Business Intelligence Report

(Bank Marketing Data-set)

Made By:
1. Ruturaj Kotwal (509)
2. Mahesh Kendre (514)

Bank Marketing

Case Study

Find the best strategies to improve for the next marketing campaign. How can the financial institution have a greater effectiveness for future marketing campaigns? In order to answer this, we have to analyze the last marketing campaign the bank performed and identify the patterns that will help us find conclusions in order to develop future strategies.

Data Set

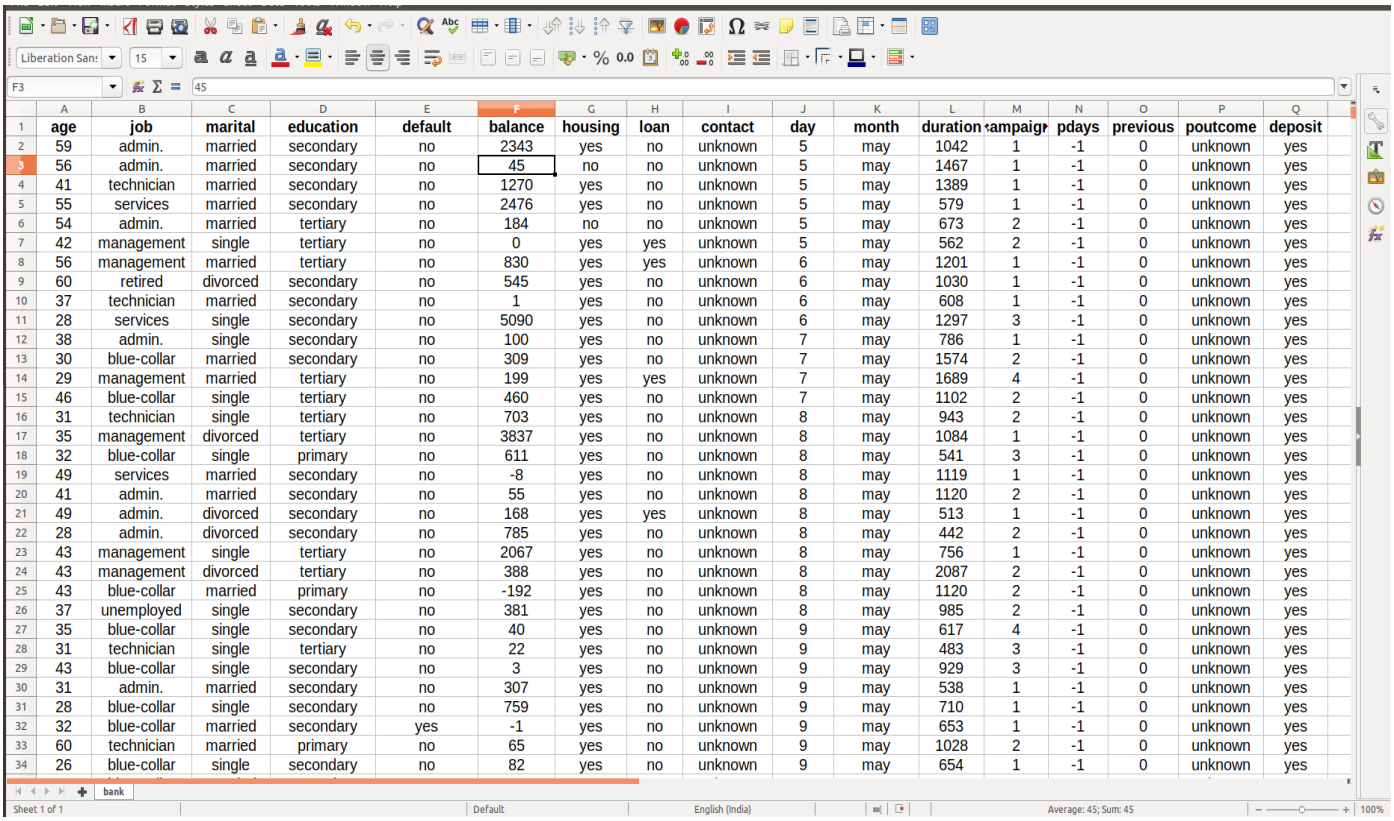
The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

There are four datasets:
1) bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]

2) bank-additional.csv with 10% of the examples (4119), randomly selected from 1), and 20 inputs.

3) bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).

4) bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).
The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
2	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	yes
3	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	yes
4	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	yes
5	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	yes
6	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	yes
7	42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2	-1	0	unknown	yes
8	56	management	married	tertiary	no	830	yes	yes	unknown	6	may	1201	1	-1	0	unknown	yes
9	60	retired	divorced	secondary	no	545	yes	no	unknown	6	may	1030	1	-1	0	unknown	yes
10	37	technician	married	secondary	no	1	yes	no	unknown	6	may	608	1	-1	0	unknown	yes
11	28	services	single	secondary	no	5090	yes	no	unknown	6	may	1297	3	-1	0	unknown	yes
12	38	admin.	single	secondary	no	100	yes	no	unknown	7	may	786	1	-1	0	unknown	yes
13	30	blue-collar	married	secondary	no	309	yes	no	unknown	7	may	1574	2	-1	0	unknown	yes
14	29	management	married	tertiary	no	199	yes	yes	unknown	7	may	1689	4	-1	0	unknown	yes
15	46	blue-collar	single	tertiary	no	460	yes	no	unknown	7	may	1102	2	-1	0	unknown	yes
16	31	technician	single	tertiary	no	703	yes	no	unknown	8	may	943	2	-1	0	unknown	yes
17	35	management	divorced	tertiary	no	3837	yes	no	unknown	8	may	1084	1	-1	0	unknown	yes
18	32	blue-collar	single	primary	no	611	yes	no	unknown	8	may	541	3	-1	0	unknown	yes
19	49	services	married	secondary	no	-8	yes	no	unknown	8	may	1119	1	-1	0	unknown	yes
20	41	admin.	married	secondary	no	55	yes	no	unknown	8	may	1120	2	-1	0	unknown	yes
21	49	admin.	divorced	secondary	no	168	yes	yes	unknown	8	may	513	1	-1	0	unknown	yes
22	28	admin.	divorced	secondary	no	785	yes	no	unknown	8	may	442	2	-1	0	unknown	yes
23	43	management	single	tertiary	no	2067	yes	no	unknown	8	may	756	1	-1	0	unknown	yes
24	43	management	divorced	tertiary	no	388	yes	no	unknown	8	may	2087	2	-1	0	unknown	yes
25	43	blue-collar	married	primary	no	-192	yes	no	unknown	8	may	1120	2	-1	0	unknown	yes
26	37	unemployed	single	secondary	no	381	yes	no	unknown	8	may	985	2	-1	0	unknown	yes
27	35	blue-collar	single	secondary	no	40	yes	no	unknown	9	may	617	4	-1	0	unknown	yes
28	31	technician	single	tertiary	no	22	yes	no	unknown	9	may	483	3	-1	0	unknown	yes
29	43	blue-collar	single	secondary	no	3	yes	no	unknown	9	may	929	3	-1	0	unknown	yes
30	31	admin.	married	secondary	no	307	yes	no	unknown	9	may	538	1	-1	0	unknown	yes
31	28	blue-collar	single	secondary	no	759	yes	no	unknown	9	may	710	1	-1	0	unknown	yes
32	32	blue-collar	married	secondary	yes	-1	yes	no	unknown	9	may	653	1	-1	0	unknown	yes
33	60	technician	married	primary	no	65	yes	no	unknown	9	may	1028	2	-1	0	unknown	yes
34	26	blue-collar	single	secondary	no	82	yes	no	unknown	9	may	654	1	-1	0	unknown	yes

Bank.csv

[Data Folder](), [Data Set Description]()

Data Set Characteristics:	Multivariate	**Number of Instances:**	45211	**Area:**	Business
Attribute Characteristics:	Real	**Number of Attributes:**	17	**Date Donated**	2012-02-14
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	1106135

Attribute Information:

Inputvariables:
bank client data:
1 - age (numeric)

2 - job : type of job (categorical: 'admin.','blue-collar', 'entrepreneur', 'housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')

3 - marital : marital status (categorical: 'divorced', 'married', 'single','unknown'; note: 'divorced' means divorced or widowed)

4 - education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school' ,'illiterate','professional.course','university.degree','unknown')

5 - default: has credit in default? (categorical: 'no','yes','unknown')

6 - housing: has housing loan? (categorical: 'no','yes','unknown')

7 - loan: has personal loan? (categorical: 'no','yes','unknown')
related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: 'cellular', 'telephone')

9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

10 - day_of_week: last contact day of the week (categorical: 'mon', 'tue', 'wed','thu','fri')

11 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

#otherattributes:
12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

14 - previous: number of contacts performed before this campaign and for this client (numeric)

