

Psychic Ads

Statistical Analysis of *Google Trends* Time Series in R Language for targeted advertising on Adwords

By Kavya M N, Nishita Rai

Advisor: Prof. Mohsin Ahmed

JKSHIM, NITTE India 574110



Nitte University
Nitte Education Trust
Quality Education - an Abiding Commitment

Agenda

- Motivation
- Introduction to targeted advertising
- Google Adwords, Trends, Forecasting
- Example of *MBA* queries
- Psychic Ads
- Demo: Processing Time Series in R Language
- Conclusion
- References
- Acknowledgements
- Question and Answers

Motivation

- An Institute needs to identify future students much before the admission season begins.
- Typical print advertising is unable to target the correct audience.
- Using Google search data analytics, we predict future students for MBA.

Introduction: Targeted Advertising

- In Traditional marketing, most of the ads are seen by people not even interested in the product.
- Internet has changed that, companies like Google and Facebook know the demographics and preferences of their users.
- Google aggregates all user queries daily and statistically analyses it to match users and advertisers.

Google Adwords

- Advertising on Google is easy for small companies, even with budgets in thousands.
- The advertiser selects “keywords” on which to advertise.
- When the user types that keyword into google, they will also see the Ad.
- If the user clicks on the Ad, google will charge the advertiser, and the advertiser gets a customer interested in their product.

Advertiser's work

- E-Marketing allows the advertisers to chose who, where, when and how will see their ads.
- Advertisers need to pick keywords in adwords and prices they are willing to pay for their ads.
- This is an art, advertisers must guess what user queries are relevant to their ads.

Examples of Google Adwords

- JKSHIM, IIMs, TAPMI can all advertise on the keywords “MBA”, “Higher Education”, “Business Classes”,

The screenshot shows a Google search for "mba college". The search bar at the top contains the text "mba college" and a magnifying glass icon. Below the search bar, the navigation tabs include "Web", "Maps", "Images", "News", "Videos", "More", and "Search tools". The search results indicate "About 136,000,000 results (0.27 seconds)".

The first organic result is titled "MBA College - AICTE approved PGDM-100% Placement" with the URL "www.dpsr.in/mba" and the phone number "098689 74097". It includes the text "Guaranteed. Apply for Scholarship" and "Apply Online - Program Details - Career Seminars - DSPSR Advantage".

The second organic result is titled "Top MBA Colleges in India - Top B Schools in India." with the URL "www.shiksha.com/top-mba-colleges-in-india-rankingpage-2-2-0-0-0". It includes the text "Search for Top MBA Colleges in India. Get a list of Top 10, 20 MBA Colleges in India categorized by Rank, location, specializations, course fees and cutoff." and "Indian Institute of Management - MBA Finance - Faculty of Management Studies".

Two paid advertisements are circled in red. The first ad is titled "MBA Admissions 2015" with the URL "www.snu.edu.in/MBA-Admission" and the text "MBA with Focus On Entrepreneurial Leadership @ Shiv Nadar University!". The second ad is titled "MBA Admissions 2015-17" with the URL "www.iibsonline.com/apply-now" and the text "Current Placement Offered 22 Lakhs" and "Get Free MBA Prospectus, Apply Now!".

Google Trends

- We use google.com/trends to statistically analyses the user queries across time.
- Example:
<http://www.google.com/trends/explore#q=MBA+Colleges>



Google trends location data



- <http://www.google.com/trends/explore#q=MBA+Colleges>

MBA College queries in India



<http://www.google.com/trends/explore#q=mba+colleges&geo=IN>

Google trends related keywords

Related searches ?

