PROJECT REPORT

ON

Statistical Analysis and Financial Survey on E-Transactions

SUBMITTED BY

Lavanya.M – (SW11)
Niroopan.E – (SW12)
Sangeeth S Menon – (SW15)
Sagar T – (SW16)
Veera Vasanth Kumar – (SW17)
Nazparween.P.H – (SW20)

Under the Guidance of

RESHMA CB **PRASOON KUMAR KG** **VIMALA MATHEW**

Submitted in partial fulfilment of the requirements for qualifying

ADVANCED DIPLOMA IN BIG DATA ANALYTICS

National Institute of Electronics and Information Technology

(An Autonomous Scientific Society of

Department of Electronics and Information Technology (Deity),

Ministry of Communications and Information Technology, Government of India)

NIT CAMPUS

Calicut 673601

ACKNOWLEDGEMENT

It's a great responsibility to express our sincere thanks to all the people, who have contributed to the successful completion of our project work through their support, encouragement and guidance.

Our first and foremost thanks go to **Dr. M P PILLAI**, Executive Director (Scientist/Engineer 'G'), NIELIT, Calicut, for all the inspiration and encouragement and gave the opportunity to associated with NIELIT by studying in the Advanced Diploma in Big-Data Program.

We owe our gratitude and thanks to **Mrs. SARITHA S**, Scientist/Engineer 'E', NIELIT, Calicut, for her constant co-ordination with the staff during our course, also for various developmental, cultural, social programs that were organized and also for the constant help she provided all through our work.

It is our privilege to place on record our indebtedness to **Mr. PRASOON KUMAR K G**, Senior Technical Officer, NIELIT, Calicut, for his immense support at every stage of the project and course work.

We owe our gratitude and thanks to, **Mrs. VIMALA MATHEW**, Scientist/Engineer 'D', NIELIT, Calicut, for her constant guidance and help all through our work.

We owe our gratitude and thanks to, **Mrs. RESHMA C B**, Senior Technical Officer, NIELIT, Calicut, for her constant guidance and help all through our work.

We also owe our gratitude and thanks to, **Mr. Hari K**, Technical Officer, NIELIT, Calicut, for his help in resolve software installation work.

We would like to express our thanks to all teaching and non-teaching staffs of our institute and our friends for their help in successful completion of this project work.

Abstract

Internet banking, also known as online banking, e-banking or virtual banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website. NEFT, RTGS, IMPS, ECS being the main online transaction mechanism used by Banks in India to conduct one to one transactions.

NEFT (National Electronic Funds Transfer) and RTGS (Real Time Gross Settlement) are the two main fund settlement mechanisms used by banks in India. These transfer protocols are maintained by the Reserve Bank of India.

IMPS (Interbank Mobile Payment Service/Immediate Payment Service) on the other hand are a mobile based payment mechanism introduced in 2010 by the National Payments Corporation of India to allow customers to transfer money instantly, facilitating instant remittance across multiple platforms.

ECS is an alternative method for effecting payment transactions in respect of the utility-bill-payments such as telephone bills, electricity bills, insurance premium, card payments and loan repayments, etc.

Since November 8, 2016 (demonetisation), every Indian has only one thing on mind: the dilemma of choosing a safe, secure, convenient and cashless payment option. Currently available cashless payment systems include IMPS, RTGS, NEFT credit/debit cards, e-wallets such as Paytm.

- The comparative analysis and performance of banks on each cashless transaction modes.
- The analytics study on transactions made before and after demonetisation.
- The final report provides a framework for smooth and speedy transition towards a digital payments economy.

Index

1.Introduction

Government of India has launched the "Digital India" program with a view to transform the country into a digitally empowered economy. Its main purpose is to connect the entire country through broadband within four years. The plan will offer various facilities like digital locker, e-education and e-health. This will help in rapid economic development of India. Digital India program offers a path for development; it has the potential to create a transformation in economy by providing universal phone connectivity. Mobile telephony is expected to play the lead role in delivering the benefits of information access and digital empowerment to the people of India. Information technology has become a necessary tool in today's world of competition. Indian banking industry has witnessed tremendous development due to the changes that are taking place in the information technology. The most important pillar of digital revolution is that it can provide parallel banking system with a physical network of banks. The modern banking brought internet revolution in Indian public and private sector banks. The RBI brought a transformational change by computerizing the bank branches, introducing MICR-based cheque clearing, modernization of payment service and settlements through Electronic Clearing Service (ECS), ATM, RTGS, NEFT were all significant landmarks in the banking technology revolution system. The current and recent trends in banking sector are moving towards E-Banking or Virtual-Banking. People are using RTGS, NEFT facilities to transfer their funds through online-banking.

Today, Banks operate in a highly competitive environment and it's difficult to survive with traditional form of technology due to the changes that are taking place in economy but with the advancement of technology we are able to reap the benefits of Big Data.

Trends in e-Banking and the Banking Industry:

Internet banking is referred as online banking or virtual banking. It is an electronic payment system that enables people to conduct financial transactions online operated by an institution. The term e-banking includes ATMs, Telebanking, mobile banking, ECS, Real Gross Settlement System (RTGS), National Electronic Fund Transfer (NEFT), credit cards, debit cards, smart carts and Telebanking. All these products and services require computerization of banking services.