15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

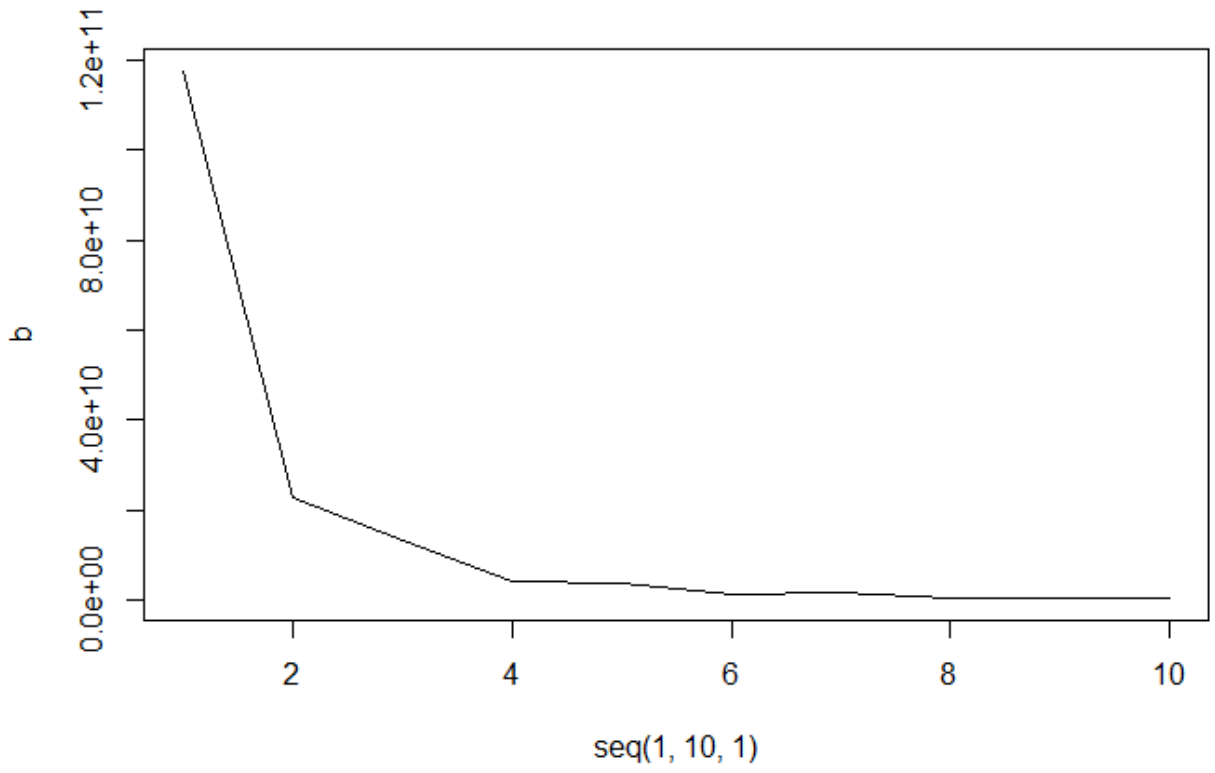
ETL Process

Extrating Data

Data is extract from bank campain data which is store in csv files

Transformation Data

Data is transforme using ***k*-means clustering** To decide number of cluder here we ues Elbow Method(The “Elbow” method to help data scientists select the optimal number of clusters by fitting the model with a range of values for K)



For transformation huge data to optimal here we taken diffrent set of cluster range form 1 to 10 and plot on the graph so we get the

appropriate Number of cluster for the bank data set the optimal value we get as 4

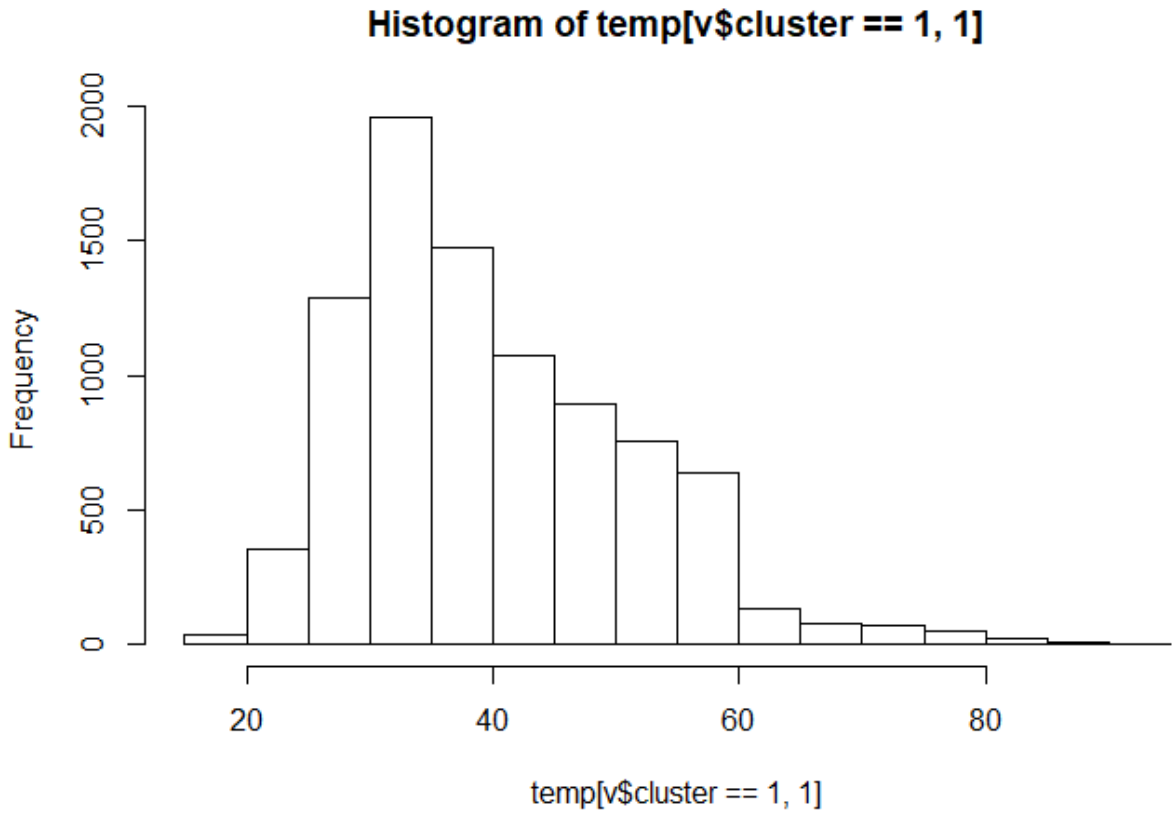
`kmeans` returns an object of class `"kmeans"` which has a `print` and a `fitted` method. It is a list with at least the following components:

