

Psychic Ads: Identifying Students for Higher Education by Analysing Google Trends Time Series for Targeted Advertising on Google Adwords

Kavya M N*
Nishita Rai**
Mohsin Ahmed***

Abstract

In this paper we show how an Institute of Higher Education can predict future MBA students for targeted advertising by analysing Google query data. We extract normalized query data from Google trends and compute correlations of time shifted time series. We then conclude that highly correlated queries across time are suitable for targeted ads in Google Adwords.

Keywords: marketing, online, google, adwords, forecast, statistics, time series.

INTRODUCTION

We begin by explaining targeting advertising and Google Adwords. We then explain Google Trends and Google Correlate data models available to us for computing targeted advertisements in Adwords. In *Choi and Varian (2011)*, show how Google Correlate can be used to predict the present *Economic Indicators*, we take this a step further by predicting the future MBA students for targeted advertising.

We collected the time series of education related queries from Google correlate (*Google Data Trends 2014*). We use R to compute correlations of these time shifted time series (*Using R for Time Series Analysis, 2014*). This gives us a list of time correlated keywords. We use these keywords in Adwords to target psychic advertisements for prospective MBA students.

TARGET ADVERTISING

The main goal of an advertiser is to reach as many customers as effectively as possible. The advertiser should also be able to understand on how a customer's mind works.

For any product to be successful in the market, it should be effectively communicated to the user. To ensure this, advertisements are used. In *Traditional marketing*, most of the advertisements are seen by those, who are not even interested in the product. This is noticed in the chase of traditional roadside hoardings and bill boards, the advertisements on boards are seen by people who are not the prospective customers.

Targeted advertising is a new technique where the advertisements are placed so as to reach the right consumers based on their demographics, at the right time and right place.

For example, consider a traditional advertisement for a local pizza shop in newspaper print seen by readers across the city, how many would actually remember this Ad at lunch time?

*Student, Justice K S Hegde Institute of Management, NITTE 574110, India. kavya.nagraj4@gmail.com

**Student, Justice K S Hegde Institute of Management, NITTE 574110, India. nishita.raii2@gmail.com

***Student, Justice K S Hegde Institute of Management, NITTE 574110, India. moshahmed@gmail.com

It would be better to *target* this Ad to the mobile phones of students between 11am and 1pm in the vicinity of the restaurant. An even more targeted ad would analyse the spending patterns of the users and their mobile location, and even their friend's data to target groups for discounts when sales are slack.

Internet advertising has become a growing trend in the market. Most of the marketers use internet advertising as a way of economical promotion. At present, most targeted new media advertising use second order proxies for targeting, such as tracking online or mobile web activities of the consumer. Google gets billions of queries on a day-to-day basis. It aggregates all user queries daily and analyses it statistically to match users and advertisers. The giants like Google and Facebook utilize the user data acquired by them on a day-to-day basis and later with the help of this data, target the customers as per the market requirement.

GOOGLE ADWORDS

Google provides advertisers a platform that can help them to find prospective customers who are searching on Google. This is known as *Google Adwords*. *Google Adwords* is Google's advertising product and its main source of revenue. It provides approximately around 80% of its sales revenue.

Adwords provide site targeted advertising for banner, text and rich media ads. The *Adwords* program includes local, national and international distribution of advertisements. It is very useful for small manufacturers, because it can help the marketers to advertise their products to the target audience at a low budget.

With *Google Adwords* the advertisers will be able to reach their customers at the precise moment when they are looking for products or services. Finding what people want to see is a very challenging task for Google, but it also provides the Adwords advertiser with a competitive edge over other products/services.

Advertiser's work

E-Marketing allows the advertisers to choose who, where, when and how see their ads. Advertisers need to pick keywords in *Adwords* and the prices which they are willing to pay for their ads. This is an art; advertisers must guess what user queries are relevant to their ads. The Google Network allows advertisers to reach millions of qualified users by distributing ads on search results and relevant Web content on a wide variety of sites and products. Advertisers can choose to show their ads on Google's search network (e.g. google.com), content network (e.g. blogspot.com), or both. All ads are opted-in to appear on the Google Network unless advertisers edit their distribution preferences. Ads are positioned on the basis of their rank in an online real-time auction.

Examples of Adwords

Bata, Nike, Reebok, Adidas can all advertise on the keywords "shoes", "running", "sports", "hiking shoes". Hilton in London can advertise on the keyword "London hotel". The ads can appear both on the right hand side of the page and occasionally above the search results. These are positioned there by Google's online advertising product, called *Google Adwords*. Ads are triggered by the user query - this is called *keyword advertising*. Keywords are terms that advertisers choose that trigger their ads on Google. When a user of Google.com types a search query into the search field, as shown in figure 1, the keyword shown here is "*London hotel*". Appropriate ads are triggered along with the search results. This keyword has

triggered many ads such as *Hilton hotel, London NYC hotel*. The advertisements are placed on the basis of the advertiser amount bid and the quality of the ad.

Keyword advertising has many distinct benefits. It allows the advertisers to target their ads specifically to users who are looking for their product or service, and avoid users who are unlikely to be their customers. As a result, the advertisers can get a better *return on investment*, also known as ROI, for the advertising budget. Targeted advertising also helps Google users find what they're looking for, ensuring that they will continue using Google and providing a valuable source of potential customers for the advertisers.