Growth of Core Banking and Role of Big Data:

Indian banking industry has accepted the advancement in information technology to operate in highly globalized, liberalized and competitive environment. Core banking is a newly developed concept adopted by banks. It is a centralised system, a transaction processing engine and a system for the financial management of the banks. Core banking application provides anywhere, anytime nonstop services. In India, both the public and private banks are moving towards core banking approach. After Core banking concept, the data management has become a critical differentiator that separates market leaders from the rest. Banks unanimously agree that managing and analyzing big data is critical to their success. Today with the advancement of technology we are able to reap the benefits of Big Data. Leveraging Big Data Analytics leads to more confident decision making, which means greater operational efficiencies, cost and risk reductions for an organization.

2.Introduction to Big Data

In the present fast developing world, especially in the developments in digital technology, the data has been generating in large volume, higher velocity and these data has been generating in variety and veracity in the developed and developing world in all field of business industries, government and non-government - profit and non-profit - institutions. Basically there are 3 kinds of data:

1. Structured Data
2. Unstructured Data
3. Semi-structured Data

To know the hidden insight patterns or unlock of these dataset's hidden patterns–more complex in nature of data sets – The traditional data processing applications are inadequate because of the constraints of time and resources. Therefore, big data tools that drive the new insight from either previously untouched data or retrieved data earlier or streaming data and integrating that insight into our business operations – big data tools will do more analysis on the data and could do manipulation, visualization and present to every one for decision making.

The big data is dominated in two kind of the technology such as:

1. Operational Big Data – System that provide the operational capability in real time (using NoSQL technology).

2. Analytical Big Data – Interactive work load where data is primarily captured and stored (Using MapReduce and Massively Parallel Processing (MPP) databases technology).

The both classes of the technology have been combined by using Hadoop. The Hadoop has merged as new technology to address Big Data challenges and enable new type of products and service delivered to the business world.

2.1 Tools Used:

We are indent to do the project work on the on Big Data for Advanced Diploma in Big Data Analytics. Since we are allowed to select the Big Data tools for mining and analysis data, based on the tools that are using in the course of the program for data analysis and predict the outcome of the analysis. We have selected the following big data tools:

1. R - Programming for Analytical, Data Manipulation and Visualization.
2. Python Tkinter for Data visualization and user interface.
3. Java MapReduce, Apache Pig, Flume, Hive - for extraction of data and store.

3. Business objective:

The information technology is developing rapped speed especially in the developed and developing world. To cope with is speedy development in the information technology lot of data integration occurs every industry – Integrating devises with internet. The Big Data analytic helps to know the insight of data pattern. It is one of the job and business opportunities in the analytical field. The government of India's make in India and Digitalization program is more focused on the development of India these programs and development of analytic industries needs more big data analyser and business opportunities.

4. Research Methodology:

Research Design: Exploratory and Empirical research Design based on Secondary Data.

Sources of Data: Annual reports of RBI, Reports on Trend and Progress of Banking in India, Reports of Internet and Reports from etaal.com

Tool of Analysis: The data collected for the study was analysed logically and meaningfully to arrive at meaningful conclusions.

Period of Study: Data for a period of 3 years, (2 Financial Years) 2014-2015, 2015-2016.

5. Software Programs:

Hive:

Apache Hive is a component of Hortonworks **Data** Platform(HDP). Hive provides a SQL-like interface to **data** stored in HDP. In the previous tutorial, we used Pig, which is a scripting language with a focus on dataflows. Hive provides a database query interface to Apache Hadoop.

CREATE TABLE IF NOT EXISTS bank (sno int,month varchar(10),Bankname varchar(50),transaction int, value int)row format delimited fields terminated by ',' lines terminated by '\n';

LOAD DATA LOCAL INPATH '/home/sw15/Documents/project/datasets/Mobile.csv ' OVERWRITE INTO TABLE bank;

INSERT OVERWRITE LOCAL DIRECTORY '/home/sw15/Documents/sangeeth' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' select bankname,avg(transaction) from bank group by bankname;

Pig:

Hive is best for structured Data & **PIG** is best for semi structured data. Hive is **used** for reporting & **PIG** for programming

```
mob = LOAD '/usr/local/hadoop/group1test/Mobile.csv'
 USING PigStorage(',') as (Sr_no:int, Month:chararray, Bank_name:chararray, Transaction:int, Amount:float);
filter_data = FILTER mob BY Bank_name == 'STATE BANK OF INDIA';
STORE filter_data into '/usr/local/hadoop/group1test/mobileanswer/answer' USING PigStorage (',');
```

R-Code:

R is a programming language and software environment for statistical computing and graphics supported by the **R** Foundation for Statistical Computing. The **R**language is widely **used** among statisticians and data miners for developing statistical software and data analysis.

```
Sys.setenv(HADOOP_PREFIX="/usr/local/hadoop")
Sys.setenv(HADOOP_CMD="/usr/local/hadoop/bin/hadoop")
Sys.setenv(HADOOP_STREAMING="/usr/local/hadoop/share/hadoop/tools/lib
/hadoop-streaming-2.7.3.jar")
library(rhdfs)
library(rmr2)

hdfs.init()
hdfs.ls('/')
#ECS

A=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/ECS1415/000000_0')
write.csv(A, "ecs1415.csv", row.names=FALSE)
ecs1=read.csv("/home/sw15/ecs1415.csv",header=FALSE)

B=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/ECS1516/000000_0')
write.csv(B, "ecs1516.csv", row.names=FALSE)
ecs2=read.csv("/home/sw15/ecs1516.csv",header=FALSE)

X1=ecs1$V1
Y1=ecs1$V2
X2=ecs2$V1
Y2=ecs2$V2

P= rbind(Y1,Y2)
P=P/1000000
Q <-
c("Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec","Jan","Feb","Mar")
AA =barplot(P,names.arg =Q,las=2,ylim =c(0,25))
text(x = AA, y = P, label = P, pos = 3, cex = 0.8, col = "red")
```

#MOBILE

```
C=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/Mobileyr1415/000000_0')
write.csv(C, "mobile1415.csv", row.names=FALSE)
mob1=read.csv("/home/sw15/mobile1415.csv",header=FALSE)

D=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/Mobileyr1516/000000_0')
write.csv(D, "mobile1516.csv", row.names=FALSE)
mob2=read.csv("/home/sw15/mobile1516.csv",header=FALSE)

X3=mob1$V1
Y3=mob1$V2

X4=mob2$V1
Y4=mob2$V2
```

#NEFT

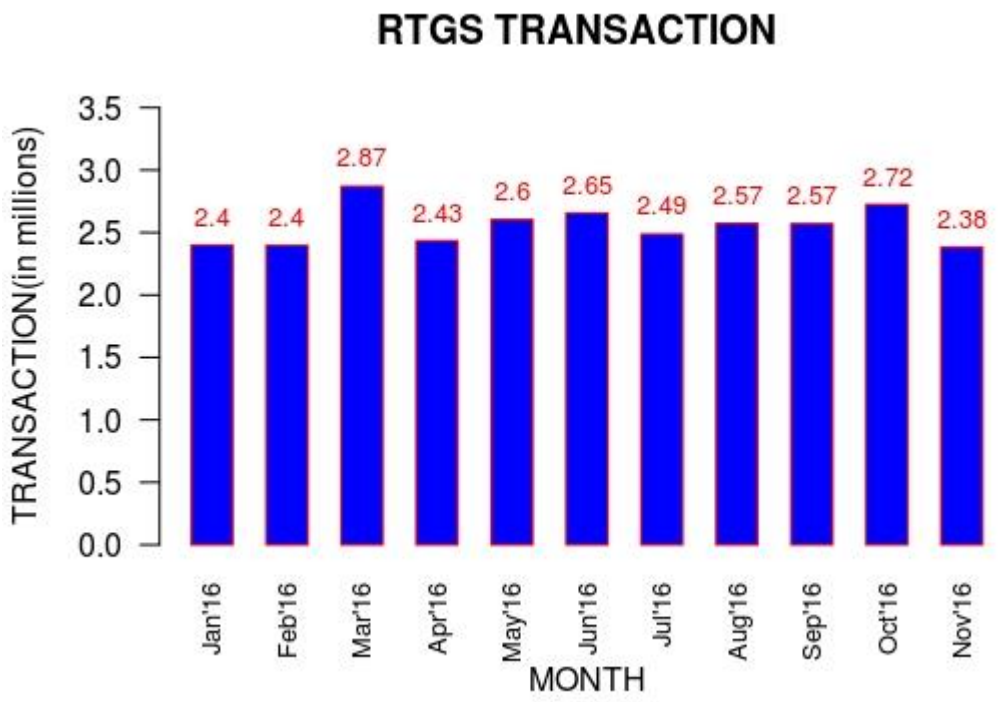
```
E=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/NEFTyr1415/000000_0')
write.csv(E, "neft1415.csv", row.names=FALSE)
neft1=read.csv("/home/sw15/neft1415.csv",header=FALSE)


neft2=read.csv("/home/sw15/neft1516.csv",header=FALSE)

X5=neft1$V1
Y5=neft1$V2

X6=neft2$V1
Y6=neft2$V2
```

#RTGS

```
H=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/RTGSyr1415/000000_0')
write.csv(H, "rtgs1415.csv", row.names=FALSE)
```