Topics

Top Rising

MBA - Degree

100



College - Degree

95



Queries

Top Rising

mba top colleges

100



mba in india

75



colleges for mba

70



best mba colleges

30



delhi mba colleges

25



bangalore mba colleges

25



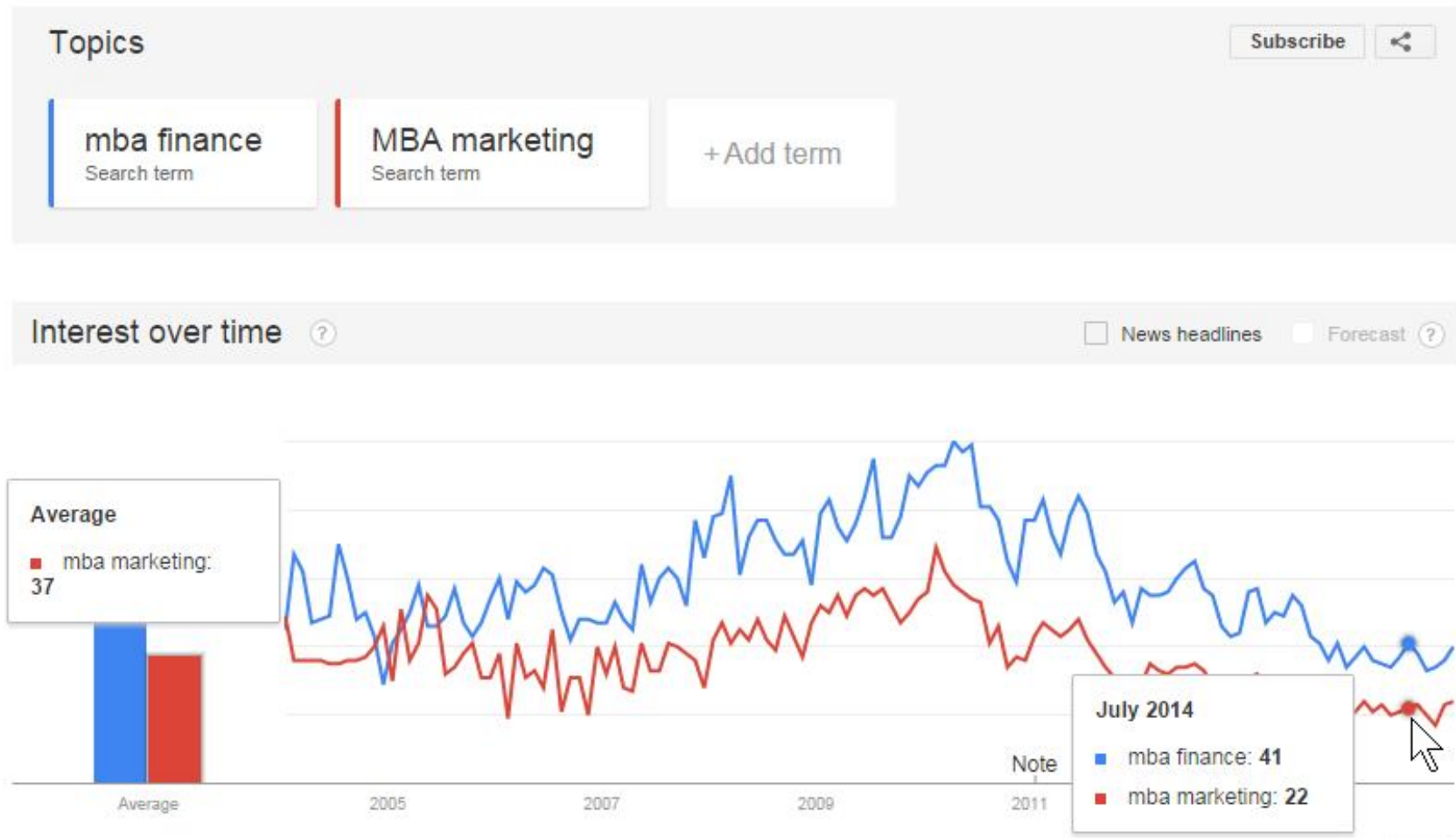
mba in delhi

25

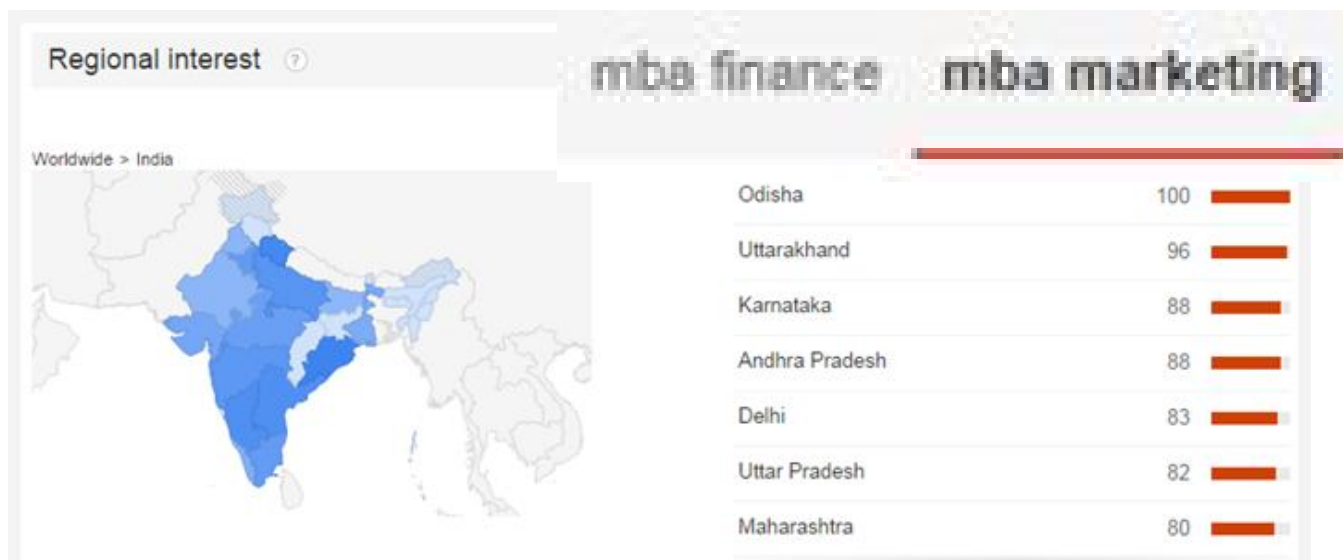


Comparing MBA Specializations

with <http://www.google.com/trends/explore#q=MBA%20finance%2C%20MBA%20Marketing&cmpt=q>



Comparing MBA Specializations



Correlating Time series

Example: What did users search 2 months before *MBA Admissions*? Answer: *Financial Risks*

The screenshot shows the Google Correlate interface. The search bar contains 'mba admissions' and a 'Search correlations' button. Below the search bar is a checkbox labeled 'Exclude terms containing mba admissions'. On the left side, there are links for 'Compare US states', 'Compare weekly time series', and 'Compare monthly time series'. The 'Compare monthly time series' link is highlighted with a red circle. Below these links, there is a 'Shift series' input field with '-2' and a 'months' label, and a 'Country' dropdown menu with 'India' selected. On the right side, there is a list of correlated terms with their correlation scores. The term 'financial risk' with a score of 0.8482 is highlighted with a red circle.

