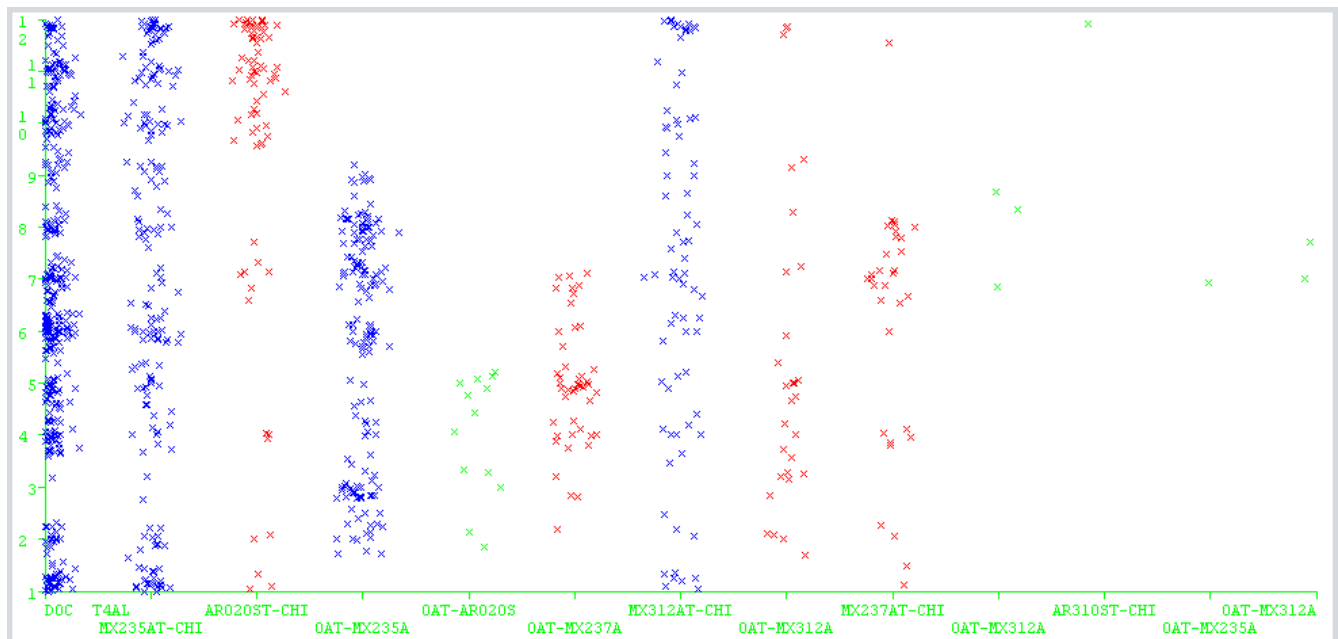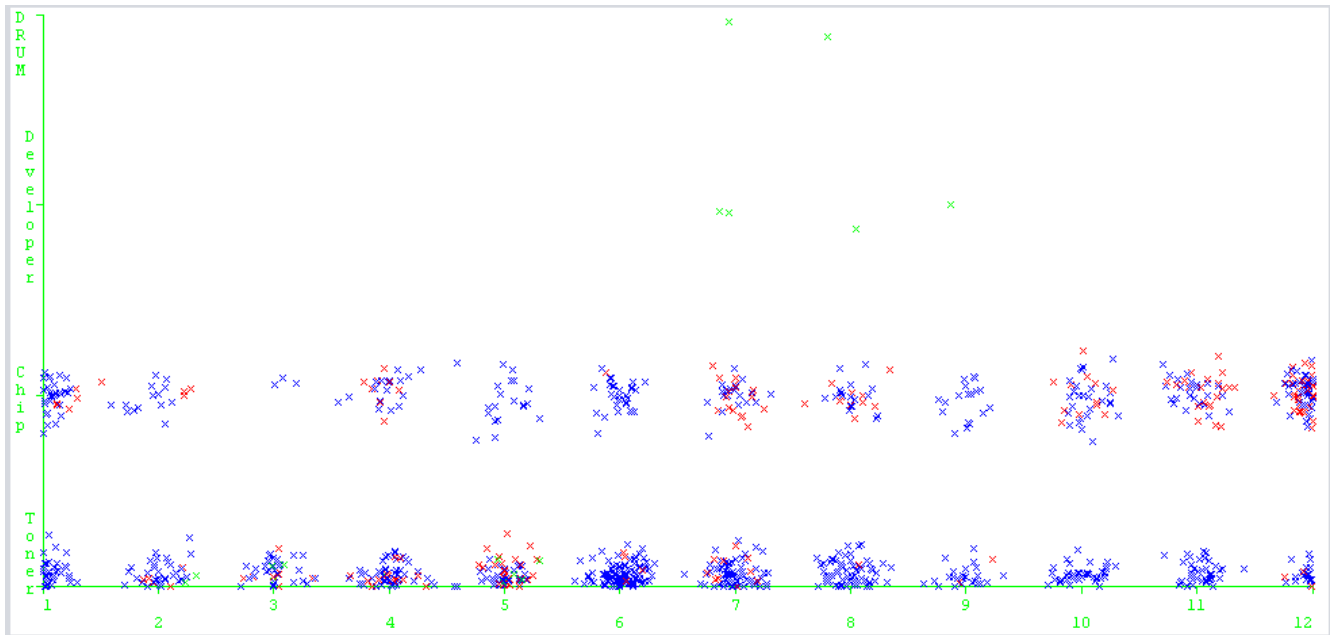

Figure 4-80: K Means Analysis – Sales Months Against Item Types

Figure 4-80 illustrates K-Means visualization in sales months against item types. Cluster color denotes from three different colors. As highlighted in previous figures in this also toner and chips has mixed in blue and red cluster while drums and developers are in green color cluster. Blue color and red color cluster points can see mostly in June and December. High toner movement shows in month of June. High chip movement shows in month of December.

The movement of items has measured based on item type, customer, sales month and usage volume of photocopier machines. From that could capture a feasible solution for fast moving items. As per the results highly moved items are chips and toners which is true because the increment of photocopies the consumable items like toners and chips also teared.

In the section of understanding data, author has used different attributes in-order to find out high usage photocopiers and fast moving items. But with the proper analysis based on the customer business environment and usage it has shown that business should focus more on usage and based on usage decide the movement of items. Based on K-Means analysis DOC T4AL and MX235AT-CHIP shows significant movement for all the combination attribute

analysis. When consider about item type, toners and chips had significant movement compared to drums and developers.

To increase the sales in the company, company can focus on DOC T4AL and MX235AT-CHIP and do promotional programs for these two items focusing on customers who are specially in finance sector and media.

Also, company should focus on reduce purchasing consumable items like developers and drums. Because in the analysis developers and drums show minimal involvement. These two consumable items are less moving. By limiting purchase of these items can reduce the wastage could happened in the division.

Summary

Based on the business environment photocopier usage may get vary. Customer who are in financial sector and media are the customer who involve with lots of paper work. This factor was highlighted in the drill down analysis. From the analysis identified Commercial Leasing and Finance PLC and citizen Development Business Finance Ltd are the customers who are highly used photocopy machines. There two customers are falling to the category of Financial Sector. From the drill- down analysis also found out customers have high usage in photocopier machines in specially in second and fourth quarter. From the use of K-Means analysis fast moving items have been identified. From the analysis found out that DOC T4AL and MX235AT-CHIP are the consumable items that fall into fast moving item category. When consider about the item type toners and chips showed significant movement throughout the analysis while developers and drums showed minimal movement which fall to the category of dead stock.

Next chapter is the final chapter of this research it consists of overall summary of the research and future enhancements that can be applied to this research.