Figure 1. Ads for MBA College in Google Search

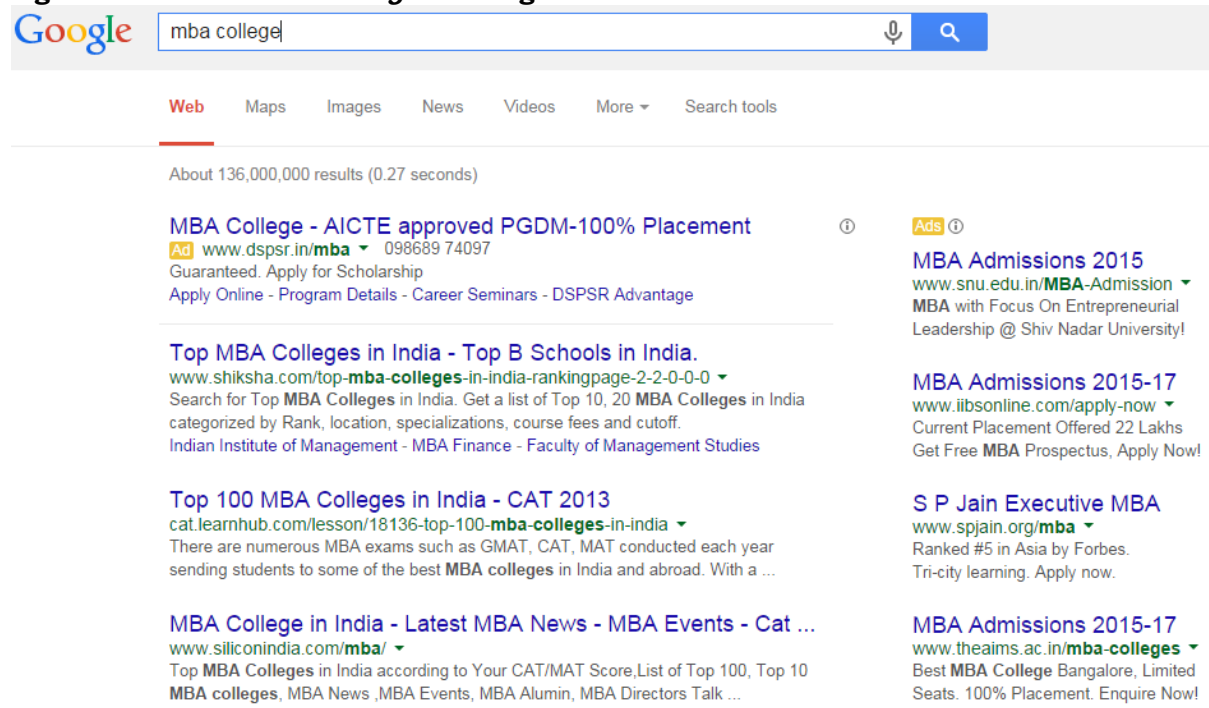
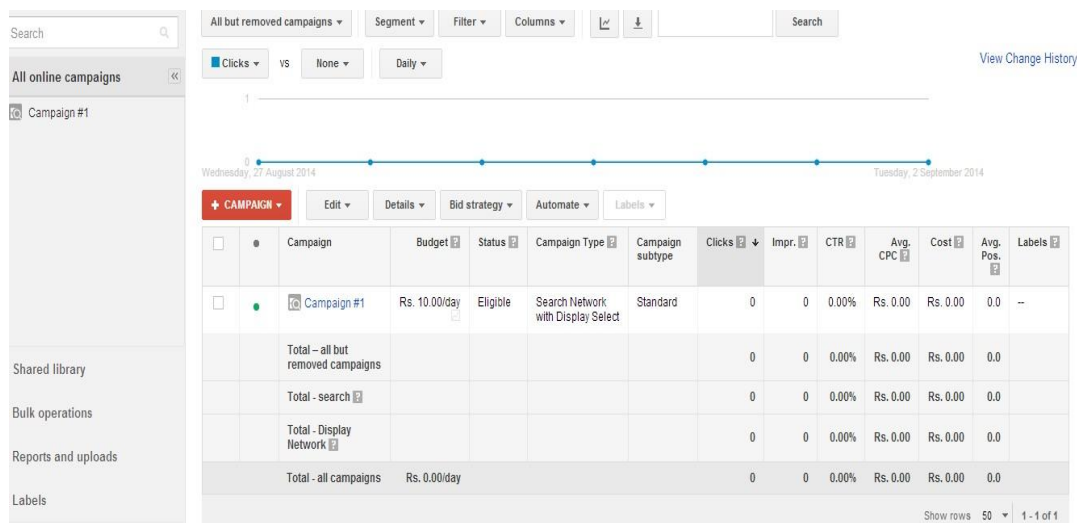


Diagram of Adword interface is shown in Figure 2. The Adword interface allows the advertiser to easily handle their Adword account. The advertiser should start the campaign by selecting the keyword and also the website that has to be advertised. The total budget for the advertisement can be spread out over time. Adwords will show the ad as often as possible within the budget guidelines. Google charges the advertiser only when Google clicks on the ad. This is known as CPC or *Cost per click*. *Impression* is the technical term for the rendering of the ad on user's. Google does not charge for impression, it will only deduct money if the clicks on the ad. *Click through rate* (CTR) is calculated by dividing number of clicks by impressions

Figure 2. Adwords Interface for Advertisers to create their ads.



Advantages of AdWords

Advertiser can always display their advertisement on Google irrespective of the total budget. The advertiser needs to pay only if people click their ads. The advertiser can choose his/her own maximum daily spending limit. The advertiser can choose how much amount he/she is willing to spend per click for every keyword. *Google Adwords* are measurable, accountable and flexible. Insights around the clock can be obtained from Google.

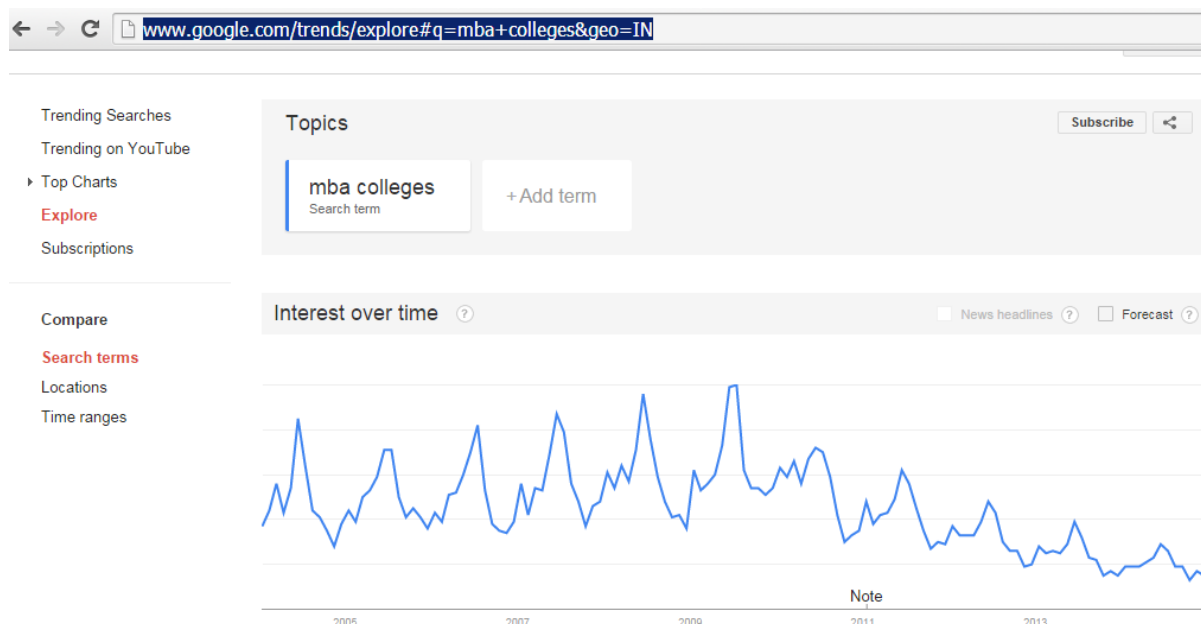
When selling AdWords, determining the customers' metrics in advance can help the advertiser to meet its objectives and ensure success. For example, if advertiser's marketing objectives are branding, building awareness or educating prospects.

Using AdWords advertisers can target the message to an audience that is likely to be receptive. If the advertiser's objective is generating leads, or identifying sales prospects, AdWords will track ads that lead to these results, for example, driving event registrations or newsletter sign-ups.