Google correlate

mba admissions

☐ Exclude terms containing mba admissions

[Compare US states](#)
[Compare weekly time series](#)
[Compare monthly time series](#)

Shift series months
Country:

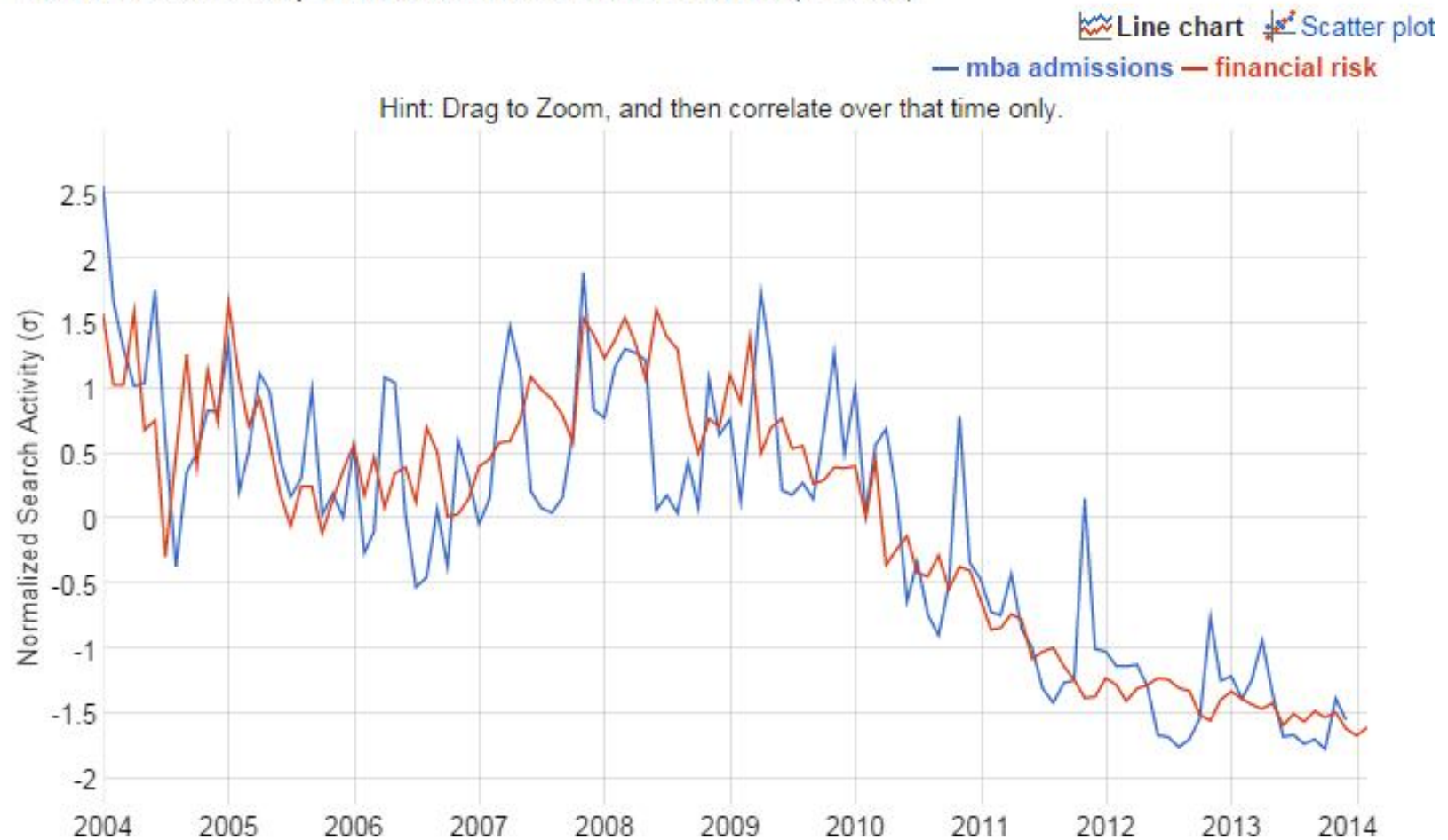
Documentation
[Comic Book](#)
[FAQ](#)
[Tutorial](#)

Correlated with mba admissions

- 0.8634 school in bangalore
- 0.8628 chipset driver
- 0.8563 cholamandalam
- 0.8560 links
- 0.8551 workflow
- 0.8537 .doc
- 0.8522 iso
- 0.8517 schools in
- 0.8501 erd
- 0.8488 acces
- 0.8482 financial risk
- 0.8474 sound drivers

