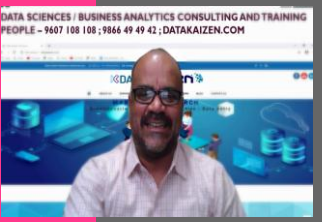
Data Analysis Using R Introduction to Business Analytics Data Sciences

Dr.M.Raghunadh Acharya

PGDCA BEd MSc(Statistics)

MBA(Marketing) PhD(Statistics)



Data Analysis Using R

The lecture Outlines

Part – I

Business Analytics ...

Data Sciences ...

Part – II

Applications ...

Nature of Data ...

Part – III

Parts of Statistics ...

Part – IV

Why Open Sources ...

Outlines of R ...

Lecture -- II

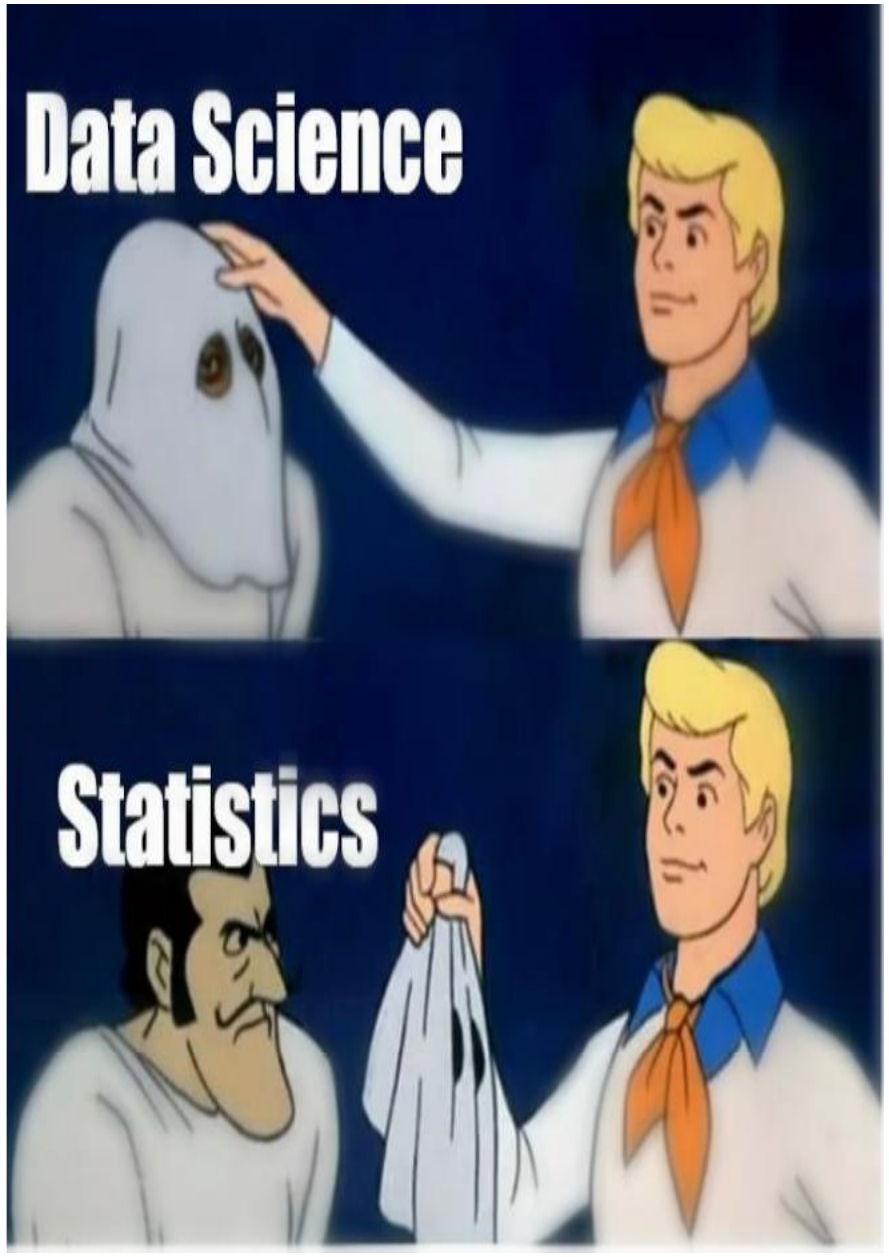
Working with R ...

- Introduction…

Data Analysis Using R

- Before we proceed to the Session let us watch the following …

- I got one interesting figure in a whatsup group



The need for knowing the three R's , reading , writing and arithmetic , is well understood.

These do not take us far unless we acquire the fourth R , reasoning under uncertainty, for taking decisions in real life.