GOOGLE TRENDS

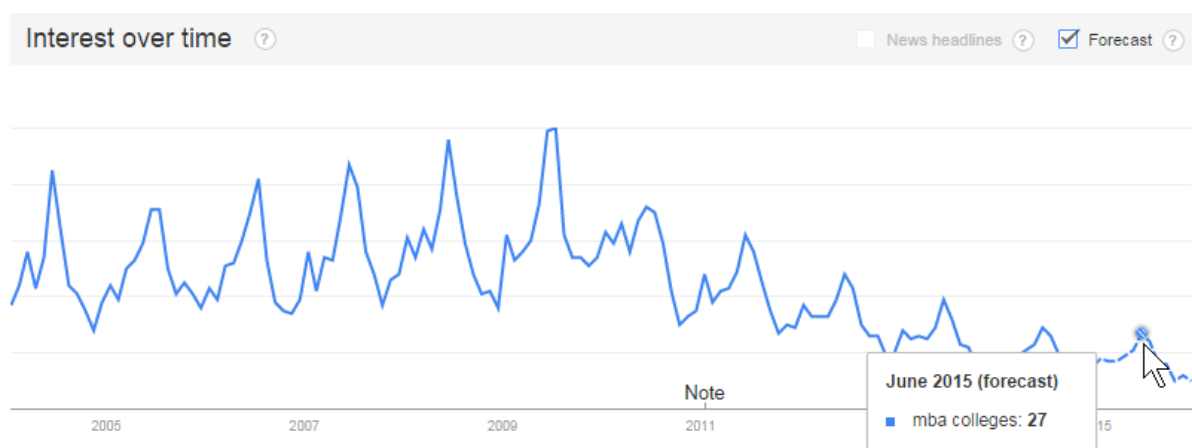
Google Trends is a public web facility of Google Incorporation. Based on the data available from Google Search, it shows how often a specific search-term is entered, compared to the total times that term is searched across various regions of the world, and in various languages. It is used to statistically analyse the queries of the user over time. Google Trends allows the user to compare the volume of searches between two or more terms. An additional feature of Google Trends is in its ability to show news related to the search-term overlaid on the chart, showing how new events affect search popularity. It helps people to know about the latest happenings in the country as well as in the world.

Figure 3. Google trends data for “MBA College”



Example: <http://www.google.com/trends/explore#q=MBA+Colleges>. Figure 3 shows the Google trends window. In this figure we have entered the words ‘MBA Colleges’. Therefore Google trends shows trend in which “MBA Colleges” was searched across various years from 2005 – 2014. (Figure 4)

Figure 4. Query volume for “MBA Colleges” from 2005 to 2014



Google trends location data

We notice that interest in *MBA colleges* peaked in 2009, and has been on decline since. However it has a seasonal component, which peaks every summer in June. If we turn the forecast checkbox, Google trends expects the query volume to fall from 29 to 27 (93% in 2015).

The figure below shows the relative percentage of the word “MBA Admissions” being searched around the world in decreasing order with India having the highest rank and UAE the lowest.

The data is normalized (is not absolute) with respect to the total query volume for each country. We see that in India has a high regional interest in “MBA Colleges”. If we had to target foreign students to enrol, we would look at Nepal.

Figure 5 Regional interest in “MBA Colleges” based on query location



The figure below the Google trends data across various India. Here Andhra Pradesh ranks the highest and Madhya Pradesh has a lower number of searches. In Town/City wise break up, we see Hyderabad leads the nation, followed by Delhi and Bangalore in MBA Admission queries.

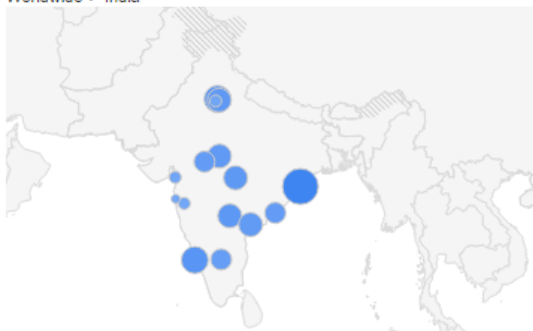
Figure 6. Regional Interest in “MBA Colleges” in India, based on query



The recruiting team could consider visiting Bhubneswar in Orissa. It is heartening to see Mangalore city is the 2nd most interested city in *MBA Colleges*. There is lot more geographical data in Google trends, which we leave to the reader to explore online: <http://www.google.com/trends/explore?q=mba+colleges&geo=IN>

Regional interest ?

Worldwide > India



Subregion | Town/City

Bhubaneswar	100	<div></div>
Mangalore	83	<div></div>
New Delhi	81	<div></div>
Hyderabad	80	<div></div>
Nagpur	79	<div></div>
New Okhla Industrial Development Ar...	78	<div></div>
Bhopal	77	<div></div>

Comparing MBA Specialisation Queries

We can even compare trends of different queries; we see that *MBA Finance* is more popular than *MBA Marketing* by a ratio of 41:22.

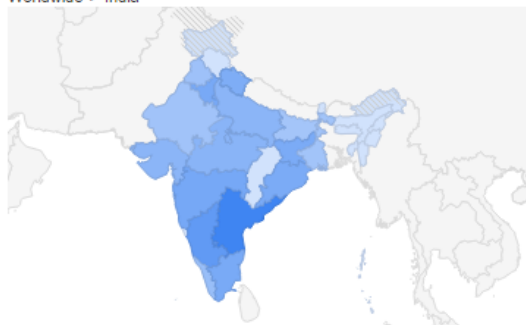


Comparing regional variation in specialisations:

Regional interest ?

mba finance mba marketing

Worldwide > India



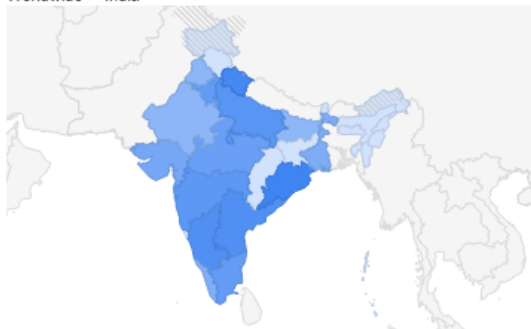
Subregion | Town/City

Andhra Pradesh	100	<div></div>
Karnataka	77	<div></div>
Maharashtra	62	<div></div>
Haryana	61	<div></div>
Tamil Nadu	60	<div></div>
Jharkhand	59	<div></div>
Odisha	59	<div></div>

Regional interest ?

