Data Science & Analytics

Hiren Deliwala & Dr. Jian ZHANG

Course Objectives

- The objective of this course is to introduce students to data analytics and its impact on business results
- Presentations, examples, assignments and in-class group exercises will be used to communicate key concepts

Topics We'll Cover

Databases	Data Modeling	SQL
Data Analytics Using Excel	Data Warehouse	Business Intelligence
Reporting & Querying	OLAP and Multidimensional Analysis	Statistical Analysis

What you want to learn…

Course Timeline

Week	Topic
Week 1	Overview of course; Introduction to Data Analysis and Science
Week 2	Database Fundamentals
Week 3, 4	Data modeling
Week 5	SQL Basics 1
Week 6	SQL Basics 2
Week 7	Introduction to Data Warehousing
Week 8	Midterm
Week 9 & 10	TBD
Week 11	Introduction to Business Intelligence
Week 12	Introduction to OLAP and Multidimensional Analysis
Week 13,14	Basic data & statistical analysis
Week 15	Final

Typical Class Format

- Lecture
- Occasional Class Exercise
- Sharing of Student Exercises

Class Etiquette

- My first time but I hope to share my passion for data analytics with you by doing our best to make the class interesting and enjoyable. I ask that you …
  - Attend all classes.
  - Participate and share your experience
  - Not use laptops in class except to take notes or as part of a class exercise
  - Turn off all cell phones and other electronic devices during class

Assignments & Grading

2 Group Assignments	30%
Homework (2-3 assignments)	20%
Mid Term	20%
Final	20%
Class Participation/In-Class Exercise	10%

Example Group Assignments

1) project to retrieve data from a few tables in database for analysis

2) project to use that data for analysis (produce charts and tables to show their interpretation about the data)

Books and Reference

1. Data Modeling Made Simple: A Practical Guide for Business and IT Professionals, 2nd Edition by Steve Hoberman, Michael Blaha, Bill Inmon and Graeme Simsion
2. The Data Warehouse Lifecycle Toolkit by Ralph Kimball, Margy Ross, Warren Thornthwaite
3. Data Science for Business: What you need to know about data mining and data-analytic thinking by Provost, Foster and Fawcett, Tom (Jul 27, 2013)
4. Healthcare Data Analytics

Introductions

- Name, etc…
- My first car was …
- I am in this class because …

Agenda

- Data Analytics in Action
- Introduction to Data Analytics Concepts

Data Analytics in Action

- Which color Binders sell the most?
- What is the sales trend for Pencils?
- Who is our best sales rep for January?

Data Analytics in Action

- Which state has the highest number of complaints?
- Which Issue gets reported the most in Florida?
- How many complaints do we receive via Web versus Phone?
- Which complaints get closed with monetary relief?

Data Analytics in Action

- Which age group votes the highest?
- Which precinct has highest Democratic voters
- Do 18-20 group vote?

“There’s a world of difference between truth and facts. Facts can obscure the truth.”

- Maya Angelou

“Essentially, all models are wrong, but some are useful

- George E. P. Box

This is Analytics

- WalMart finding out what sells in a hurricane
- Netflix finding out what movies a customer might want to watch
- An investor finding out anomalies exist in the stock market in order to make a profit to his/her customers
- Amazon personalizing and customizing websites
- Sprint finding out that a customer might want to drop its service before the customer even knows it
- Finding the best route for a packet in a network

It helps Answers Key Questions

It helps Answers Key Questions

- What movies (or books) customers would like to watch (or read)?
- What movies to order from studio and how many?
- **What is the number one reason for the success of baseball player?**
- **Why should you always defer to the 2nd half to get the ball in college football?**
- When is there a flu epidemic in region in the country?
- What is the one item you want to have in your store in case of a hurricane?
- Which customers are most likely not to have an accident?
- When a customer is likely to jump ship & go to a competitor?
- Who are our best customers?
- When should we tell a customer to quit gambling?
- What is the best criteria that predicts success when hiring a new Ph.D. student to become a faculty member?
- What is the one thing that will improve a lawyer's chance to win a case?
- What are some questions one can answer with a loyalty card?