Padma Vibhushan Prof. C.R.Rao
(Statistics and Truth)

Calyampudi Radhakrishna Rao giving a talk at Indian Statistical Institute. Chennai in April 2012

Definitions of **Statistics** and **Applications**

" The quite statisticians have changed the world not by developing new facts or technical developments but by changing the way we reason , experiment , and form our opinions about it "

- Hacking

" Scientific laws are not advanced by the principle of authority or justified by faith or medieval philosophy ; statistics is the only court of appeal to new knowledge "

- P.C.Mahalanobis

"Statistics is the technology of finding the invisible and measuring the immeasurable "

Analytics is Analysis + Techniques

The terms and Data Sciences and Business Analytics to be understood before we proceed further

Data Analysis Using R

- The term Data Sciences in day to day sense I understand it as

"Applied Statistics "

" in broader sense , which include Statistics Operations Research Data Mining & Deep Learning "

A Branch of Data Sciences applied to Business is referred to as Business Analytics

Statistics

Summary Measures ; Probability ; Sampling ; Relationship analysis ; Multivariate Data Analysis Techniques;

Operations Research

Mathematical Modeling ; LPP ; TP ; AP ; IPP; DP; GP …

Data Mining

Associations ; Naïve Bayes ; CART & CHAID …

Deep Learning

Genetic Algorithms ; Machine Learning ; Artificial Intelligence

Data Analysis Using R

- Different Areas of Business Analytics
- Descriptive Analytics ....
- Diagnostic Analytics ....
- Predictive Analytics ....
- Prescriptive Analytics ....

Data Analysis Using R

- Quick brief on each of the areas
- Descriptive Analytics ....

  this will indicate on what happened in the past based on the past data

  this can be in the form of tables , graphs , dash boards etc.,

  for example ... for a particular brand how many are purchasing on line and how many are purchasing off line.

  basically to understand the profile of the customers who are purchasing online and who are purchasing offline.
- Diagnostic Analytics ....

  this can be understood as a step further to Descriptive analytics , this undertakes a deeper analysis to answer a question ,

  in the earlier examples the company wishes to find out on why certain profile customers are preferring to shop online ...

  this will make use of advanced data analysis techniques like ..., Multivariate Techniques ... Data Mining ....

Data Analysis Using R

- Quick brief on each of the areas
- Predictive Analytics ....

  this makes use of past data and develops models to predict the future , which can be used for planning organizing etc.,

  for example the company can use the models to predict the likely customers who are likely to purchase on line and who likely to purchase off line etc., when a customer profile is given , the event that he is likely to purchase online or offline., Will make use of Machine Learning , Deep Learning , Artificial Intelligence
- Prescriptive Analytics ....

  this takes the predictive analytics to the next level , it not only predicts the future and helps the companies on future steps to be taken

  for example : how many customers likely to purchase on line and how much staff to be recruited in future etc., this will make use of Data Mining , Machine Learning , Operations Research Techniques.

- **Applications of Data Sciences / Business Analytics**

Data Analysis Using R

Applications of Business Analytics / Data Sciences in short

I wish to make one point clear that bigger or fancy techniques not much required

Simple / Basic techniques are sufficient for solving most of the BA problems / Situations.

For advanced analysis , advanced model building activities , one need advanced techniques.

In a nutshell I can state that BA applications are present every where , the more you

apply the more scope you notice for the applications.