mba finance mba marketing

Worldwide > India

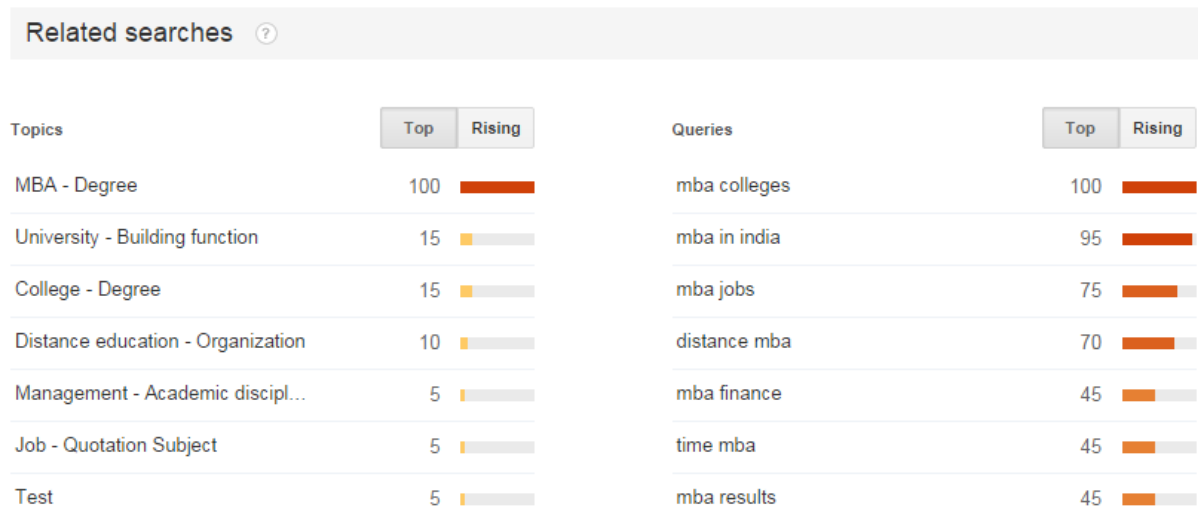


Subregion | Town/City

Odisha	100	<div></div>
Uttarakhand	96	<div></div>
Karnataka	88	<div></div>
Andhra Pradesh	88	<div></div>
Delhi	83	<div></div>
Uttar Pradesh	82	<div></div>
Maharashtra	80	<div></div>

In the figure below we can see that Google Trends is showing the related keywords that were typed by people searching “MBA” on Google. This shows that most people in India searched for *MBA colleges* followed by *MBA in India* (as opposed to going abroad). The right side of the figure shows the different queries through which people searched the word ‘MBA’.

Figure 7. Searches related to “MBA Colleges”



Advantages of Google Trends

Google Trends is free. Even though there are lot of advantages for individuals who use Google Trends, they need not pay any money to access the real time information which is updated on a regular basis. Google trends also have a fast method to compare the single and multiple search engine keywords and allow the user to compare the stories that are on recent searches for past days. A marketer can get a visual look at the number of visits to his sites and also the geographical location of the majority of the visitors. This helps them to find daily unique visitors, and track them over time. They can then take measures to improve their website traffic. It helps to analyse those patterns that can be helpful, especially for a start-up business.

Another advantage for the start-up firm is that it is free. With the Help of Google Trends a company can look into its competitor’s performance based on the number of queries to the rivals website. This helps them to know the comparison between both of them so as to enable them to get some ideas on how to improve. Google Trends helps to obtain real-time insight into market sentiment, and this information can be used strategically to devise profitable trading strategies for businesses.

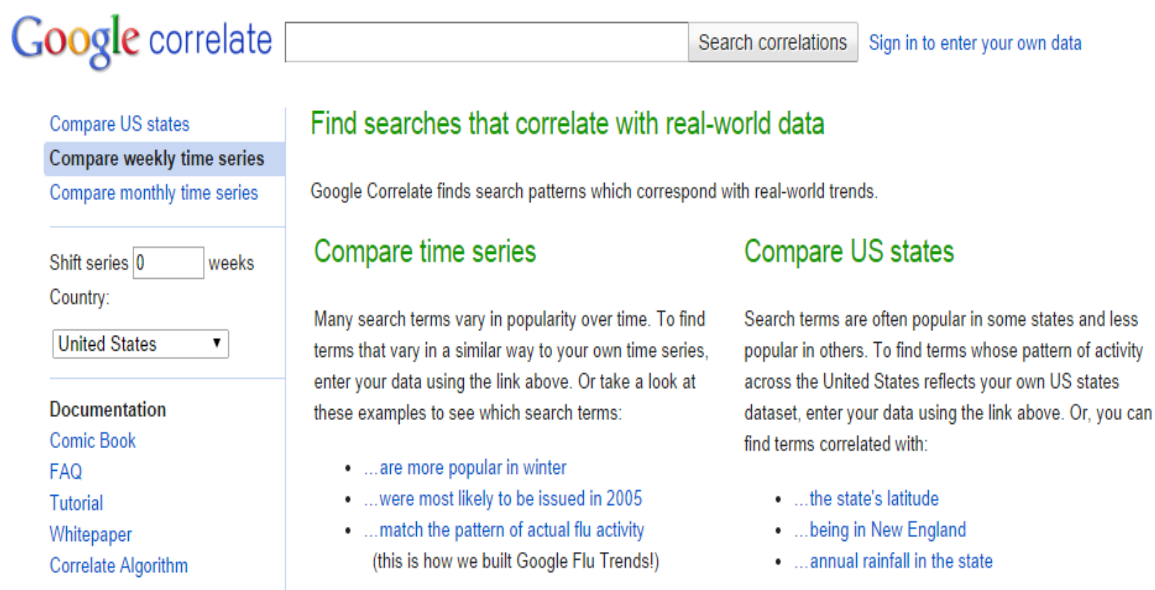
If investors carefully look into the data provide Google Trends more deeply they can get the daily or hourly trends which may help them to predict a rise or fall in stock prices.

It could also give people warning of impending crises, as in the case of flu trends in USA. Google Trends, with its access to search data from the most popular search engine - ‘Google’ is probably the most powerful such tool currently available.

GOOGLE CORRELATE

Google Correlate is a tool on Google Trends which enables the user to find queries with a similar pattern to a target data series. The target can either be a real-world trend or a query that one enters. The results can be viewed on the Google Correlate website or downloaded as a CSV file for further analysis. Google Correlate is like Google Trends in reverse. Google Correlate contains web search activity data from January 2003 to present. This data is updated weekly.

Figure 8 Google Correlate home screen



Google correlate [Sign in to enter your own data](#)

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

Shift series weeks

Country:

Documentation

[Comic Book](#)
[FAQ](#)
[Tutorial](#)
[Whitepaper](#)
[Correlate Algorithm](#)

Find searches that correlate with real-world data

Google Correlate finds search patterns which correspond with real-world trends.

Compare time series

Many search terms vary in popularity over time. To find terms that vary in a similar way to your own time series, enter your data using the link above. Or take a look at these examples to see which search terms:

- ...are more popular in winter
- ...were most likely to be issued in 2005
- ...match the pattern of actual flu activity (this is how we built Google Flu Trends!)

Compare US states

Search terms are often popular in some states and less popular in others. To find terms whose pattern of activity across the United States reflects your own US states dataset, enter your data using the link above. Or, you can find terms correlated with:

- ...the state's latitude
- ...being in New England
- ...annual rainfall in the state

Using Google Correlate, for example, “*What did users search 2 months before MBA in India?*” The answer from Google Correlate is *School in Bangalore,.., Financial Risk*. There will be many spurious matches like (*chipset driver*), that the user must filter. See Figure 9.

Figure 9 Finding queries correlated to “MBA admissions” two months earlier in India.