Strategic factors for the use of Analytics

- More difficult to find and sustain competitive advantages (geographical barriers gone, product differentiation reduced, etc.)
- Becomes increasingly more important to execute on strategy and become operationally excellent particularly in serving customers
- Many more business are now data-driven (virtual companies)
- Speed of change and risk in marketplace
- Evidence of success by other companies (Monkey see .. Monkey do)

Data Analytics

- Data Analytics is a process of systematically applying statistical and logical techniques to describe, summarize and compare data.

- Analytics is the process of collecting and analyzing data in order to make better business decisions, develop better products and serve the customers better.

Analytics is:

- Providing Right Information at the Right Time to enable managers to make informed Business Decisions
- It fact-based rather than gut based decision making

Types of Data Analytics?

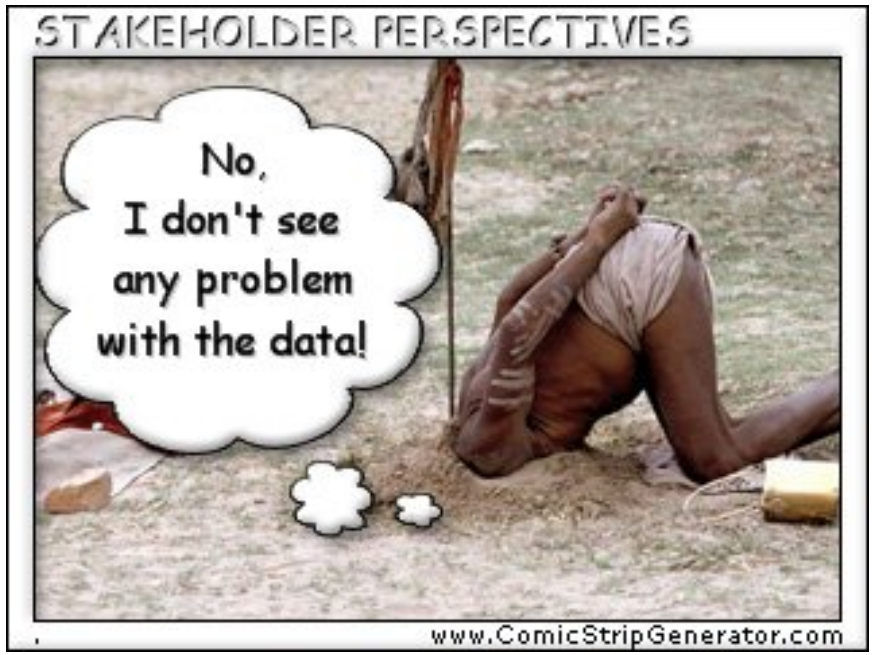
- Quantitative
  - Data – expressed as numbers
  - Analysis – numerical methods to ascertain size, magnitude, amount
- Qualitative
  - Data – difficult to measure sensibly as numbers, e.g. count number of words to measure dissatisfaction
  - Analysis – expresses the nature of elements and is represented as themes, patterns, stories

Another way to look at Data Analytics?

- Exploratory – looks for patterns, differences, matches and anomalies in data
  - Airline ex. “Are certain geographic locations more prone to lost luggage?”
- Confirmatory – confirm or reject whether expected patterns in data exist
  - Airline ex. – “Are duty free sales on international flights increasing or decreasing on a per passenger basis?”

Before your can do Data Analytics…

- Must access client data – connect to it and query out what you need
- Transfer – ODBC, Internet, Disk
- Store – Where? Security? Access?
- "Massage" – get data into useable form



What software is used in Data Analytics?

- Excel & Access
- Scripting Languages – write your own programs in VB, Perl
- Business Intelligence Tools
- Visualization Tools
- Statistical Tools such as SPSS and SAS