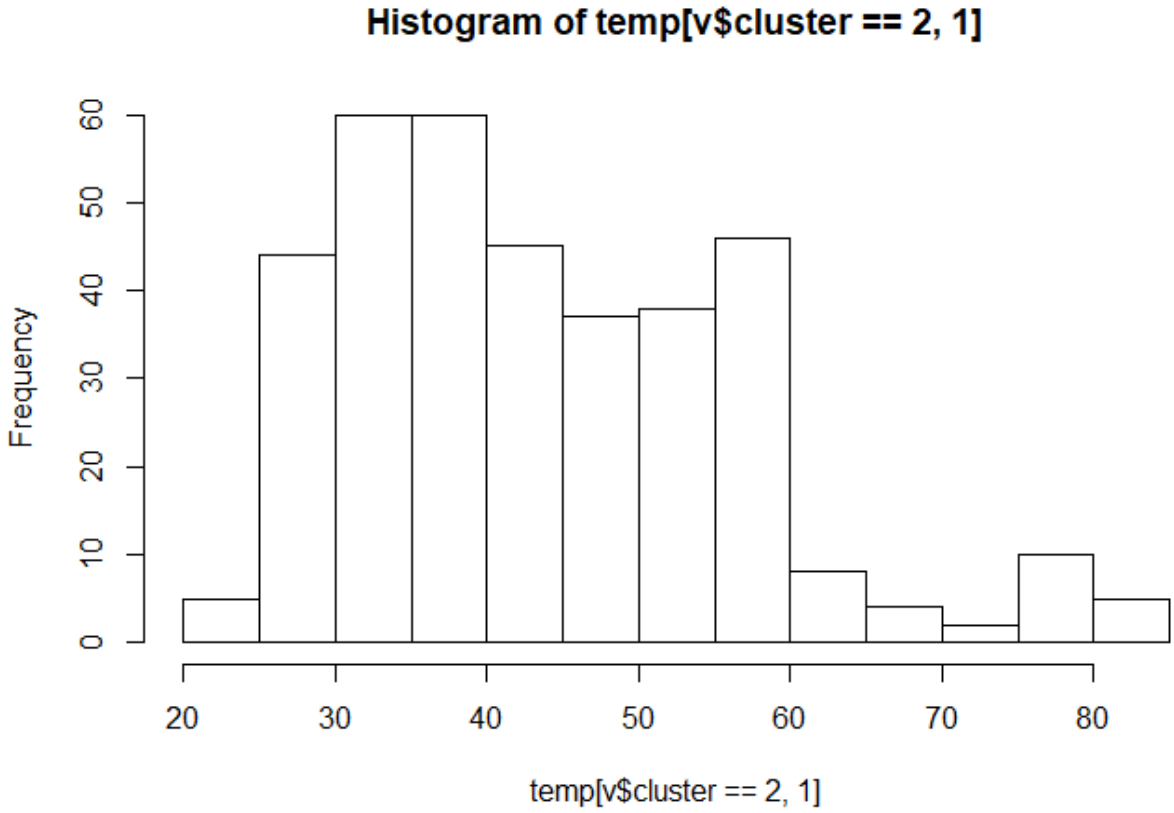
`cluster`	A vector of integers (from `1:k`) indicating the cluster to which each point is allocated.
`centers`	A matrix of cluster centres.
`totss`	The total sum of squares.
`withinss`	Vector of within-cluster sum of squares, one component per cluster.
`tot.withinss`	Total within-cluster sum of squares, i.e. `sum(withinss)`.
`betweenss`	The between-cluster sum of squares, i.e. `totss-tot.withinss`.
`size`	The number of points in each cluster.
`iter`	The number of (outer) iterations.
`ifault`	integer: indicator of a possible algorithm problem – for experts.

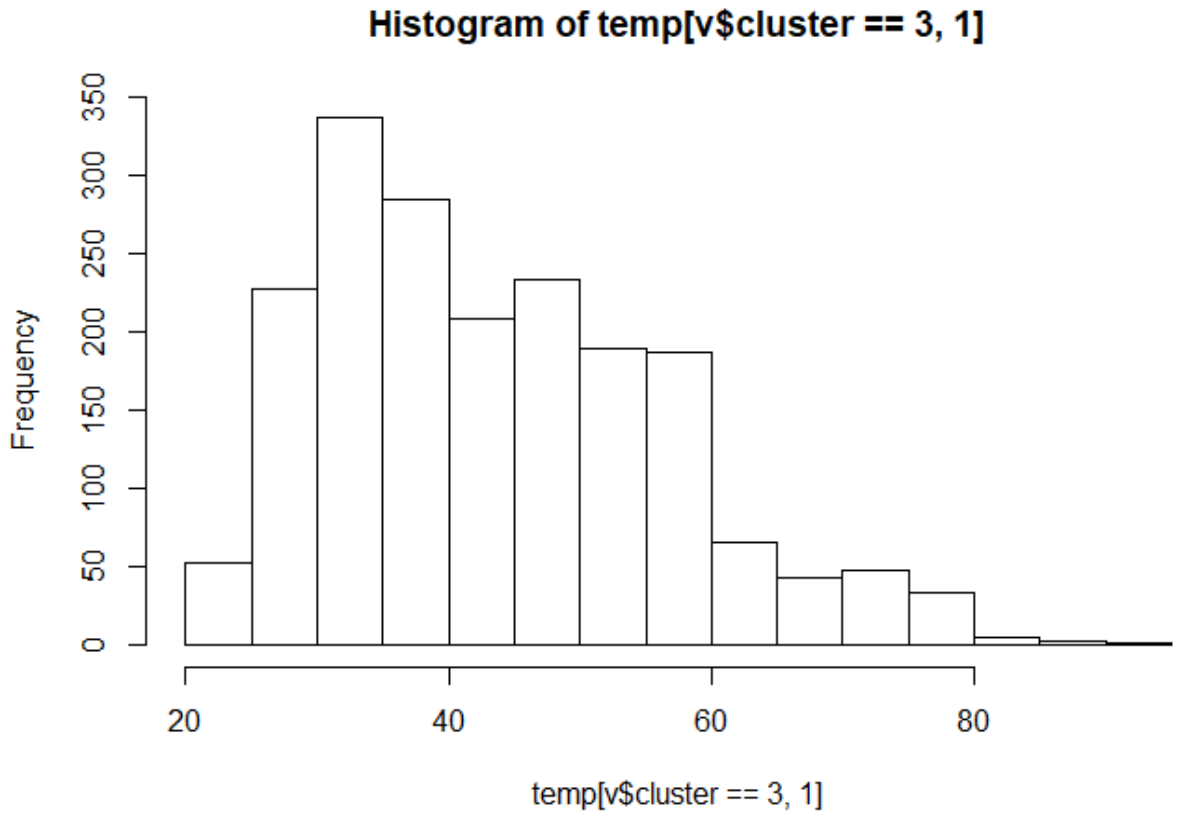
so we use the kmean formula and pass the data set useing that kmean object we put the data on sheet