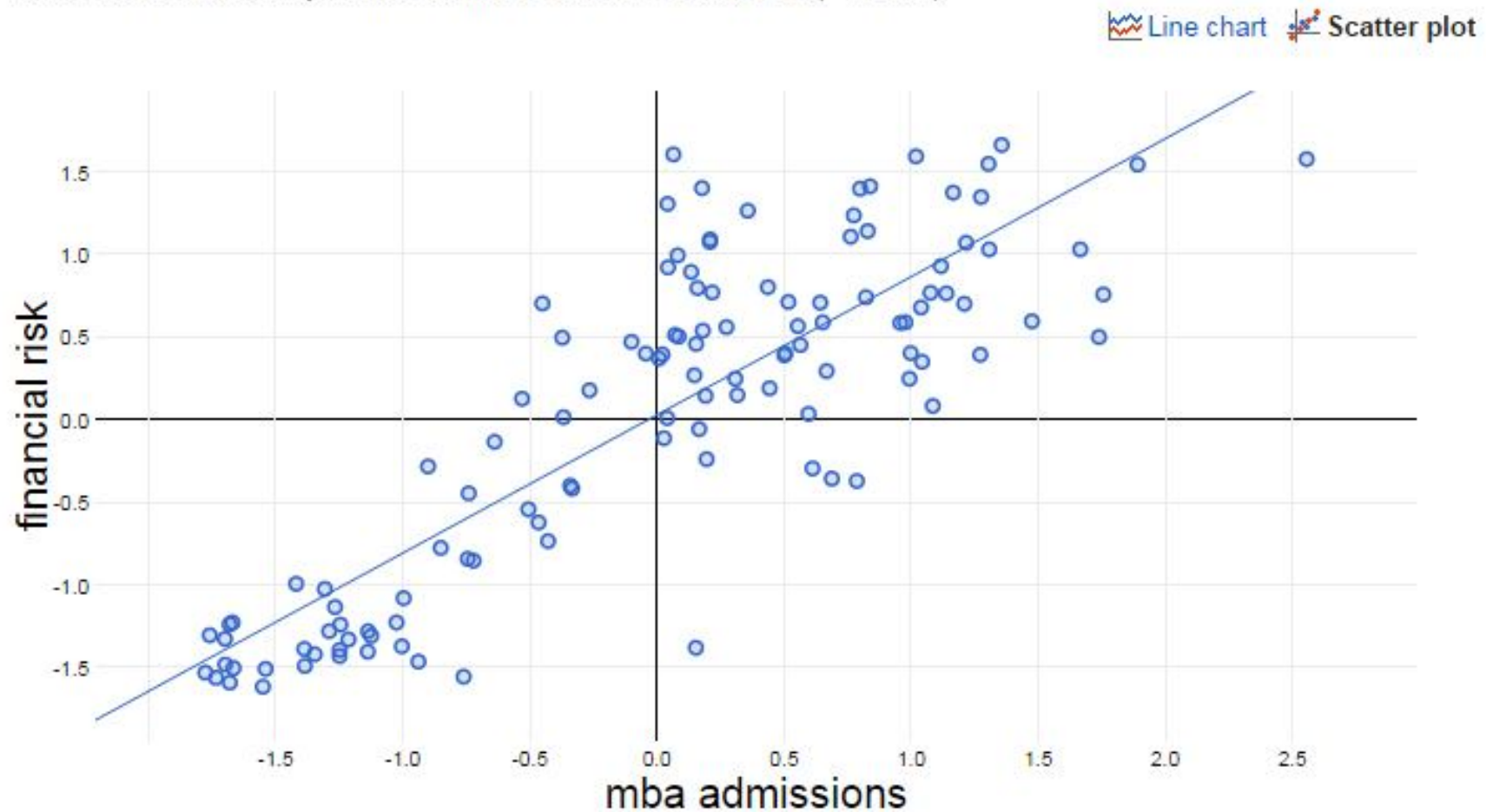
Correlation of *MBA Admissions* and *Financial Risk* (2 months earlier)

India Web Search activity for mba admissions and financial risk ($r=0.8482$)



Correlation of *MBA Admissions* and *Financial Risk* (2 months earlier)

India Web Search activity for **mba admissions** and **financial risk** ($r=0.8482$)



Psychic Ads

- So to advertise 2 months before for MBA course, the College can target students searching for “*Financial Risk*”.
- Note: There will be many correlated phrases, we must decide which ones are likely to target MBA students.
- Because the users have not thought of applying for MBA yet.
- This is psychic – predicting what our users will want in 2 months’ time.
- Next we show how to do this analysis in R.

Using R Code in R Studio

- Demo of correlating time series in R.