5.0 Chapter 5: Conclusion and Future Work

In the previous chapter discussed about key evaluation and results found out from the evaluation.

This is the last chapter of this thesis. This chapter concludes with the findings of the research which has discussed in the chapter 4. Also in this chapter, will be discussed about problems encountered during the time of this research and the steps took to resolve those problems. Furthermore, in this chapter will be discussed about future enhancements which can apply to the research.

5.1 Conclusion

The main objective of this research was to find out fast moving consumable items based on the combinational attributes such as photocopy usage volume, customer, sales month, item type etc. Apart from that, identification of patterns in photocopier usage based on different customers has carried out in the main research area of this project. The findings of the research have been discussed in chapter 4.

In chapter 4 initial demand forecasting has done based on above finding and has suggested business strategies to increase profits and reduce the wastage in the division.

From Chapter 4 able to identify customers who has high usage volume is different quarters and months. By drilling down the detail, it gave better picture of the behavior of the usage volumes in different businesses.

In-order to identify high volume usage and fast moving items, used the model that explained in section 3.3.4.

Commercial Leasing and Finance PLC and Citizen Development Business Finance Ltd identified as the customers who use the highest number of photocopiers. Use of K-Means analysis, DOC T4AL and MX235AT-CHIP has identified as the fast-moving items. Furthermore, found out apart from DOC T4AL and MX235AT-CHIP other toners and chips categorized as slow moving items while developers and drums fell to category of dead stock because movements in drums and developers were minimal compared to toner and chips. These findings are the main objectives of this research. There are sub objectives such as forecasting and strategy suggestions. These two has conducted based on the main findings. This also has descriptively discussed under chapter 4.

The main challenge encountered in this study was the lack of domain knowledge in-order to carry out the research. Author has overcome this challenge by referring research papers, books and other material which helped to conduct this type of research.

Another main challenge of this research was to understanding of the dataset. This task was challenging because this was the basement of the research. Without having a better understanding of the data, it is a difficult task to conduct the research. Hence that author spent lot of time with analyzing data using different methods and tools such as use of MS Excel, WEKA tool, IBM Cognus Insight and IBM SPSS.

The next challenging part was get the "right" dataset removing the garbage data in the data set. 3.3.1 DATA PRE-PROCESSING chapter discussed about how the "right" data set took to conduct the research.

It was difficult to process large number of records for all three years. Hence that analysis conducted to each year separately.

5.2 Future Work

For this research, only five years of data has been considered to conduct the research. By having more data would be able to find out more patterns than in this research. This would help to make better decisions than this research.

By applying demand forecasting algorithms could help to do a better demand forecast for consumable items. Since this is not the main objective of the research there is no high weightage has put on demand forecasting area.

K-Means data mining technique has been used to identify fast moving items in the dataset. By using hierarchical clustering and neural network will allow to explore many more findings and improve the research.

Summary

In concluding the finding of this research, identification of patterns of usage volumes and from that identify high usage photocopier users is one of the major discovery of this research. Identification of fast moving items use of K-Means analysis is the key discovery of this research and it has been done using the domain knowledge which gathered throughout the research.

These findings can be used in better way use of demand forecasting algorithms which helps to increase the profits in the division.

6.0   References

[1] ABC Company. 2015, Annual report 2015, ABC Company

[2] M. J. Zaki and M. Wagner, "*Data and analysis*" in "*Data Mining and Analysis – Fundamental Concepts and Algorithms*", NY: USA, Cambridge University Press, 2014, pp.1.

[3] M. Steinbach, L. Ertöz & V. Kumar, "*The Challenges of Clustering High Dimensional Data*", 2000. [Online]. Available:
http://www-users.cs.umn.edu/~kumar/papers/high_dim_clustering_19.pdf
[Accessed:10th September 2016].