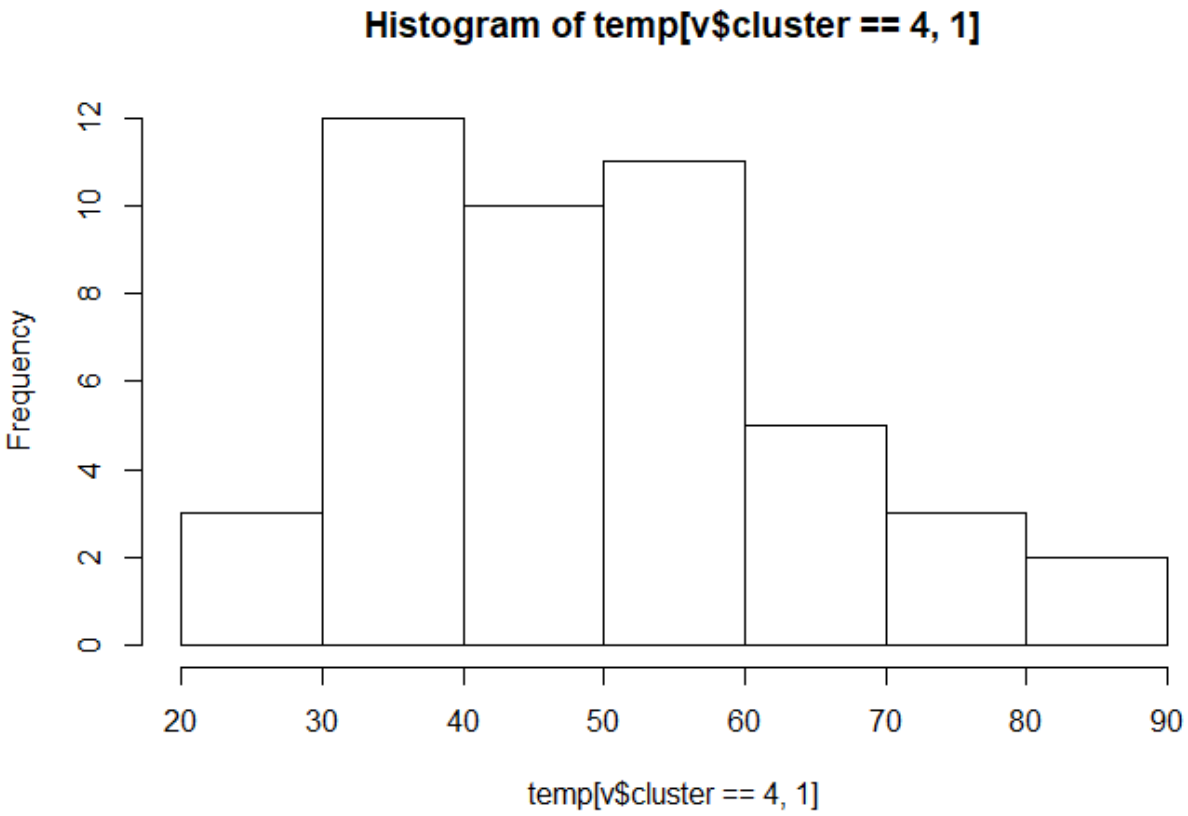
	A	B	C	D	E	F
1		**Description**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
2	**age**	Range of Max Freq	25 to 60	25 to 60	25 to 60	30 to 60
3	**admin**	% of Jobs in Cluster	12.7	8.5	9.3	4.3
4	**blue_collar**	% of Jobs in Cluster	18.5	11.5	13.8	6.5
5	**entrepreneur**	% of Jobs in Cluster	2.9	2.5	3	6.5
6	**housemaid**	% of Jobs in Cluster	2.4	1.9	2.5	2.1
7	**management**	% of Jobs in Cluster	21.7	29.9	27	36.9
8	**retired**	% of Jobs in Cluster	5.8	10.4	11.26	17.4
9	**self_employed**	% of Jobs in Cluster	3.5	4.7	4	4.3
10	**student**	% of Jobs in Cluster	3.2	3.3	3.23	4.34
11	**services**	% of Jobs in Cluster	8.9	4.12	6.1	2.2
12	**technician**	% of Jobs in Cluster	16.2	18.68	16.21	15.21
13	**unemployed**	% of Jobs in Cluster	3.4	3.3	2.55	0
14	**unk_job**	% of Jobs in Cluster	0.56	1.09	0.83	0
15	**single**	% of Marital in Cluster	32.34	31.31	27.94	23.91
16	**married**	% of Marital in Cluster	55.96	59.89	60.37	67.39
17	**divorced**	% of Marital in Cluster	11.69	8.79	11.67	8.69
18	**primary**	% of Education in Cluster	13.24	12.91	14.39	15.22
19	**secondary**	% of Education in Cluster	51.4	36.81	41.1	26.1
20	**tertiary**	% of Education in Cluster	31.16	44.5	38.95	54.33
21	**unk_education**	% of Education in Cluster	4.2	5.77	5.52	4.35
22	**default**	% of Defaulters	1.879	0	0.104	0
23	**balance**	Range of Max Deposit	-500 to 2000	6000 to 14000	2000 to 6000	20000 to 30000
24	**housing**	% of People having homes	49.536	35.44	39.93	21.74
25	**loan**	% of People having loans	14.81	4.12	6.93	6.52
26	**cellular**	% of Contact_mode in Cluster	71.69	72.8	73.52	73.91
27	**telephone**	% of Contact_mode in Cluster	6.1	10.44	9.91	15.22
28	**unk_contact**	% of Contact_mode in Cluster	22.2	16.76	16.57	10.87
29	**day**	Ranges of Max Freq	2-8,10-20,28-30	10-21	4-6,11-21,30	10-15,21,26
30	**month**	Ranges of Max Freq	Feb,May-Aug,Nov	Feb,Apr-Aug,Nov	Feb,Apr-Aug,Nov	Feb,May-Aug,Nov
31	**duration**	Range of Max Freq	0-400	0-400	0-400	0-400
32	**campaign**	Max Freq Range	1-2	1-5	1-4	1-2
33	**pdays**	Max Freq Range	-1,50-100,150-200	-1,50-100,150-200	-1,50-100,150-200	-1,50-100,150-200
34	**previous**	Max Freq Range	0-6	0-3	0-5	0-3
35	**failure**	% of camp_outcome in Cluster	10.73	10.99	12.35	6.52
36	**other**	% of camp_outcome in Cluster	4.61	5.49	5.47	10.87
37	**success**	% of camp_outcome in Cluster	8.88	13.74	11.99	13.04
38	**unk_poutcome**	% of camp_outcome in Cluster	75.77	69.78	70.17	69.56