R Language Details

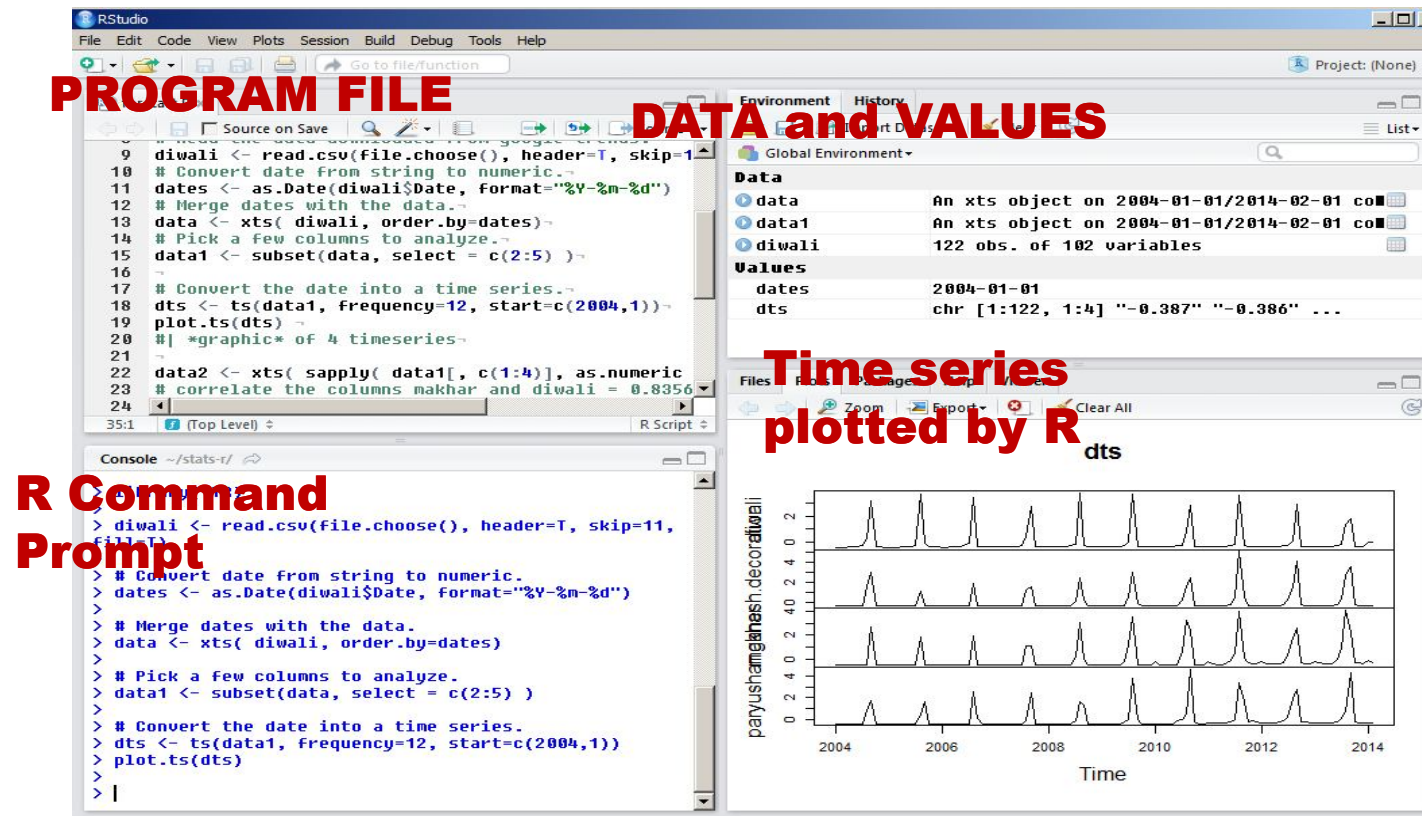
- We show how to use R to process Google Trends and Correlate data.
- Download the “*mba admissions*” query time-series and 2 months earlier queries for India, as csv files from <http://google.com/trends>
- Install R 3.1, R Studio to process the time-series csv data on Windows.
- Use the insights to plan targeted Ad campaigns on Google Adwords for MBA admissions 2 months early.

Analyse the Time Series in R

- Download the Google correlate data as a csv file.
- Analyze the data with R libraries:
 - XTS (Extendible Time Series)
 - TTR (Technical trading Rules)

The Time Series in R Studio

R Studio



Analyse the Time Series in R

Load the libraries to process the data.