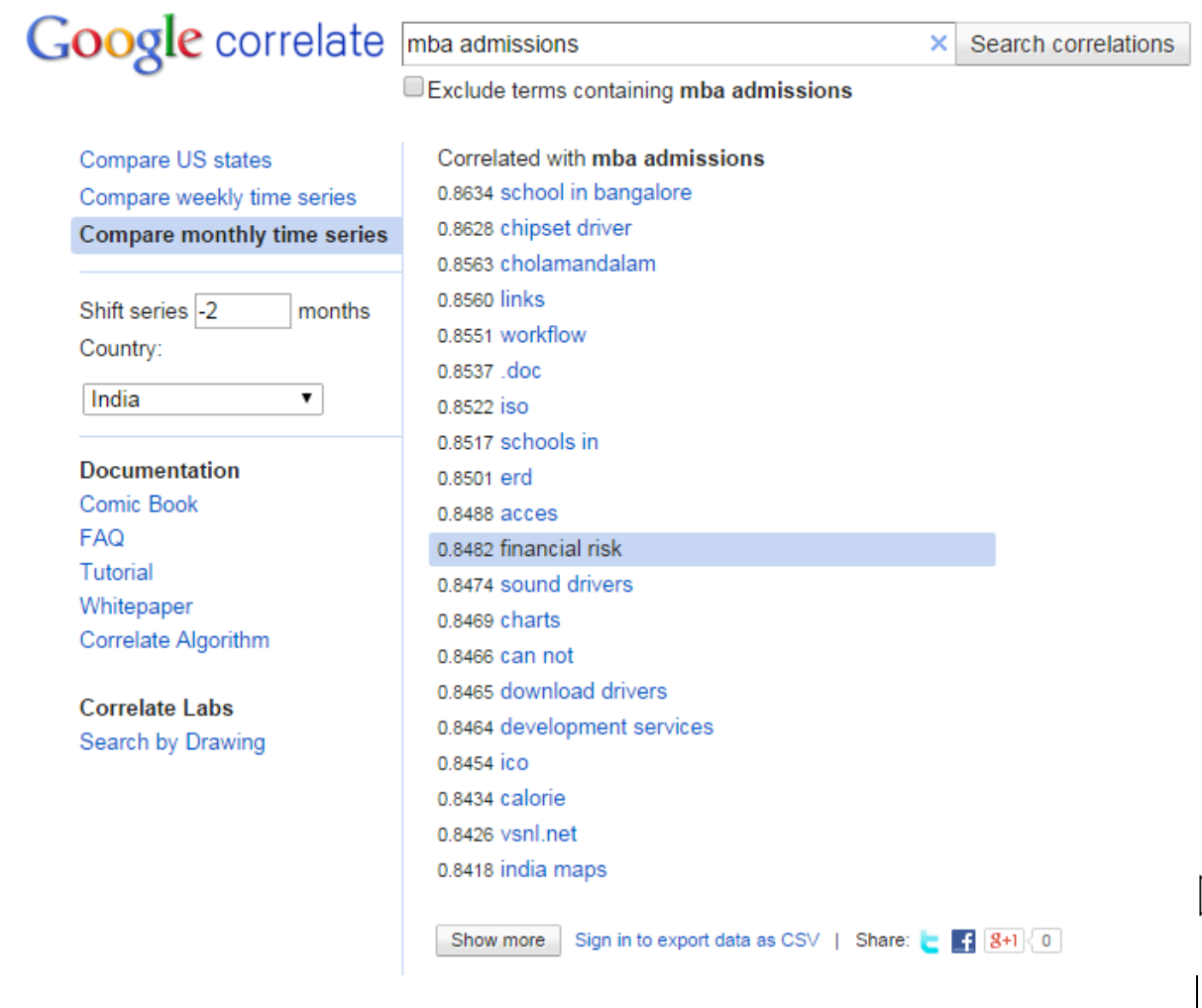


Figure 10. Correlation of query “MBA admissions” and “Financial Risk” two months earlier, using the URL

<http://www.google.com/trends/correlate/search?e=mba+admissions&e=financial+risk&t=monthly&p=in&shift=-2#default,30>

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

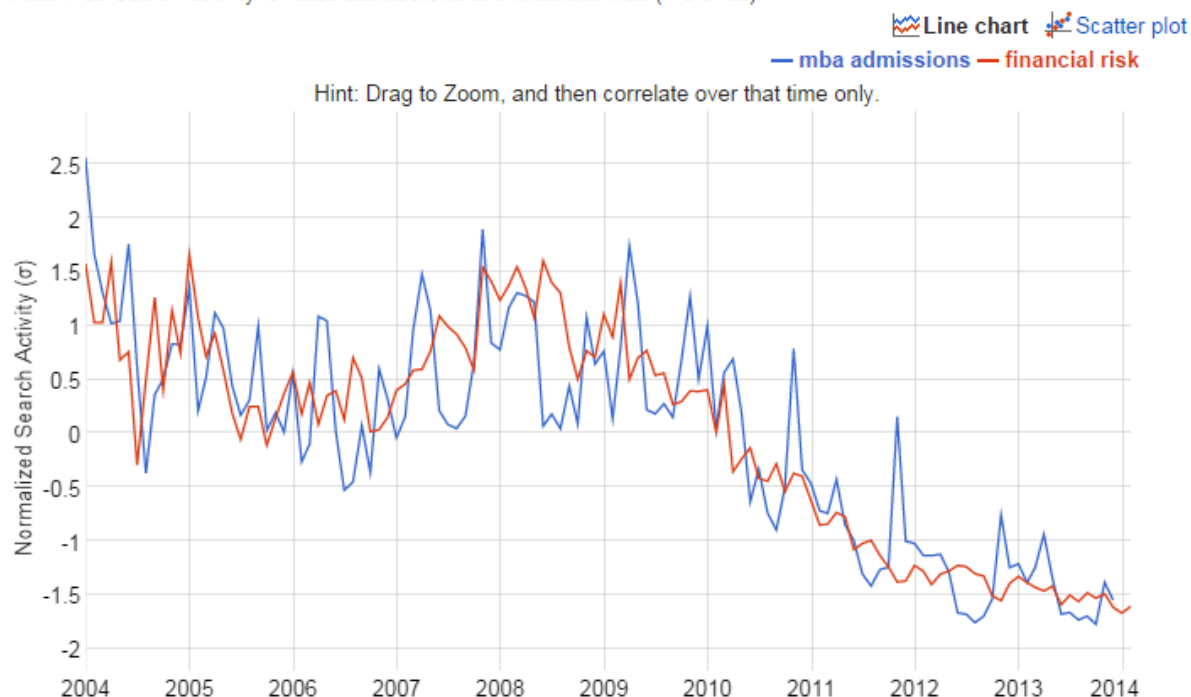
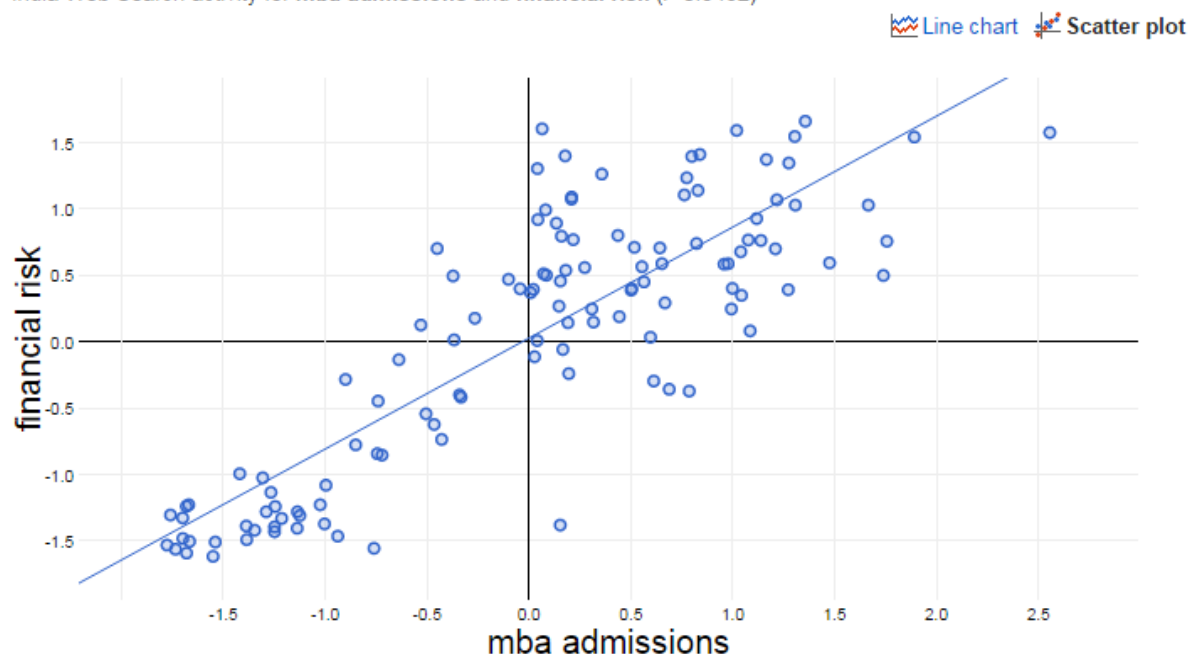