plot thes parameter on graph with respect to age similarly on each field we plot a graph and analyse the data






Histogram of temp[v$cluster == 3, 1]
Frequency
0
50
100
150
200
250
300
350
20
40
60
80
temp[v$cluster == 3, 1]


Histogram of temp[v$cluster == 4, 1]
Frequency
0
2
4
6
8
10
12
20
30
40
50
60
70
80
90
temp[v$cluster == 4, 1]

Observations

1. Age:
   - Maximum Customers belong to the age group of 25 to 60 years except Cluster 4, where lower limit is 30 years.

2. Jobs:
   - Admin, Blue-Collar, Management, Technician are the most popular jobs among all the Clusters.
   - In Cluster 1, admin consists of 12.7% of total cluster population. Blue-Collar jobs and Technicians constitute similar frequency of 18.5% and 16.2% respectively. Maximum jobs are held by the Management sector. It has a high percentage of population who are Unemployed, 3.4%. It has the maximum no. of customers working in Services, out of all the Clusters(8.9%).
   - Maximum Customers in Cluster 2 are working in the Management sector(29.9%). It has the maximum no. of customers working as Technicians, out of all the Clusters(18.68%). This cluster also has a high unemployment rate of 3.3%.
   - Cluster 3 has one of the highest number of people working in blue-collar jobs(13.8%), of all the clusters. It has high percentage of retired population (11.26%) and 16.21% are Technicians. It also one of the highest percentages of Customers working in Admin sector(9.3%), among the 4 sectors.
   - Cluster 4 has the lowest percentage working in the Admin sector(4.3%). 36.9% Customers are from the Management sector. It has the lowest percentage (2.2%) in the Services. There are no unemployed and unknown-job Customers in this cluster.

3. Marital Status:
   - Highest percentage(32.34%) of singles is in cluster 1. Maximum people in Cluster 1 are married(55.96). Highest percentage(32.34%) of divorcees are also in cluster 1.
   - Cluster 2 has high percent of singles(31.31%). 59.89% of the people in this cluster are married. This cluster has one of the lowest numbers of divorcees(8.79%).
   - Cluster 3 has the second-highest percentage of divorcees(11.67%). It has a high percent of married Customers, 60.37.
   - Cluster 3 has the highest percent of Married Customers among all the four clusters. And it has the lowest Divorce(8.69%) rate within all the clusters. It also has the lowest number of Singles(23.91%).

4. Education:
   - Maximum Customers from cluster 1 have pursued Secondary(51.4%) education, which is the highest among all the four clusters. Second-highest percentage is of Tertiary education(31.16%) in this cluster.
   - 44.5% people of cluster 2 have completed Tertiary education. And 36.81% have secured up to Secondary Education.
   - Cluster 3 has high percent(14.39%) of people having secured Primary Education. This also has a high percent of people whose Education details are unknown.
   - Cluster 4 has the maximum number of people having secured Primary Education(15.22%) and Tertiary Education(54.33%) out of all the Clusters. This cluster has the lowest percent of people securing up to Secondary Education(26.1%) and Unknown Education(4.35%).

5. Defaulters:
   - Cluster 1 has 1.879% and Cluster 3 has 0.104% of defaulters.
   - There are no defaulters in Cluster 2 and Cluster 4.

6. Balance:
   - Cluster 1 has Customers who deposit very low amounts(-500 to 2000).
   - Cluster 2 has heavy deposits ranging from 6000 to 14000.
   - Cluster 3 has mediocre level deposit amounts ranging from 2000 to 6000.
   - Cluster 4, by far is the most productive Cluster with deposits in ranges 20,000 to 30,000.

7. Housing:
   - 49.536% people of Cluster 1 have Housing available.
   - 35.44% people of Cluster 2 have Housing available.
   - 39.93% people of Cluster 3 have Housing available.
   - 21.74% people of Cluster 4 have Housing available.

8. Loans:
   - 14.81% people of Cluster 1 have taken loans.
   - 4.12% people of Cluster 2 have taken loans.
   - 6.93% people of Cluster 3 have taken loans.
   - 6.52% people of Cluster 4 have taken loans.