library(xts) # eXtended Time Series library

library(TTR) # Technical Trading Rules library

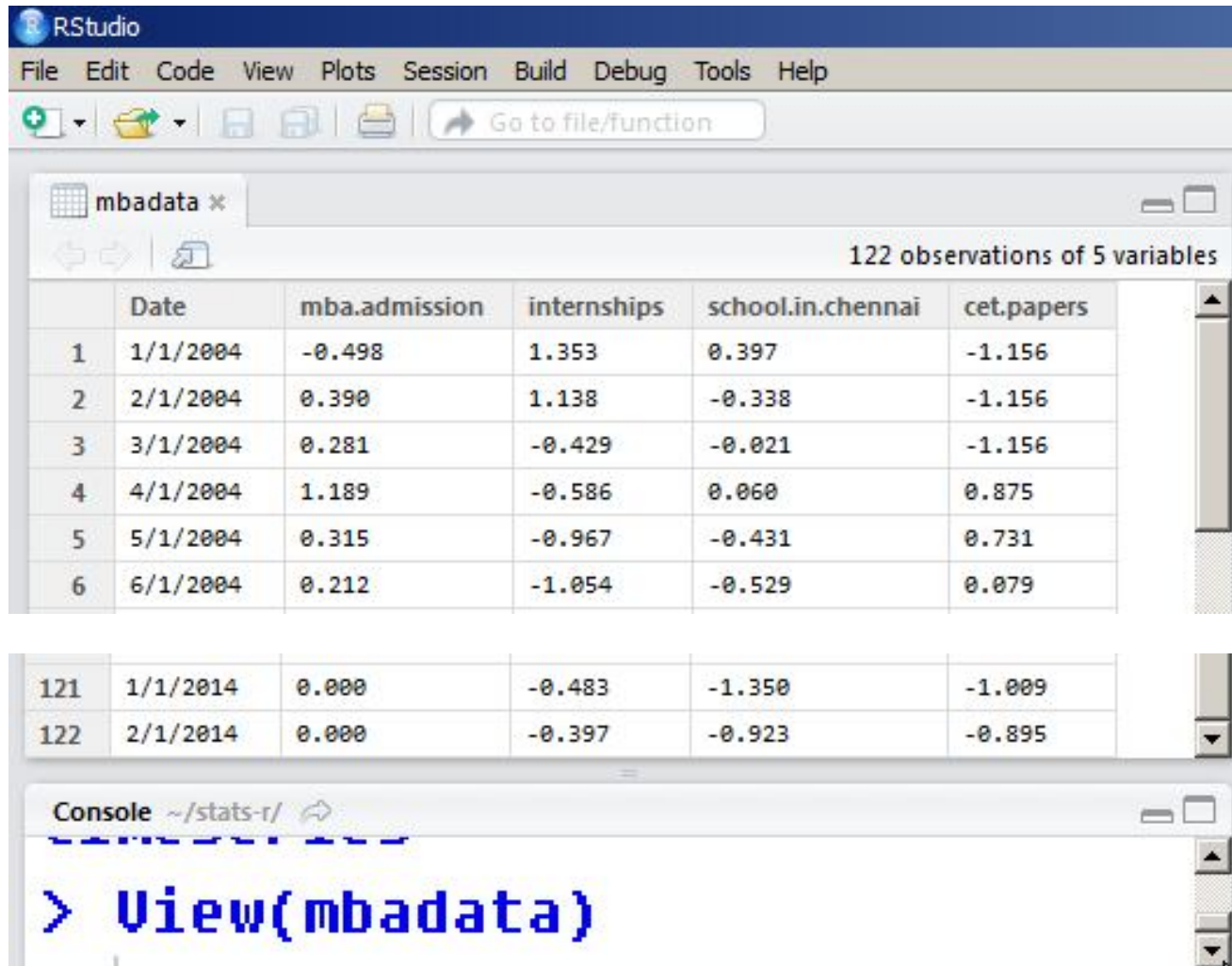
Read the data downloaded from Google trends.

Data correlates the query 'mba admissions'

with queries -2 months before.

```
mbadata <- read.csv(file.choose(), header=T,  
  skip=11, fill=T)
```

View the data we loaded from csv file



The RStudio interface displays the 'mbadata' data frame. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The toolbar shows icons for file operations and a search bar labeled 'Go to file/function'. The 'mbadata' window shows 122 observations of 5 variables. The data is displayed in a table with columns: Date, mba.admission, internships, school.in.chennai, and cet.papers. The console shows the command 'View(mbadata)'.

	Date	mba.admission	internships	school.in.chennai	cet.papers
1	1/1/2004	-0.498	1.353	0.397	-1.156
2	2/1/2004	0.390	1.138	-0.338	-1.156
3	3/1/2004	0.281	-0.429	-0.021	-1.156
4	4/1/2004	1.189	-0.586	0.060	0.875
5	5/1/2004	0.315	-0.967	-0.431	0.731
6	6/1/2004	0.212	-1.054	-0.529	0.079
...
121	1/1/2014	0.000	-0.483	-1.350	-1.009
122	2/1/2014	0.000	-0.397	-0.923	-0.895

```
> View(mbadata)
```

Analyse the Time Series in R

Convert date from string to numeric.

```
dates <- as.Date(mbadata$Date,  
  format="%Y-%m-%d")
```

Merge dates with the data.

```
data <- xts( mbadata, order.by=dates) )
```


Process the Time Series in R

Pick two columns to analyze.