Figure 11. Scatter plot of correlated search words.

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



The graphs in Figure 10 and 11 show the correlation between *MBA Admissions* and *Financial Risks* through a line chart and a scatter diagram.

Advantages of Google Correlate

Results can be surfaced in real time. There is no need for any prior knowledge to make a good choice of queries that match the phenomenon. As this application is user friendly and simple to use we no training is required to use Google Correlate.

It is helpful for Search Engine Optimisation (SEO) keyword research. For example, there are many marketers who are advertising on Google. So when a person types the word “*MBA*”, Google only shows the most important advertisement. This is why each marketer strives hard to make their page attractive so that Google selects their website in organic search results. But Google selects the best matches by its page rank algorithm. The page rank algorithm looks at how many other websites link to the matching website, how many people clicked on the result earlier, the type of people, duration of browsing the sites, etc. This is called SEO and Google trend helps to optimise its efficiency.

Though Google returns only positive correlations, one can see the negative correlations also if they want to. For this, the user has to multiply their input data by negative one (-1) in their spreadsheet program before uploading it to Google Correlate.

Privacy: None of the information in Google Correlate can be linked to a specific person. The data does not include any information concerning the identity, specific location, or IP address of the user. Apart from this, Google would anonymize any original web search logs that are more than nine months old. Only if many people have typed a particular keyword on Google, then it is used by Google Correlate or else it is treated as private information and not shown to users. Google has to balance between private and public information.

Filtering: Google Correlate filters the Queries which may not be interesting. This includes rare or misspelled queries, *stop words* (stop words are common words like ‘the’), queries of a correlation value that is less than $r=0.6$, and pornographic queries.

PSYCHIC ADS

Psychic ads are those advertisements which will be shown to the customers, who might need the product in future, but have not thought of purchasing the product at present. The advertisements are shown prior to the requirement. This is in contrast to traditional marketing strategy discussed by *Wu and Lin (1997)*.

For example, in order to advertise *MBA College* two months early, the advertiser has to search for an event that has taken place preceding two months. In this case, one such query is *Financial Risk*. So to advertise two months before for *MBA admissions*, the College can target users searching for *Financial Risk*. The advertiser uses the technique of *recall*, as in *Marder, (1997)*. Even though the customers have not thought of buying *MBA* yet, the need for it is triggered prior to its requirement.

This is *psychic* – predicting what our users will want in the future. For psychic advertisements, the advertiser can collect the data pertaining to the user queries from Google trends. Later he/she can select the required keyword for the ad using Google AdWords. For instance, in case of *MBA* advertisement, the advertiser can opt for the keyword *Chalomandalam* (It is a financial company, as Google tells us) or the phrase *Financial Risk*. Even though the advertisement is for *MBA*, but the ad will be shown when

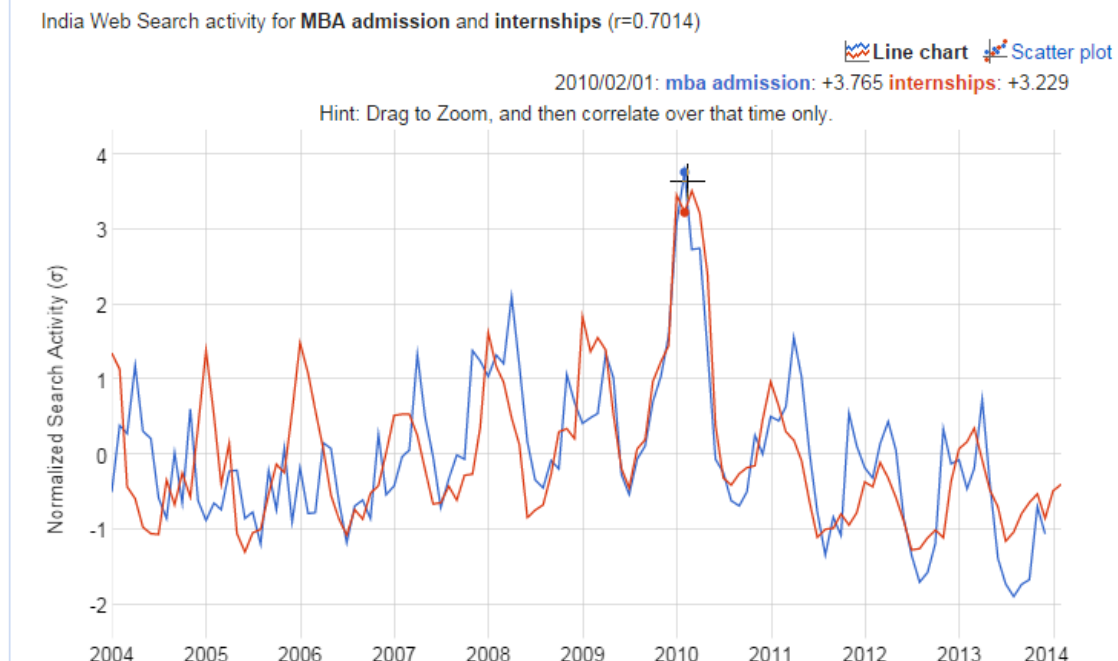
the users type “*Financial Risk*”, i.e. When a user types *Financial Risk* on Google, an MBA College advertisement will be shown.

There is one caveat, the advertiser has to use his/her own judgement to decide if the demographics of the two temporal separated search queries overlap, because Google Correlate discards user ids before generating its time-series. We tackle the problem of matching demographics across queries in a subsequent paper, by analysing mobile query data.