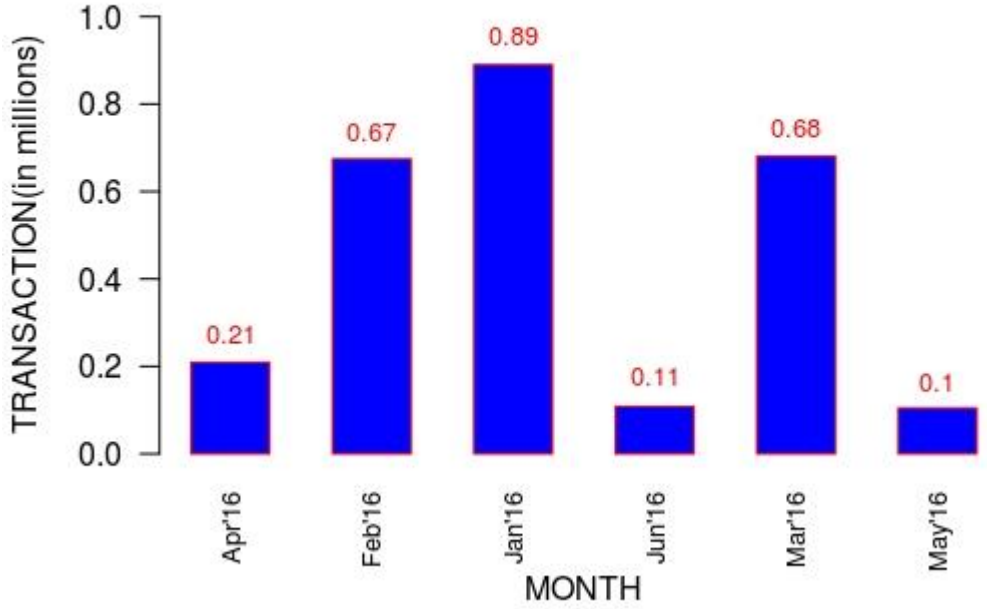
```
rtgs1=read.csv("/home/sw15/rtgs1415.csv",header=FALSE)

I=hdfs.read.text.file('/usr/local/hadoop/sw20/projectyear/RTGSyr1516/000000_
0')
write.csv(I, "rtgs1516.csv", row.names=FALSE)
rtgs2=read.csv("/home/sw15/rtgs1516.csv",header=FALSE)

X7=rtgs1$V1
Y7=rtgs1$V2

X8=rtgs2$V1
Y8=rtgs2$V2
```

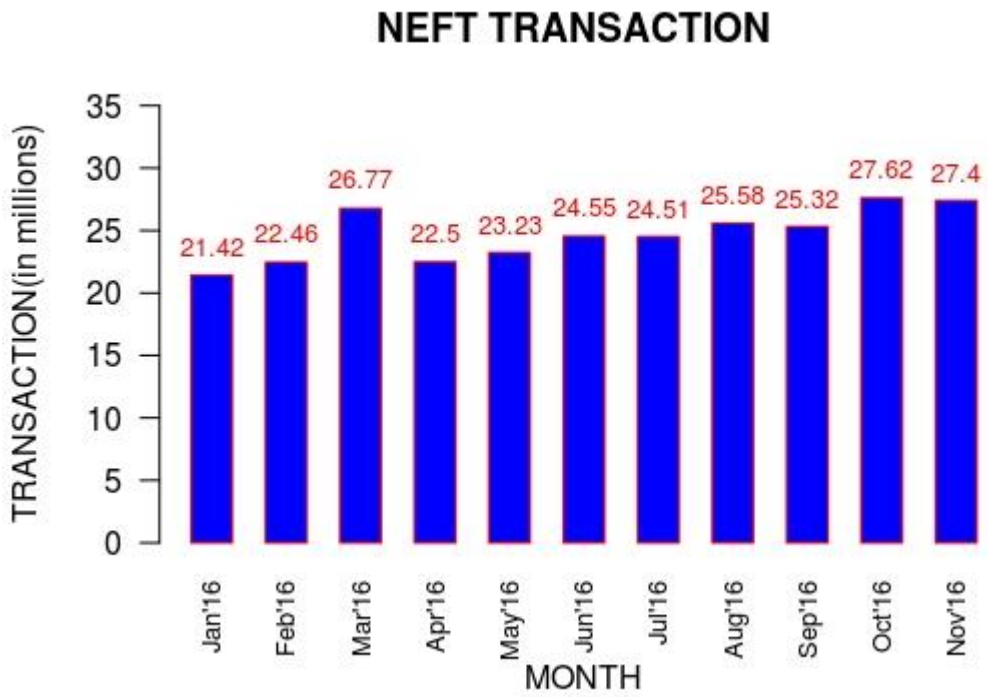
Month-wise Transaction:



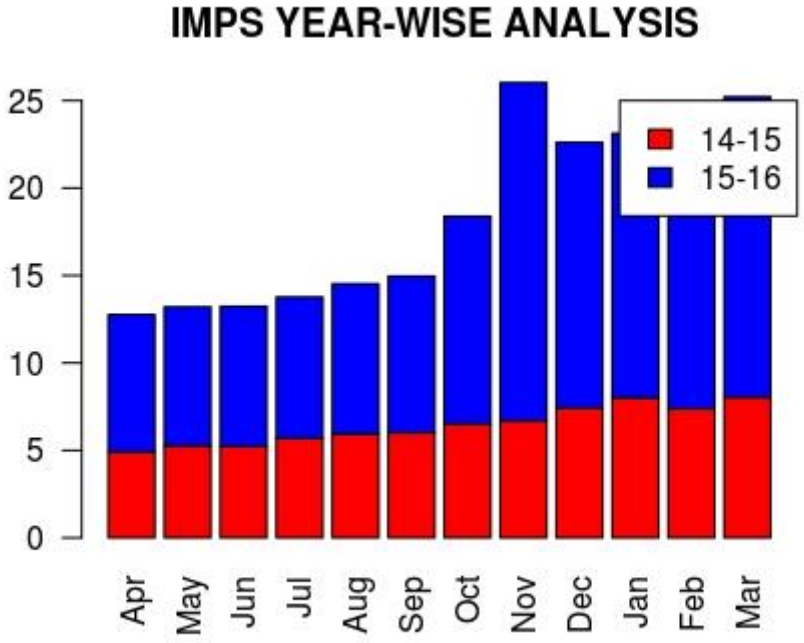
MOBILE TRANSACTION

ECS TRANSACTION

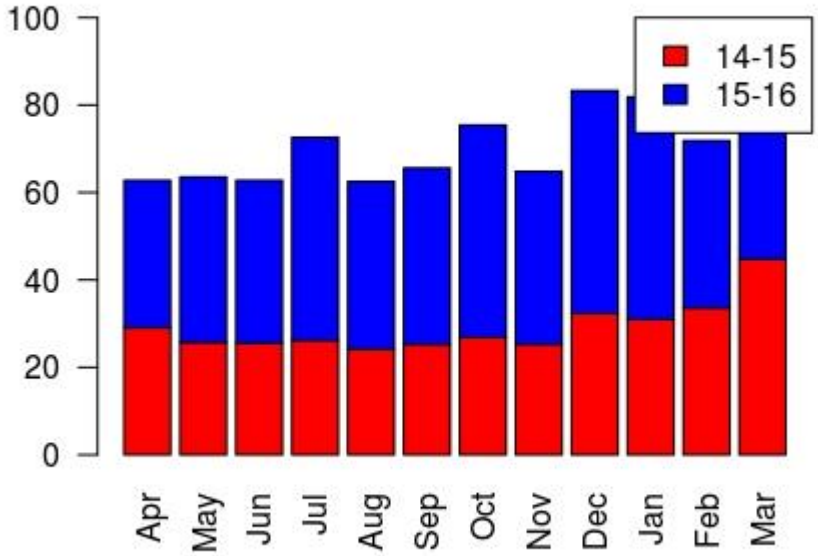




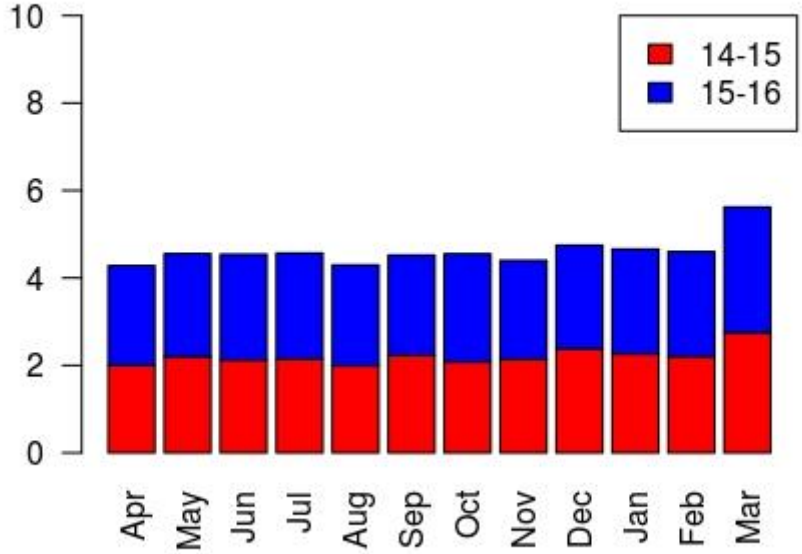
Year-wise Comparison:

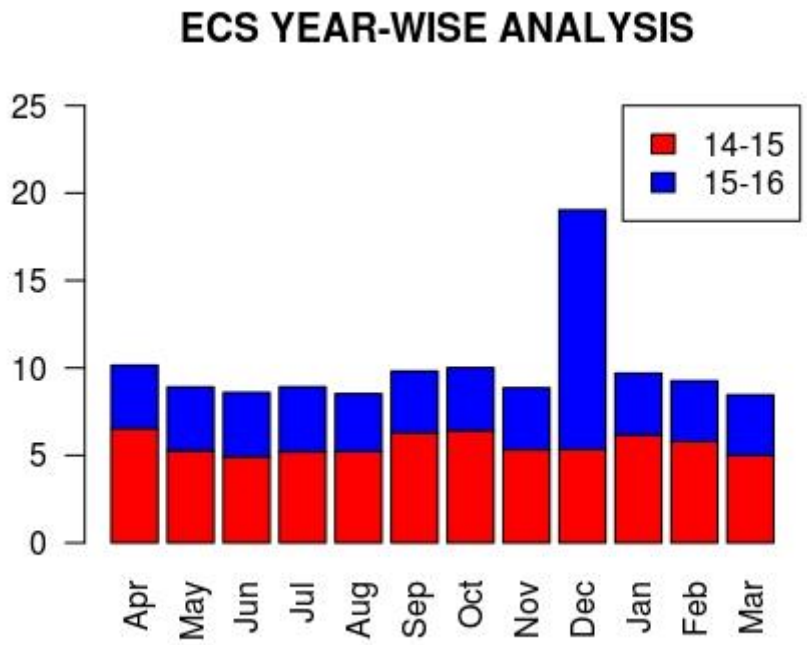


NEFT YEAR-WISE ANALYSIS

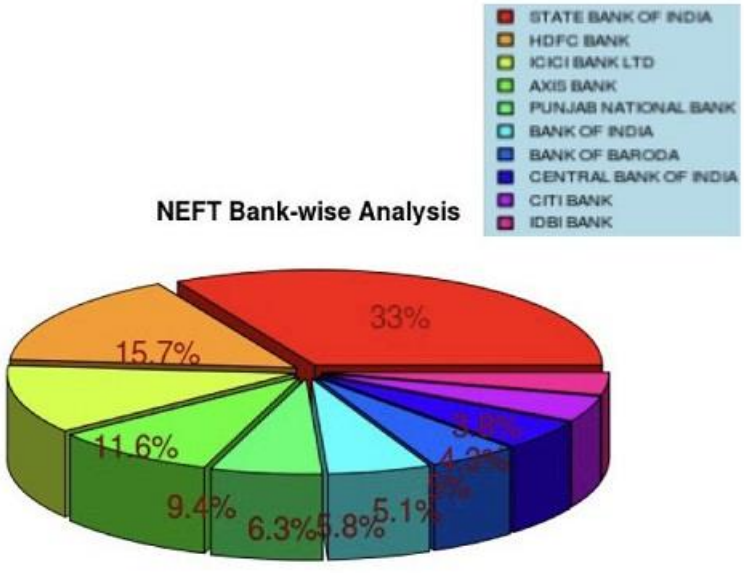


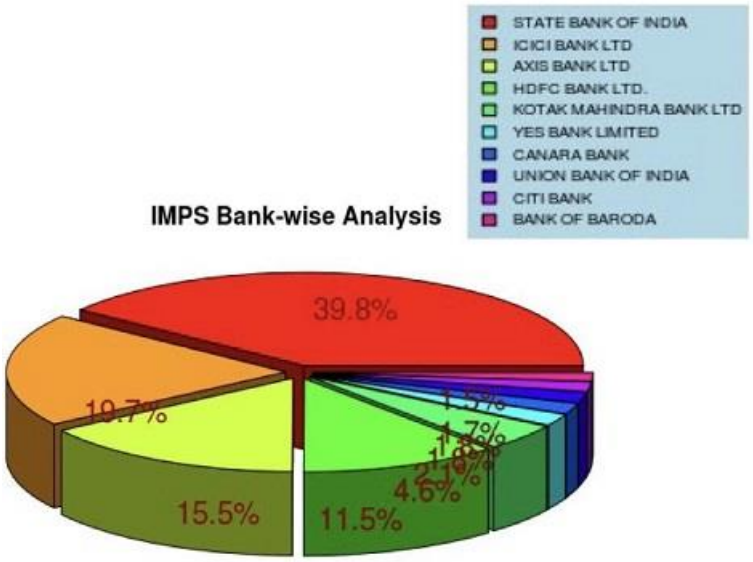
RTGS YEAR-WISE ANALYSIS

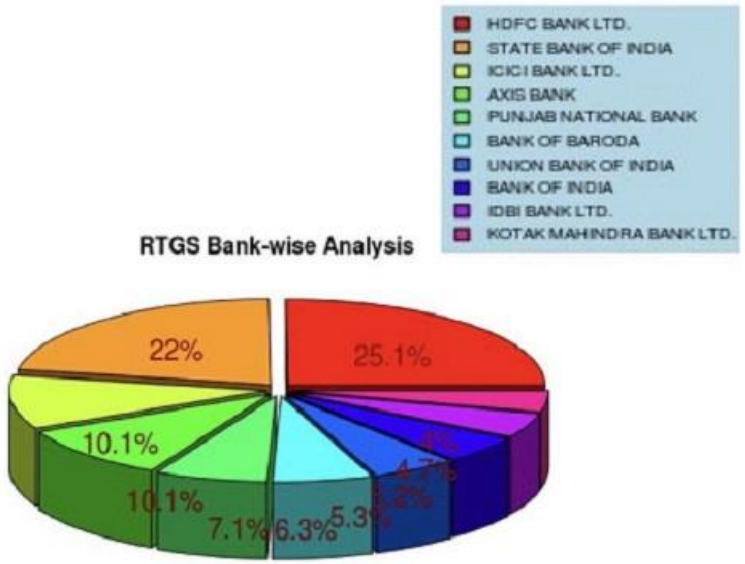




3D – Bar Plot – Top 10 Banks:




IMPS Bank-wise Analysis
STATE BANK OF INDIA
ICICI BANK LTD
AXIS BANK LTD
HDFC BANK LTD.
KOTAK MAHINDRA BANK LTD
YES BANK LIMITED
CANARA BANK
UNION BANK OF INDIA
CITI BANK
BANK OF BARODA
39.8%
15.5%
11.5%
4.6%


RTGS Bank-wise Analysis
HDFC BANK LTD.
STATE BANK OF INDIA
ICICI BANK LTD.
AXIS BANK
PUNJAB NATIONAL BANK
BANK OF BARODA
UNION BANK OF INDIA
BANK OF INDIA
IDBI BANK LTD.
KOTAK MAHINDRA BANK LTD.
25.1%
22%
10.1%
10.1%
7.1%
6.3%