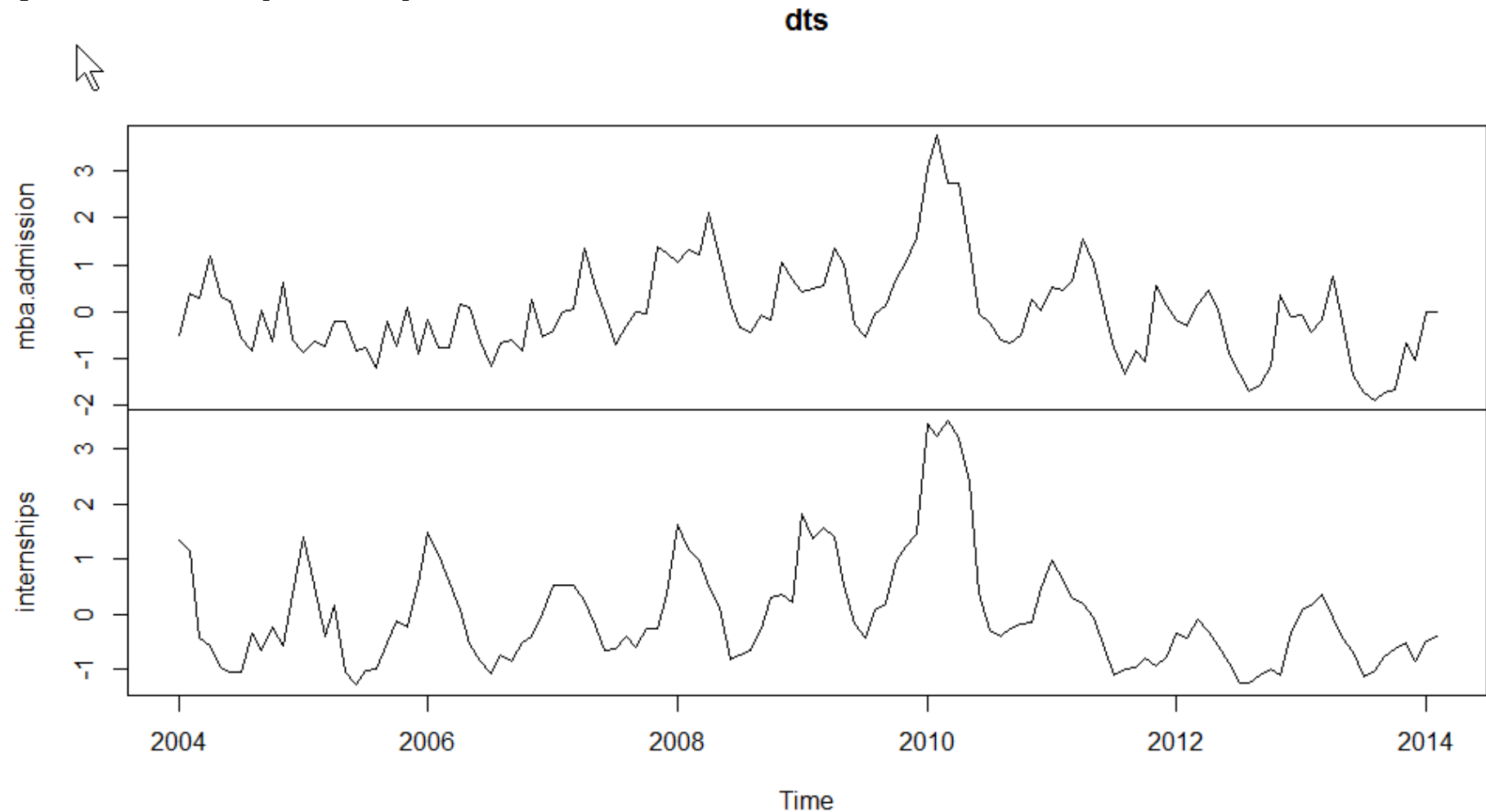
```
data1 <- subset(data, select = c(2:3) )
```

Convert the date into a time series.

```
dts <- ts(data1, frequency=12, start=c(2004,1))
```

Plot the two Time Series

```
plot.ts(dts)
```



Correlate two columns of Time Series

```
data2 <- xts( sapply( data1[, c(1:2)],  
  as.numeric ), order.by=dates)
```

```
# Correlate the columns
```

```
# 'mba admissions' and 'internships'.
```

```
cor(data2$mba.admission, data2$internships)
```

```
# We get r=0.7
```

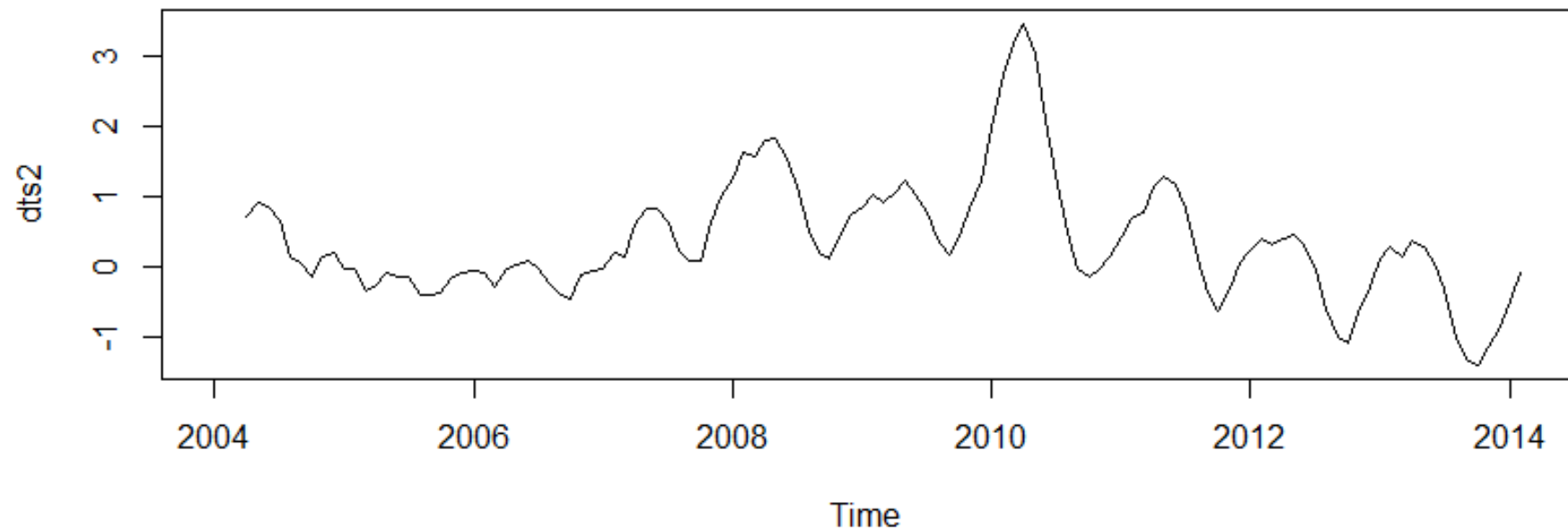
Smooth the time series to remove random fluctuations

Use the SMA package to Smooth the series.

```
dts2 <- SMA( ts(data2,frequency=12,  
  start=c(2004,1)), n=4)
```