TIME SERIES COMPUTATION IN R

Here we show how to target MBA students 2 months before they search for “*MBA Admissions*”. We find a strong correlation between the searches for “*MBA admissions*” and “*Internship*” 2 months before that. So to target MBA students two months before admission begins, we can advertise “MBA” on the keyword “internship”.

Figure 12 India Web Search activity for MBA admission and internships ($r=0.7014$)



The “*MBA Admissions*” correlate data is available for download as CSV files from (*Google Correlate Dataset, 2014*) by navigating from (*Google Data Trends 2014*). We process these CSV files in R 3.1 (*The R Project for Statistical Computing, 2014*) inside R-Studio on Windows 7 and 8. We will not delve into the R language here, R and R-Studio was taught in a workshop by *Kavya, Nishita, Nikhita, Jovita, Meet and Laxmi (2014)*. Given the portability of R, we expect consistent results on Linux (*The R Project for Statistical Computing, 2014*). We use the *xts* (extended time series) and *TTR* (Technical Trading Rules) packages of R to analyse time (*Using R for Time Series Analysis, 2014*). Without further ado, we present our R code to analyse the time series.

```
# R program to analyse
# Google trends forecast data.
library(xts) # Extended Time Series.
library(TTR) # Technical Trading Rules.
```

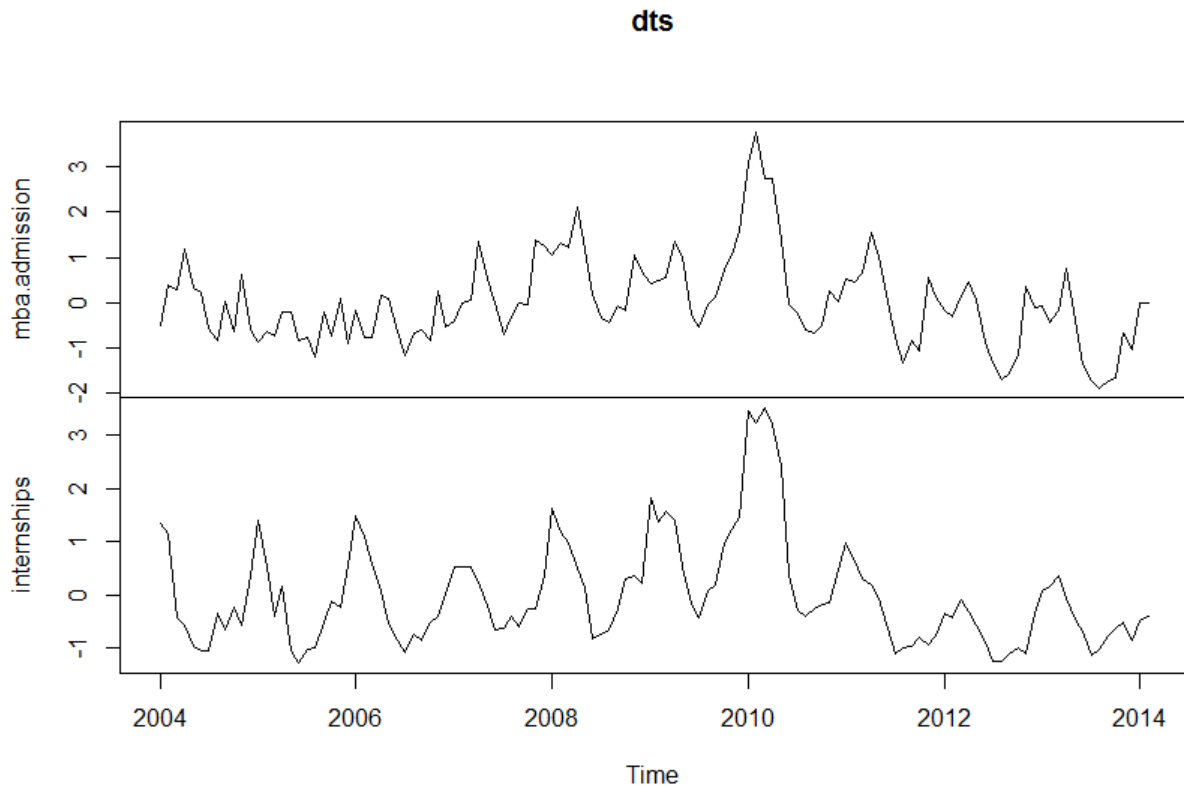
```
# Read the data downloaded from Google trends.
# Data correlates the query 'mba admissions'
# with queries -2 months before for India.
mbadata <- read.csv(file.choose(), header=T, skip=11, fill=T)

# 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)

# Pick two columns to analyse.
data1 <- subset(data, select = c(2:3) )

# Convert the date into a time series.
dts <- ts(data1, frequency=12, start=c(2004,1))
plot.ts(dts) # graphics of 2 time series. Figure 12.
```


Figure 13. Query Data for *MBA Admissions* and *Internship* from 2004 to 2014.