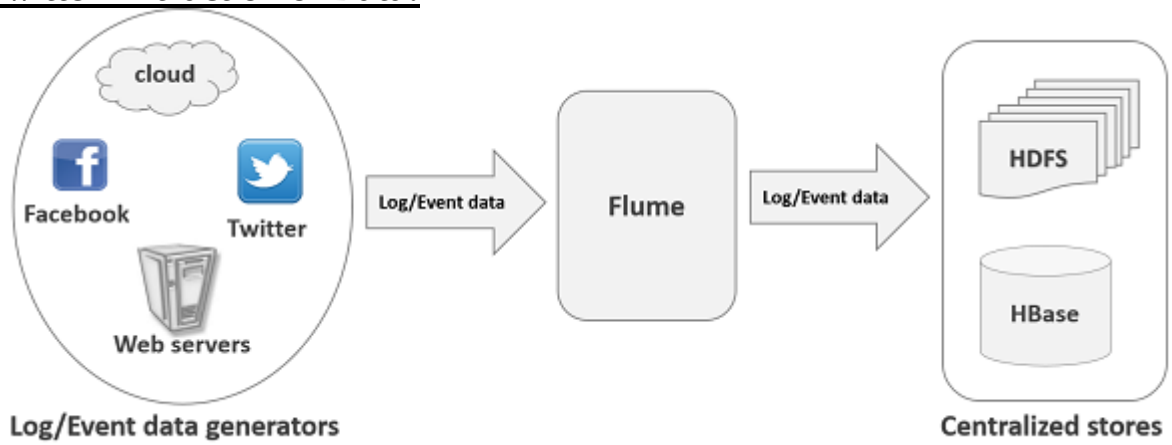
Flume – MapReduce:

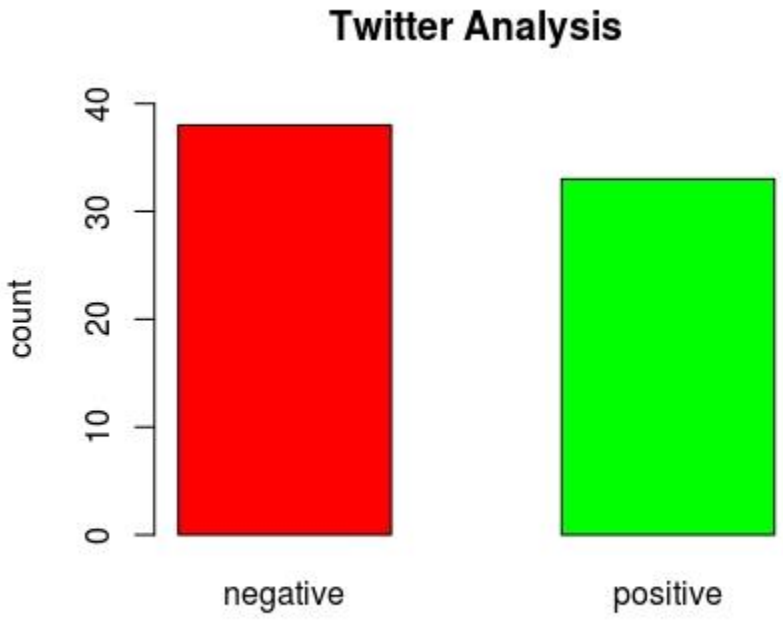
Apache **Flume** is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the**Hadoop** Distributed File System (HDFS).

```
import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class SentiAnalysis {

	/**
	 * @param args
	 */
	@SuppressWarnings("deprecation")
	public static void main(String[] args) {
		// TODO Auto-generated method stub
			try {
			Job job = new Job();
			job.setJarByClass(SentiAnalysis.class);
			job.setJobName("Analysis");

			Path path1 = new Path(args[0]);
			FileInputFormat.addInputPath(job, path1);
			Path path2 = new Path(args[1]);
			FileOutputFormat.setOutputPath(job, path2);

			job.setMapperClass(SentiMapper.class);
			job.setReducerClass(SentiReducer.class);

			job.setOutputKeyClass(Text.class);
			job.setOutputValueClass(IntWritable.class);

			System.exit(job.waitForCompletion(true)?0:1);

		} catch (IOException e) {
			// TODO Auto-generated catch block
			e.printStackTrace();
		} catch (ClassNotFoundException e) {
```

```
                    // TODO Auto-generated catch block
                    e.printStackTrace();
                } catch (InterruptedException e) {
                    // TODO Auto-generated catch block
                    e.printStackTrace();
//MAPPER
public class SentiMapper   extends Mapper<LongWritable, Text, Text, IntWritable>
{

  public void map(LongWritable key, Text value, Context context)
     throws IOException, InterruptedException {

        // TODO Auto-generated method stub
        String data = value.toString();
        int textLocStart = data.indexOf("\"text\":\"") + ("\"text\":\"").length();
        int textLocEnd = data.indexOf("\"contributors\":") -2;
                    //data.indexOf(",contributors");
        String textContent = data.substring(textLocStart, textLocEnd);
        context.write(new Text(textContent), new IntWritable(1));


//REDUCER
public class SentiReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
         @Override
         public void reduce(Text key, Iterable<IntWritable> values,
            Context context)
            throws IOException, InterruptedException {
                int cnt = 0;
                Iterator<IntWritable> iterator = values.iterator();

                while(iterator.hasNext())
                {
                        cnt = cnt +   iterator.next().get();
                        }
                IntWritable cntIntW = new IntWritable(cnt);
                context.write(key, cntIntW);
```