 [4] A. Joshi, N.Pandey, R. Chawla et. Al, "*Use of Data Mining Techniques to Improve the Effectiveness of Sales and Marketing*", Vol. 4, pg.81 – 87, April 2015.

 [5] J. Han & M. Kambler, "*Data Mining Concepts and Techniques*", Morgan Kaufmann Publishers: CA, 2006, Pg. 5.

[6] A. Khan, B. Baharudin and K. Khan, "*Mining Customer Data For Decision Making Using New Hybrid Classification Algorithm*", Vol. 27, Pg. 56-57, May 2011.

 [7] V. Ware & H. N Bharathi, "*Decision Support System for Inventory Management using Data Mining Techniques*", Vol.3,Pg. 164-165,August 2014.

[8] M.Rajeswari & Y.Ramu, "*Frequent Patterns Mining Of Stock Data Using Hybrid Clustering Association*", Vol.1, Pg.33, December 2011.

[9] G. K. Guptha & J. Ghosh, "*Detecting Seasonal Trends and Cluster Motion Visualization for Very High Dimensional Transactional Data*". [2001]. Online. Available: http://www.siam.org/meetings/sdm01/pdf/sdm01_08.pdf [Accessed: 10th September 2016].

[10] A. Aponso, K. Karunarathne, N. Madubashini et Al "*Analysis and Prediction Framework: Case Study in Fast Moving Consumer Goods*", Vol.9, Pg. 68-73, December 2015.

[11] P. Chapman, J Clinton, R Keber et Al "*CRISP – DM- Step by Step Data Mining Guide*", [2000]. Online. Available: https://www.the-modeling-agency.com/crisp-dm.pdf [Accessed: 3rd December 2016].

[12] J. Anderson & A. Kotsiopulos "*Enhanced Decision Making using Data Mining: Applications for Retailers*", Vol.2, Pg. 1-13, Summer 2002.

[13] K. Walinjkar, F. Shah & S. Maskeen "*Business Support System using Hybrid Classification Algorithm*", Vol. 5, Pg. 469-471, February 2015.

[14] S. Rajagopal, "*Customer Data Clustering Using Data Mining Technique*", Vol. 3. Pg. 1-11, November 2011.

[15] Photocopier Prices. 'Types of Photocopier and Their Use', 2013. [Online]. Available: http://www.photocopierprices.com/types-of-photocopiers-and-their-uses/   [Accessed 16th December 2016].

[16] Lanka Sale. 'Toner'. [Online]. Available: https://www.lankasale.com/item-229-printer-cartridge-and-toner-re-filling-.html [Accessed 16th December 2016].

[17] Olx. 'Developer'. [Online]. Available: https://www.olx.com.ng/ad/sharp-photocopier-developer-ID15Ji1P.html [Accessed 16th December 2016].

[18] Ali baba. [Online]. Available: https://www.alibaba.com/product-detail/Compatible-Sharp-AR-207-AR-M160_1899954266.html [Accessed 16th December 2016].

[19] Alphatec [Online]. Available: http://www.alfatec.co.uk/copier-printer-fax-supplies/copiers-spares/ricoh-a2309510-copier-drum-340-350-355-450-455-genuine.html [Accessed 16th December 2016].

[20] Oracle [Online]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/img/dm_process.gif [Accessed 10th January 2017].

[21] F Chen, P Deng, J Wan et. Al, "*Data Mining for the Internet of Things: Literature Review and Challenges*", Vol. 1, Pg.2, March 2015.

[22] ICAR [Online]. Available: http://iasri.res.in/ebook/win_school_aa/notes/Data_Preprocessing.pdf [Accessed 15th January 2017].

[23] E. Rahm H. Do [Online] Available: http://betterevaluation.org/sites/default/files/data_cleaning.pdf [Accessed 15th January 2017].

[24] D Prathima & N. Nimmakanti "*Pattern Recognition Algorithms for Cluster Identification Problem*", Vol. 2. Pg. 25-32, 2008.