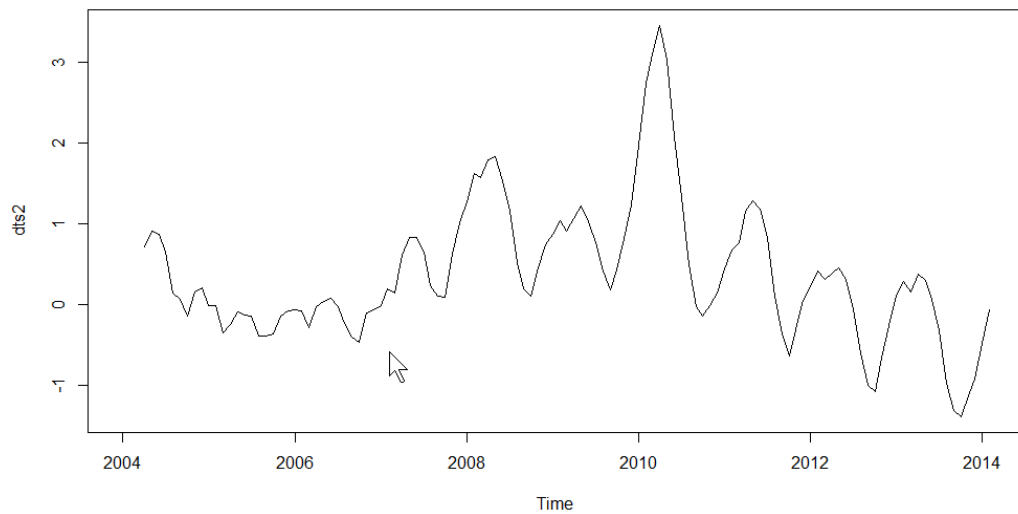


```
data2 <- xts(sapply(data1[,c(1:2)],as.numeric),order.by=dates)

# Correlate the columns 'mba admissions' and 'internships'.
cor(data2$mba.admission, data2$internships)
# R computes the Correlation as 0.7

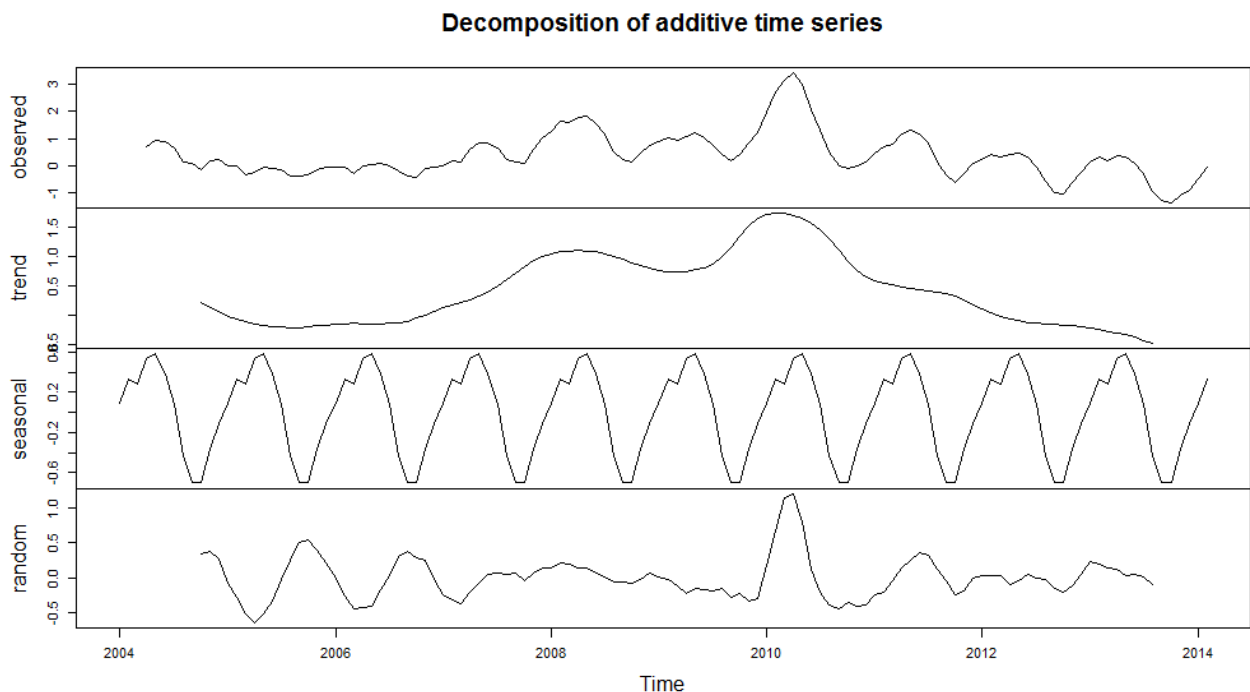
# Use the SMA package to smooth the series.
dts2 <- SMA(ts(data2,frequency=12,start=c(2004,1)),n=4)
plot.ts(dts2) # graphics of dts2, Figure 13.
```

Figure 14. Smoothed time series, with $n=4$.



```
# Decompose the time series into
# Seasonal and non-seasonal components.
ddts2 <- decompose(dts2)
plot(ddts2) # graphics with seasonal trend, Figure 14.
```

Figure 15. Time series decomposed into its components



We get additional insight into the query for “*MBA admissions*” trend over time. It has a seasonal peak in summer, when students are looking for colleges. The trend had a peak in 2010 when MBA was very popular.

CAVEATS AND EXTENSIONS

There are nuances and pitfalls in machine generated Adwords. The queries with temporal correlation may be from a disjoint set of users, as Google forecast discards the user id before the computation of the forecast data, as described by Vanderkam, Schonberger, Rowley, and Sanjiv Kumar (2011).

Advertiser creativity and common sense is still necessary to generate the final ad, as many spurious queries show up in the correlation. We do not have Adwords Analytics data to know whether spurious matches are effective for cheaper Adwords.

For example, Google trends data does not tell us that the users searching for “*MBA Admissions*” maybe different from the users searching for “*Engineering Admissions*”. We rely on the judgement of the Adwords marketers to pick the right keywords in such cases. However with access to private query data it is possible to compute such disambiguation automatically. Mobile query data is much more accurate, because it comes from a single authenticated user with a known location. In the mobile query stream there are more signals that can be correlated, the user’s location history, contact graph, and demographics can be accurately inferred, increasing the effectiveness of the Ad, however that is out of the scope of this paper.

The extension would work as follows, given access to a mobile query stream; we would first cryptographically hash the ids for protecting the user privacy in our database. The stop-words will be filtered out. Then the database would be indexed by words and ids. For each word pair (w_1, w_2) of interest to the advertiser, we would compute S_1 the set of ids querying w_1 , and S_2 similarly. If $(|S_1| \text{ intersection } |S_2|) / (|S_1| + |S_2|) < 0.5$, we will reject the correlation as spurious.

In our example, we can target “MBA Admissions” to queries of “CA admissions”, if at least 50% of the mobile users who have queried for one-term have also queried for the other term.

CONCLUSION

Targeted advertising is much more effective than traditional medium of advertising, with much higher CTR (click through rates) for advertisers. This will keep both the user and advertiser happy.

We take this a step further in *Psychic Ads* by predicting how to target users for future marketing, giving the advertiser a big advantage over their competitors in terms of early reach and cheaper ads on Adwords.

Future work

We will plan to launch Google AdWords Ad campaigns targeting students and compare the CTR improvement.

Acknowledgements

We thank the JKSHIM, NITTE faculty for valuable guidance. In particular we thank our director Prof. Sankaran for his constant encouragement and detailed comments on the paper. Prof. Krishna Prasad for help with the referencing style, Prof Saravan for organizing the conference, Prof. Krishna Kumar for piquing our interest in R, Prof. Sheetal Kumar for organizing the *R for MBA* workshop as a Techbugs activity, and our colleagues Nikhita, Jovita, Meet, Laxmi in the R workshop.

References

- Google Trends Data*, Retrieved Oct 1, 2014 from Google website, <http://google.com/trends>
- Vanderkam, Schonberger, Rowley, and Sanjiv Kumar, “*Nearest Neighbor Search in Google Correlate*”, Tech Report 41694, Google Inc, Retrieved Oct 1, 2014 from <http://research.google.com/pubs/pub41694.html>
- Using R for Time Series Analysis*, retrieved Oct 1, 2014 from <http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html>
- Kavya M N, Nishita Rai, Jovita M, Nikitha F, Laxmi N, Meet A, “*R for MBA*”, Techbugs workshop, JKSHIM, NITTE, India 8/2014.
- Marder, Eric, *The Laws of Choice: Predicting Customer Behavior*,. New York: The Free Press, A Division of Simon and Schuster, 1997.
- Wu, Sheng; Lin, Cathy S; Lin, Jean. *An empirical investigation of online users' keyword ads search behaviours*. Online Information Review 35.2(2011): 177-193.
- Hyunyoung Choi, Hal Varian, “*Predicting the Present with Google Trends*”, 2011. Retrieved from Google Research blog, <http://googleresearch.blogspot.in/2009/04/predicting-present-with-google-trends.html>
- Google Correlate Dataset, for MBA admissions in India*. Retrieved Oct 1, 2014 from Google website, <http://www.google.com/trends/correlate/search?e=MBA+admission&e=internships&t=monthly&p=in&shift=-2#default,20>
- “*The R Project for Statistical Computing*”, Retrieved from <http://www.r-project.org/>