Let us dwell on some naive applications of BA

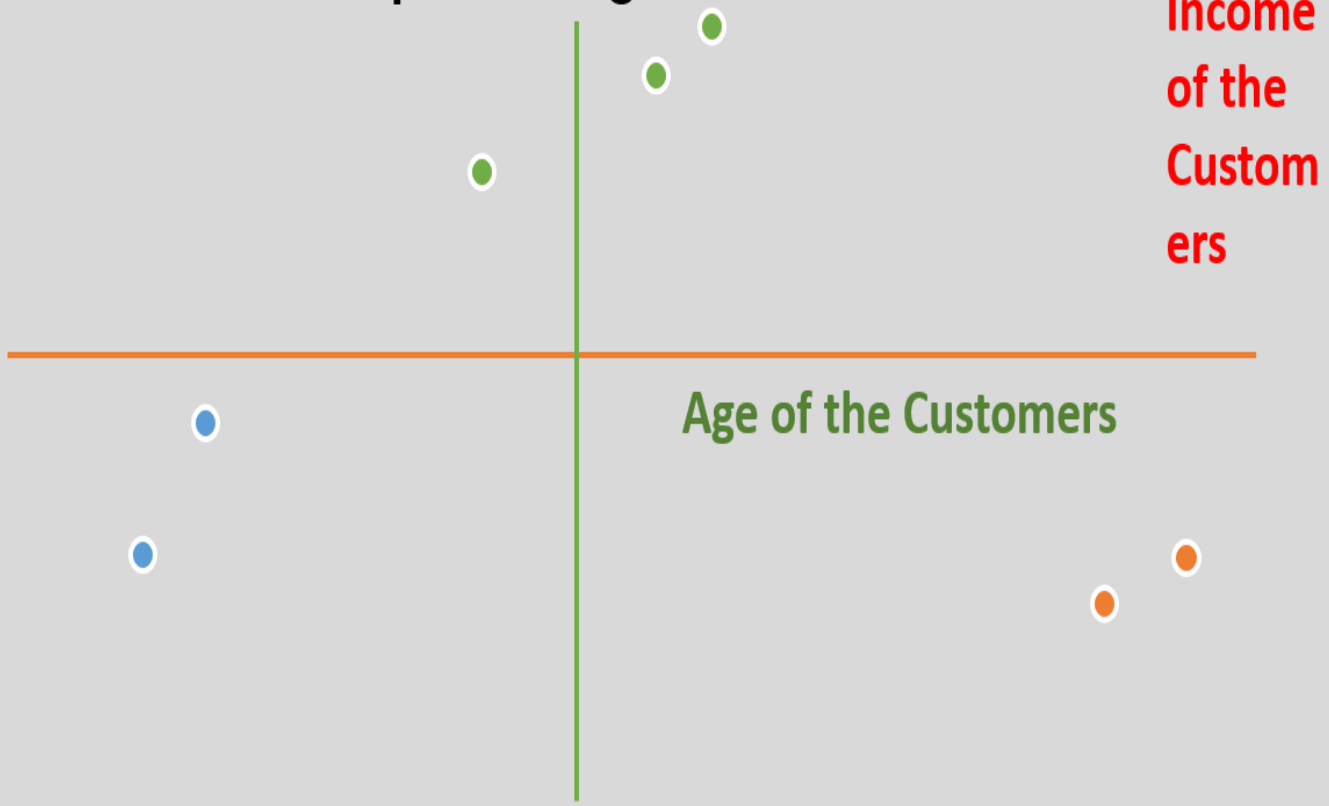
Data Analysis Using R

1

A company wishes to segment its customers so as to serve them better and gain loyalty thus facilitating repeated purchase , references etc.,

For example you have the data of customers with respect to age and income

S.No.	Customer	Age	Income (p.m.)
1	Customer_1	23	21000
2	Customer_2	32	24000
3	Customer_3	37	48000
4	Customer_4	62	34000
5	Customer_5	90	12000
6	Customer_6	71	18000
7	Customer_7	63	45000
8	Customer_8	54	65000
9	Customer_9	53	54000
10	Customer_10	79	10000
11	Customer_11	59	48000
12	Customer_12	28	23000
13	Customer_13	52	45000



Data Analysis Using R

2

Identifying the Churning of Customers … by using Customers data making use of Statistical and Data Mining tools

Can I know few days / weeks / months before the customers likely to leave your service.

If the telco knows the likely churning customers much before it can take measures to retain the customers … since acquiring new customers is very expensive

One use basic tools like Cross tabulations , Chi square tests ; etc. or one can take up slightly advanced tools like Logistic regression , Discriminant Analysis etc.,

Data Analysis Using R

 # 3

Which new product has the highest chances of success

This is very interesting as a Company wish to release a product in to market and ten of their Competitors as well willing to introduce a similar product , Viz., CORONA gear

Company wishes to know on what is their success rate in relation to release of the competitors products.

Competitors analysis … Positioning analysis etc.,

 # 4

Finding prospective customers from a large group of customers

5

BA / DS capabilities connect people with trusted, relevant information, enabling them to confidently predict, plan and act.

Which transactions are fraudulent … to indicate among the transactions made which of them are not genuine , fraudulent.

They work better in real time on enterprise data , the mechanism can alert the customers / vendors real time thus save customer money and for companies they save lot of trust and value.

Data Analysis Using R

6

Let us see some heath care applications , a patient has seen with some vertigo or neuro problems and he has been falling down often.

The past historical data of the patient and historical data of similar patients and gives an alert to the patient and the attendees and the doctors thus averting the fatalities as well giving the health care with immediate effect.

Jobs and Careers in Data Sciences / Business Analytics

Data Analysis Using R

Let us see some roles in the industry prevailing

There are different roles in this area

Programmers ... *they develop program / code for execution of a task viz., R Programmers , PYTHON Programmers , SAS Programmers , tec.,*

Data Administrators or Data Base Administrators ... *Persons involved in managing the data , data based , Big data etc., their job is to maintain the data , retrieve the data etc.,*

Business Analysts ... *this role requires to analyse the requirements of the business at clients end , come back and explain the requirements to the associates.*

Business Analytics Professionals / Data Scientist *... this is the most heard and most exciting job among all the above mentioned.*

All the above profiles are routine except the last one and you get tired after doing it for some time ...