9. Mode of Contact:
   - Maximum contacts are done through cellular mode of communication (more than 70%) in each cluster.
   - Minimum contacts are carried out through Telephone (less than 20%) in each cluster.

10. Other Parameters:

Remaining Parameters demonstrate similar traits between all the clusters and have been mentioned together below

- Day - Customers were mostly contacted in the range 10-20 in dates.
- Month - Customers were most active in February, May to August and in November.
- Duration of contact - The Maximum duration range was 0 to 400 seconds per call on average. Very less calls were held beyond this point.

Efficiency of Clusters

Sr No	**No. Of Deposits(no)**	**Size of Cluster(len)**	**Efficiency(no / len)**
Cluster 1	3940	8834	0.4460041
Cluster 2	209	364	0.5741758
Cluster 3	1115	1918	0.5813347
Cluster 4	25	46	0.5434738

Conclusions from the Above Table:

- Cluster 1 has the lowest efficiency (44.6%)
- Cluster 2 and Cluster 3 have similar efficiency 57.42% and 58.13% respectively.
- Cluster 4 has a moderate efficiency with 54.35%.

Conclusions

We observed that Cluster 1 has the lowest efficiency and their deposit amounts are also very low (-500 to 2000). So, the banking institution should invest lesser efforts in the Applicants resembling the traits from this organization.

Cluster 3 has high efficiency(58.13%) but their deposit contributions are also very low as compared to other clusters.

The Banking Institution should concentrate more on customers who have traits from the Cluster 2 and 4.

These two clusters have people depositing large amounts and the number of persons in the cluster are lower as compared to cluster 1 and 3. In the next marketing campaign, the institution should work towards increasing the persons from cluster 2 and 3.

Traits of an Ideal Applicant:

- Age should be between 25 to 60 years.
- Job profile should be management / Technician.
- Marital Status should be Married.
- The applicant should have completed Tertiary level of Education.
- He should not be defaulter and his housing and loans should be minimal.
- Preferred mode of communication should be Cellular and must occur largely in mid-year(May to August).
- Duration of every phone call from customer care executive should be below 400 seconds.

The person possessing all of these traits will be highly profitable for the banking institution.

