Data Analytics & Tools

Internal Audit Refresher Course

at

WIRC of ICAI

Day 2 – Technical Session 5 on 27-Aug-16

A special thanks to all the members in Internal Audit background who continuously contribute towards process and efficiency improvement of business and contribute towards nation building. A special salute to all the family members of members in Internal Audit background who sacrifice their personal and family life and thus contribute towards nation building.

Did you know?

1. India never lost a test match, when Dada scored a century.
2. India never lost an ODI when Shikhar Dhawan scored a century (before 20-Jan-16).
3. India never lost against Pakistan in Cricket World Cup

These statements are nothing but inferences made from data analytics.

Industry : Retail (Shopping Malls / Grocery Supermarkets)

Have you ever advised your clients regarding which product to be placed near a product in a shelf in a retail outlet?

Industry : Telecom

Have you ever advised your clients regarding which data plan to be offered to which users?

Industry : Airlines

Have you ever advised your clients regarding the optimum path between two airports?

Industry : Media

Have you ever advised your clients regarding the selection of topic / cast / director / timing of broad cast of a television show?

Industry: Banking
Vertical : Credit Cards

Have you ever advised your clients regarding offering credit cards with particular attractions to a particular category of clients?

Data Analytics – Meaning in simple terms

- https://www.youtube.com/watch?v=SSRBDO0J5UI
  (Disclaimer: this is a useful video as well as a promotion of a course by an academy. There is no intention to promote any organization by way of this presentation. The only intent is to spread knowledge)

- Breaking the problem into pieces

- Dive in the data (population, not just sample)

- To make inferences / to have insights (regarding correlation, association, and other relationships of various variables / behaviors)

Data Analytics – Few more examples

- Archimedes and King's gold crown

https://www.youtube.com/watch?v=jp25RyoLcvY&list=PLliISwqa3YzSDZHnqtc116WWPgbsnimZz

- Gold medal winning strategies of Olympics coaches (rowing boat race)
- https://myactivity.google.com/

Benefits of DA in Internal Audit

- Population instead of samples
- Insights of business / industry
- Predictive instead of corrective
- Required as risk, complexities, competition, and sophistication has increased.

Classification of DA

On the basis of industry	On the basis of functions / Domain Analytics	On the basis of Insights offered
• Cr. card, Insurance, Retail, FMCG, Telecom, etc.	• HR analytics, Finance analytics, sales analytics, Supply chain analytics, Risk analytics, etc. (virtually anything)	• Descriptive analytics (retail grocery example – why behind the buy – info about past) • Predictive analytics (Product Affinity Analysis / Association Analysis/ Market basket Analysis) • Prescriptive Analytics (Aviation example)

Few important methods of analysis

- Hypothesis testing (various distributions)
- Linear Regression (trend line – predicting Y based on X)
- Logistic Regression – (Y or N)
- Cluster Analysis – (Amazon)
- Cohort Analysis (Comparing Groups)

Data

- Structured Data
- Semi Structured Data (at least some organization – like tag, etc.)
- Unstructured Data (FB photos, etc.)

Big Data – IBM definition

- Volume
- Volatile
- Variety

DA – Tools and Techniques

- Traditional
  - Excel
  - SQL
  - Access

DA – Tools and Techniques

- Sophisticated
  - SAS
  - WPS
  - IBM Packages
  - R
  - Apache Hadoop
  - Tableau
  - PowerPivot Excel AddIn
  - Microsoft HDInsight
  - ACL
  - Idea

SAS and WPS

- History of SAS institute Inc (journey from University for analysis in Agriculture to SAS Institute)
- More than 200 packages
- Certifications
- SAS Enterprise Miner
  - GUI based yet powerful
  - Costly – all companies cannot afford
- Base SAS package (including other tools)
  - Code based interface
  - Cheaper than SAS E-Miner
- WPS
  - Similar to Base SAS
  - Much cheaper
  - Had to fight a copyright battle for code with SAS Institute

IBM Packages

- Have spent huge money to acquire other companies
- Two popular packages are SPSS and SPSS Modeler
- SPSS
  - Equivalent of Base SAS
  - More popular in market research (data is not huge) whereas Base SAS is more popular in business analytics (huge data and complex computations)
- SPSS Modeler
  - Can be compared with GUI based SAS Enterprise Miner
  - Regression, clustering, decision tree, etc.

R

- Most popular open source analytics tool
- Evolved from a language – S (4th Gen PL)
- S was later bought by a company and S+ was developed
- S+ stands for S + GUI
- Although parent language (S and S+ were commercial), R evolved as an open source analytic tool
- Around 3000 packages (collection of functions) from few 100s in 2001- any one can produce a package in R
- R is a concise and elegant language – but not easy to learn – but once mastered, super complex analysis can be done
- Excellent visualization and statistical capabilities
- Problem with large data – however with integration with Hadoop it is now possible to manage big data by using R

Apache Hadoop and Big Data

1. Meaning of Big Data – IBM definition – Volume, Volatile, Variety
2. Hadoop
   1. Open source & Apache License
   2. Not a software – but a set of tools
   3. Distributed computation

Many more…

- Tableau (easy to use, drag and drop)
- PowerPivot Excel AddIn
- Microsoft HDInsight
- ACL
- Idea

Access and SQL

1. Used for structured data
2. RDBMS
3. Tables are linked with each other
4. Specific analytical queries can be made in user friendly language

Excel Lessons

1. Data exploration, data preparation, and data dictionary
2. VLOOKUP() and VLOOKUP() on left
3. Simple Pivots and GETPIVOTDATA()
4. Using logic in Excel and data exploration combined with SUMIFS() and COUNTIFS()
5. Advanced Filters
6. OFFSET nested with XIRR() to calculate IRR of several clients with several cash flows
7. Understanding the usefulness of INDEX, and MATCH functions
8. Understanding the usefulness of Macros

Sources and Useful links (no intent to advertise any academy):

- http://www.indiatimes.com/sports/did-you-know-india-never-lost-a-test-match-where-dada-scored-a-century-the-story-of-sourav-ganguly-s-16-test-hundreds-257047.html
- http://www.cricbuzz.com/cricket-news/70324/icc-cricket-world-cup-2015-statistical-round-up-shikhar-dhawans-feat-india-cricket-team-makes-a-statement-and-so-do-the-associate-cricket-teams
- http://www.forbes.com/sites/bernardmarr/2016/08/09/how-big-data-and-analytics-help-athletes-win-olympic-gold-in-rio-2016/#7ad78ab04205
- https://community.tableau.com/docs/DOC-1236
- https://www.youtube.com/watch?v=jp25RyoLcvY&list=PLliISwqa3YzSDZHnqtc116WWPgbsnimZz
- https://www.youtube.com/watch?v=V7st1Rgl74w
- https://www.youtube.com/watch?v=Pq3OyQO-l3E
- https://www.youtube.com/watch?v=HwAuiA39jGA

Thanks for giving your precious time to this session.