They are like any other IT jobs.

- ***Frame Work of Business Analytics / Data Sciences***

Data Analysis Using R

Data sciences has the following stages

Data Collection

Data Storing

Data Retrieving

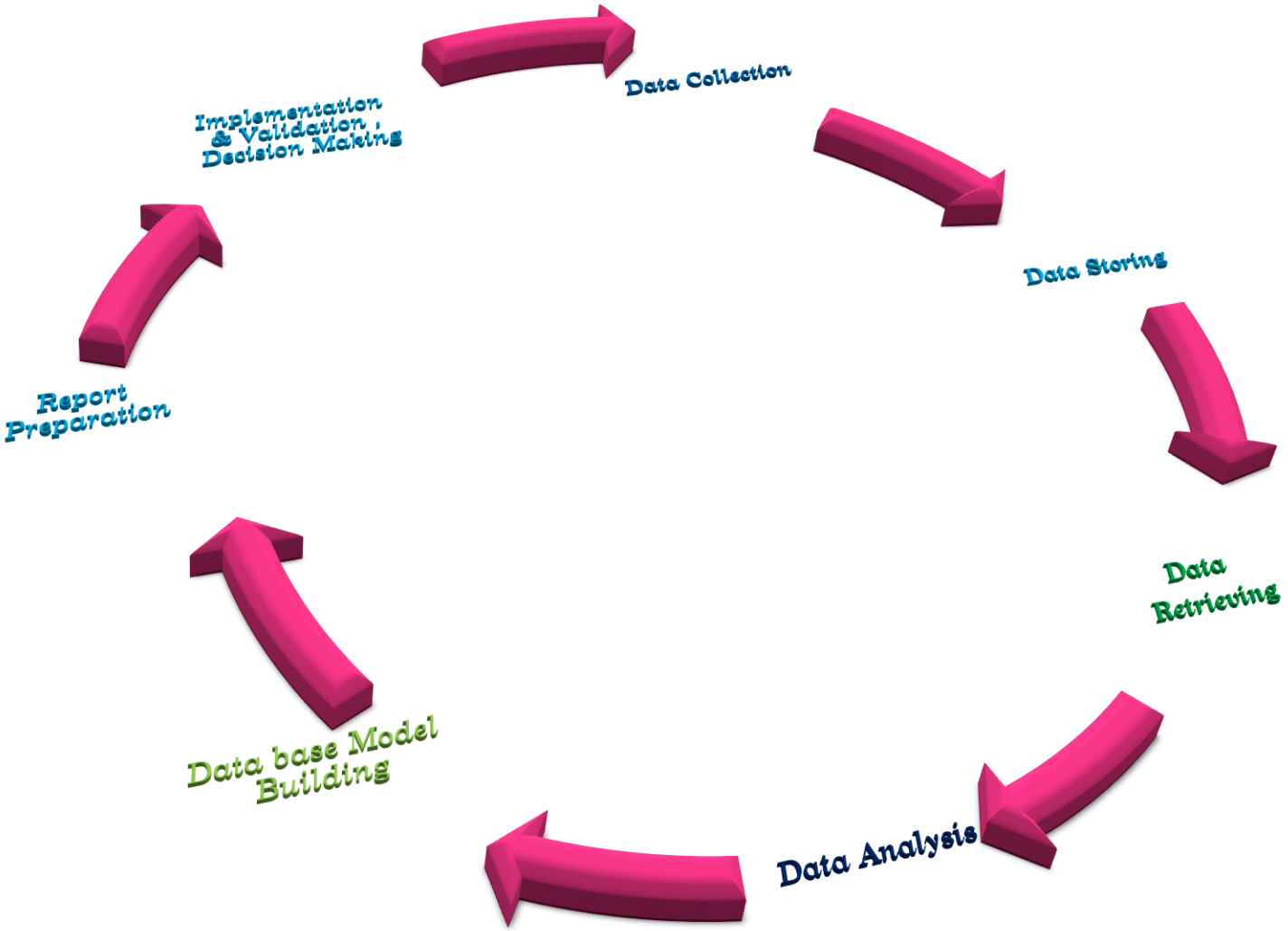
Data Analysis

Data base Model Building

Report Preparation

Implementation & Validation , Decision Making

Data Analysis Using R



Data Analysis Using R

- Why Open sources

  Architecture of R , R studio and PYTHON

  Demonstration of R , R Studio and Python

Data Types

- Data by means of Collection …

  Primary

  Secondary

- Data by means of its Nature …

  Measurement and Scaling