Plot the smoothed TS

```
plot.ts(dts2)
```



Decompose the TS into Seasonal and Non-Seasonal Components

```
# Decompose the time series into  
# seasonal and  
# non-seasonal components.  
ddts2 <- decompose(dts2)
```

Plot Seasonal/Non-Seasonal TS

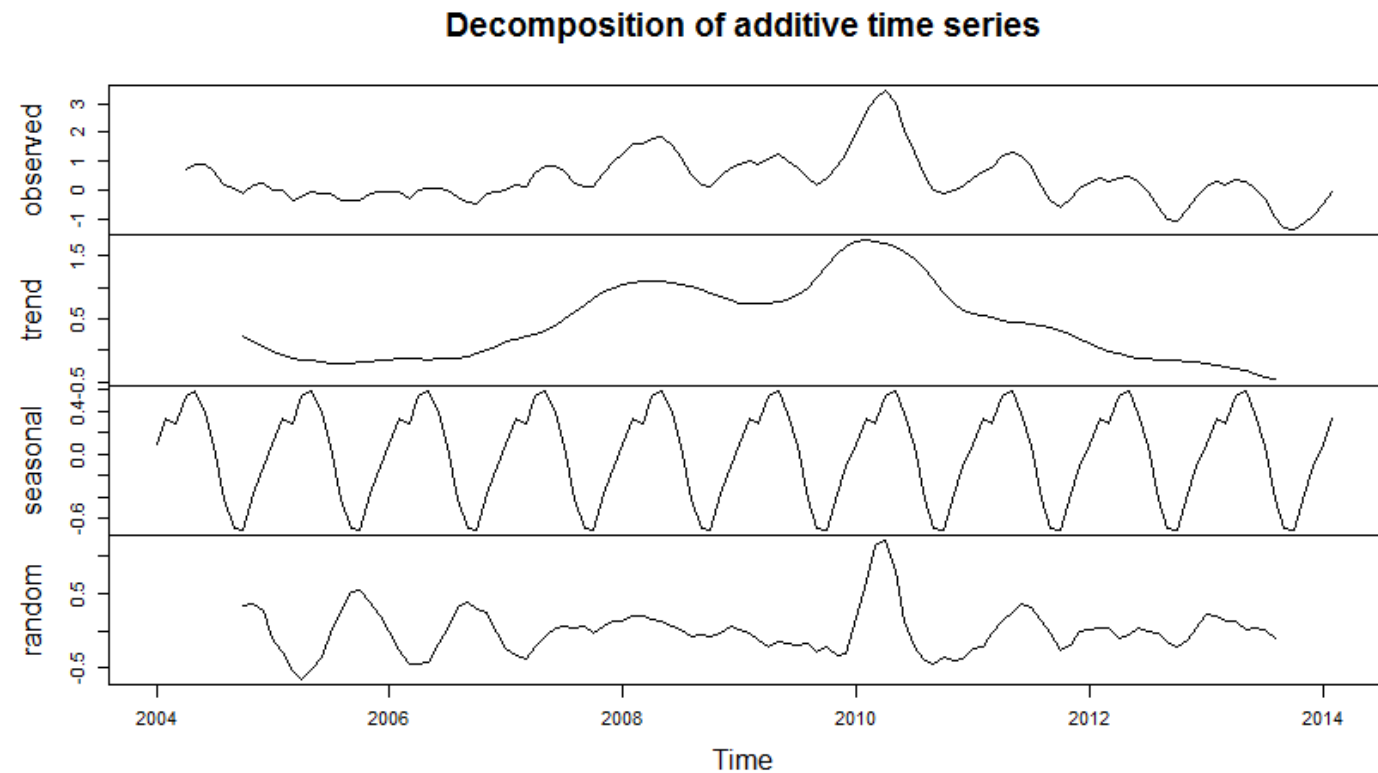
`plot(ddts2)`

Observed

Trend

Seasonal

Random



Conclusion

- Analyzing Time Series in R is easy.
- Visualizing the query trends over time gives us an insight into user behavior
- We use these techniques to find keywords for targeted Ads in Google Adwords.

Conclusion

- Targeted advertising is much more effective than traditional medium of advertising, with much higher CTR (click through rates) for advertisers. This keeps both users and advertisers happy.
- We take this a step further by predicting how to target users for future marketing, giving the advertiser a big advantage over their competitors.

References

1. Google Trends Data, from <http://google.com/trends>
2. Vanderkam, Schonberger, Rowley, Sanjiv Kumar “*Nearest Neighbor Search in Google Correlate*”, Tech Report 41694, Google Inc, from <http://research.google.com/pubs/pub41694.html>
3. Using R for Time Series Analysis, from <http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html>
4. Kavya M N, Nishita Rai, Jovita M, Nikitha F, Laxmi N, Meet A, “*R for MBA*”, Techbugs workshop, JKSHIM, NITTE.

Acknowledgements

- We thank our director Prof Dr.K.Sankaran for constant encouragement on the work.
- Prof. Krishna Prasad for help with the References and Publications.
- Prof. Sandhya Rao for Marketing lessons.
- Prof. Krishna Kumar for encouraging us to use R Language.
- Prof. Sheetal Kumar for organizing Techbugs' “*R for MBA*” workshop.

Question and Answers

- Thank you.
- Please email us your questions and suggestions.