R Script

This script contains all the functions and values used for transforming, cleaning and analysing the data. Plotting functions like plot(),hist() have been extensively used for visulizations and deriving useful observations.

```
deposit <-
rep(0,length(x$deposit))
table(deposit)
for(i in 1:length(x$deposit))
  if(x$deposit[i]=="yes")
    deposit[i]=1
table(deposit)

default <-
rep(0,length(x$default))
table(default)
for(i in 1:length(x$default))
  if(x$default[i]=="yes")
    default[i]=1
table(default)

housing <-
rep(0,length(x$housing))
table(housing)
for(i in
1:length(x$housing))
  if(x$housing[i]=="yes")
    housing[i]=1
table(housing)

loan <-
rep(0,length(x$loan))
table(loan)
for(i in 1:length(x$loan))
  if(x$loan[i]=="yes")
    loan[i]=1
table(loan)

admin <-
rep(0,length(x$job))
table(admin)
for(i in 1:length(x$job))
  if(x$job[i]=="admin.")
    admin[i]=1
table(admin)
```

```
blue_collar <-
rep(0,length(x$job))
table(blue_collar)
for(i in 1:length(x$job))
  if(x$job[i]=="blue-collar")
    blue_collar[i]=1
table(blue_collar)

entrepreneur <-
rep(0,length(x$job))
table(entrepreneur)
for(i in 1:length(x$job))

if(x$job[i]=="entrepreneur"
)
    entrepreneur[i]=1
table(entrepreneur)

housemaid <-
rep(0,length(x$job))
table(housemaid)
for(i in 1:length(x$job))
  if(x$job[i]=="housemaid")
    housemaid[i]=1
table(housemaid)

management <-
rep(0,length(x$job))
table(management)
for(i in 1:length(x$job))

if(x$job[i]=="management
")
    management[i]=1
table(management)

retired <-
rep(0,length(x$job))
table(retired)
for(i in 1:length(x$job))
  if(x$job[i]=="retired")
    retired[i]=1
table(retired)
```

```
self_employed <-
rep(0,length(x$job))
table(self_employed)
for(i in 1:length(x$job))
  if(x$job[i]=="self-
employed")
    self_employed[i]=1
table(self_employed)

services <-
rep(0,length(x$job))
table(services)
for(i in 1:length(x$job))
  if(x$job[i]=="services")
    services[i]=1
table(services)

student <-
rep(0,length(x$job))
table(student)
for(i in 1:length(x$job))
  if(x$job[i]=="student")
    student[i]=1
table(student)

technician <-
rep(0,length(x$job))
table(technician)
for(i in 1:length(x$job))
  if(x$job[i]=="technician")
    technician[i]=1
table(technician)

unemployed <-
rep(0,length(x$job))
table(unemployed)
for(i in 1:length(x$job))

if(x$job[i]=="unemployed")
    unemployed[i]=1
table(unemployed)

unk_job <-
rep(0,length(x$job))
table(unk_job)
for(i in 1:length(x$job))
  if(x$job[i]=="unknown")
    unk_job[i]=1
table(unk_job)

single <-
rep(0,length(x$marital))
table(single)
for(i in 1:length(x$marital))
  if(x$marital[i]=="single")
    single[i]=1
table(single)

divorced <-
rep(0,length(x$marital))
table(divorced)
for(i in 1:length(x$marital))

if(x$marital[i]=="divorced"
)
    divorced[i]=1
table(divorced)

married <-
rep(0,length(x$marital))
table(married)
for(i in 1:length(x$marital))

if(x$marital[i]=="married")
    married[i]=1
table(married)

table(x$education)

primary <-
rep(0,length(x$education))
table(primary)
for(i in
1:length(x$education))
```

```
  if(x$education[i] ==
"primary")
    primary[i]=1
table(primary)

secondary <-
rep(0,length(x$education))
table(secondary)
for(i in
1:length(x$education))
  if(x$education[i] ==
"secondary")
    secondary[i]=1
table(secondary)

tertiary <-
rep(0,length(x$education))
table(tertiary)
for(i in
1:length(x$education))
  if(x$education[i] ==
"tertiary")
    tertiary[i]=1
table(tertiary)

unk_education <-
rep(0,length(x$education))
table(unk_education)
for(i in
1:length(x$education))
  if(x$education[i] ==
"unknown")
    unk_education[i]=1
table(unk_education)

table(x$contact)

cellular <-
rep(0,length(x$contact))
table(cellular)
for(i in 1:length(x$contact))
  if(x$contact[i] ==
"cellular")
    cellular[i]=1
table(cellular)

telephone <-
rep(0,length(x$contact))
table(telephone)
for(i in 1:length(x$contact))
  if(x$contact[i] ==
"telephone")
    telephone[i]=1
table(telephone)

unk_contact <-
rep(0,length(x$contact))
table(unk_contact)
for(i in 1:length(x$contact))
  if(x$contact[i] ==
"unknown")
    unk_contact[i]=1
table(unk_contact)

data<-
data.frame(x$age,admin,bl
ue_collar,entrepreneur,

housemaid,management,re
tired,self_employed,

student,services,technician,
unemployed,unk_job,

single,married,divorced,pri
mary,secondary,

tertiary,unk_education,defa
ult,x$balance,housing,

loan,cellular,telephone,unk
_contact)
View(data)
head(data)
```

```
temp <- data
head(temp)
b = rep(0,10)
for(i in 1:10)
{
  v <- kmeans(temp,i)
  b[i] = v$withinss
}
plot(seq(1,10,1),b,"l")

v <- kmeans(temp,4)
v$centers
v$size

table(x$day)
table(x$month)

month <-
rep(0,length(x$month))
for(i in 1:length(x$month))
{
  if(x$month[i] == "jan")
    month[i] = 1
  else if(x$month[i] ==
"feb")
    month[i] = 2
  else if(x$month[i] ==
"mar")
    month[i] = 3
  else if(x$month[i] ==
"apr")
    month[i] = 4
  else if(x$month[i] ==
"may")
    month[i] = 5
  else if(x$month[i] ==
"jun")
    month[i] = 6
  else if(x$month[i] ==
"jul")
    month[i] = 7
  else if(x$month[i] ==
"aug")
    month[i] = 8
  else if(x$month[i] ==
"sep")
    month[i] = 9
  else if(x$month[i] ==
"oct")
    month[i] = 10
  else if(x$month[i] ==
"nov")
    month[i] = 11
  else if(x$month[i] ==
"dec")
    month[i] = 12
}
table(month)
table(x$pdays)
table(x$poutcome)

failure <-
rep(0,length(x$poutcome))
table(failure)
for(i in
1:length(x$poutcome))
  if(x$poutcome[i] ==
"failure")
    failure[i]=1
table(failure)

other <-
rep(0,length(x$poutcome))
table(other)
for(i in
1:length(x$poutcome))
  if(x$poutcome[i] ==
"other")
    other[i]=1
table(other)

success <-
rep(0,length(x$poutcome))
```

```
table(success)
for(i in
1:length(x$poutcome))
  if(x$poutcome[i] ==
"success")
    success[i]=1
table(success)

unk_poutcome <-
rep(0,length(x$poutcome))
table(unk_poutcome)
for(i in
1:length(x$poutcome))
  if(x$poutcome[i] ==
"unknown")
    unk_poutcome[i]=1
table(unk_poutcome)

data<-
data.frame(x$age,admin,bl
ue_collar,entrepreneur,

housemaid,management,re
tired,self_employed,

student,services,technician,
unemployed,unk_job,

single,married,divorced,pri
mary,secondary,

tertiary,unk_education,defa
ult,x$balance,housing,

loan,cellular,telephone,unk
_contact,x$day,month,

x$duration,x$campaign,x$
pdays,x$previous,failure,

other,success,unk_poutcom
e)

temp <- data
head(temp)
b = rep(0,10)
for(i in 1:10)
{
  v <- kmeans(temp,i)
  b[i] = v$withinss
}
plot(seq(1,10,1),b,"l")

v <- kmeans(temp,4)
v$centers
v$size
#,xlim=c(1,200),ylim=c(-
1,1000)
hist(temp[v$cluster==1,1])
table(temp[v$cluster==4,1
])

table(x$previous)
table(x$pdays)

table(x$job)
table(x$marital)

final <-
data.frame(data,deposit)
View(final)

f <- kmeans(temp[1:37],4)
f$size
f$centers

c1_outcome <-
final[f$cluster==1,38]
sum(c1_outcome)/
length(c1_outcome)

c2_outcome <-
final[f$cluster==2,38]
sum(c2_outcome)/
length(c2_outcome)
```

```
c3_outcome <-
final[f$cluster==3,38]
sum(c3_outcome)/
length(c3_outcome)

c4_outcome <-
final[f$cluster==4,38]
sum(c4_outcome)/
length(c4_outcome)

?kmeans

sum(c4_outcome)
f$size
```