Twitter Extraction of Data:





Python – Tkinter:

Tkinter provides various controls, such as buttons, labels and text boxes **used** in a GUI application. These controls are commonly called widgets. The Button widget is **used** to display buttons in your application. The Canvas widget is **used** to draw shapes, such as lines, ovals, polygons and rectangles, in your application.

```
import Tkinter as tk
import ImageTk
from Tkinter import *
def show_image1():
    x = canvas.create_image(500,300, image=tk_img1)
    while True:

        canvas.itemconfigure(x, state=tk.NORMAL)
        button1.configure(text = 'FIND')
        yield

        canvas.itemconfigure(x, state=tk.HIDDEN)
        button1.configure(text = 'FIND')
        yield

root = tk.Tk()
canvas = tk.Canvas(root, width=1000, height=600)
canvas.grid(row=5, column=8)

label1 = Label( root, text=" TWITTER SENTIMENT ANALYSIS on PAYTM")
label1.grid(row =0,column = 0,sticky=tk.NW)

tk_img1 = ImageTk.PhotoImage(file='/home/sw15/Documents/project/twitter plot.jpeg')

button1 = tk.Button(
    root, text="FIND", command=show_image1().next, anchor='n',
    width=10, activebackground="#33B5E5")

button1.grid(row=3, column=0,sticky=tk.NW,)
```

```
root.mainloop()
```

Monitoring customers feedback through Twitter Sentiment Analysis:

Sentiment analytics helps to analyses opinions about the products and services available on social media, blogs and review sites in text format. Extracted dataset with help of Apache Flume and used PIG analytics tool to monitor customers' feedback. It analyses original comments and breaks them in groups as negative and positive sentiments. Later, scores are assigned accordingly in the ranges from highly negative (-1) to neutral (0) to highly positive (1). This scoring helps the bank to view the overall sentiment without having to read all the comments posted about their products or brand. Banks generate humongous data such as purchase history including online transactions, customer profile data, browsing history, social media data on every single day. Using this information, banks can employ data analytics tools to reach effectively their main marketing objectives.

Demonetization Effect – New Payment Platforms Grows:

We have seen mobile payments (IMPS) in India picking up quite well and currently at about 20% of the ECS transactions.

Paytm:

The Indian RBI has bucked the overall Economic Trends and given wide scope for Digital Payment Platform. In this backdrop it is very important to note that Paytm Digital Payment Platform (Wallet using Pre-Paid Cash or Money Added via Debit / Credit Card, has now acquired over 600,000 merchants (compared to 300,000 for SBI, the largest among bank acquirers)) within 9 months of launch. Paytm is already conducting ~ 10mn transactions a month (plans to double it over the next few months) with these merchants, compared to 40mn a month at SBI & HDFC Bank. With much lower onboarding costs relative to POS and Zero MDR Merchant Discount Rate charge, Paytm is on boarding even small merchants on its network. Paytm plans to offer customers onboarding facility through its business correspondent network of more than 10,000 outlets. The payment bank would tie up with multiple service providers to offer most of the facilities being provided by a universal bank. For customers with balance more than Rs100k (not allowed with a payment bank), the amount would be automatically swept to a partner bank.

Paytm targets to rapidly on board its wallet customers to the bank through the e-KYC facility. Reliance JIO recently onboarded 16 million customers in a month using e-KYC. Paytm is likely to launch its payment bank this month. Meanwhile, it already has a wallet customer base of 150mn (similar to the total no. of SBI deposit customers), processes 85m transactions monthly and already merchant transaction volume is exceeding the mobile recharge volumes.

Digitalization is going to be the core guiding principal going forward and all these new platforms will give Banks a serious run for their money and profits if they fail to innovate and retain customers.

Conclusion and Future Scope:

Most of the banks have already implemented the e-banking facilities, as these facilities are beneficial to both i.e. banks as well as consumers. But the adoption of e-banking by the consumers is still at the early stage due to various challenges. The challenges such as security risk, privacy risk, trust factor and less awareness among consumers about e-banking are acting as hurdle in the adoption of e-banking facilities. Considering the challenges and risk related to e-banking, the Government of India along with various government agencies is making an effort to make e-banking more safe, secure and reliable.

- The comparison of NEFT, RTGS, Mobile Banking and ECS for 2 Financial Years (2014-15, 2015-16) shows an increase of 3% to 5% in all the modes of transactions.
- Month-wise comparison (2016) shows the following bank had the maximum transactions –
    - NEFT - HDFC – 24667892
    - RTGS - HDFC – 2553820
    - ECS - RBI – 44447604
    - Mobile Transactions - SBI – 16820286

- Top 10 Bank-wise analysis on e-Transactions shows that the private banks are more supportive and inclined to the same.
- The twitter sentiment analysis on Paytm indicates that such e-Transaction app has better chance of survival supportive to demonetisation hit.

The future studies may be conducted to analyse the various factors which influence the consumer intention to adopt internet banking services.