IS5126
Hands-On With
Business Analytics
(HowBA):
Beyond Finance
Lecture 1 - Introduction to BA
January 12, 2015

Admin
• Pick up syllabus and schedule
• Purchase HBS Case from http://hbsp.harvard.edu
  o Data.gov: FP-410-015
• Sign up team of 4 on IVLE by Jan. 30
  o Use IVLE forums to find team mates

Course Staff
• Lecturer: Dr. Tuan Q. PHAN
  o Email: disptq@nus.edu.sg
  o Phone: 6601-1054
  o Office location: Com2 #04-08
  o Office hours: Mondays, 5:30-6:30pm, or by appointment

• TA: Ms. Tianhui Tan
  o Email: tianhui.tan@nus.edu.sg

What is This Course About?
• Expose students to practical Business Analytics, common tools, and techniques
• “Data Science”
  o Scientific method for interpretation for business applications
• Target students:
  o Technically strong and business experience
  o Various industries, eg. Financial services, travel industry, retail, marketing, consulting services
• Present Final Project at StePS
  o VCs, industry leaders, professors, colleagues, and friends!

Learning Objectives
• Introduce analytic methods and techniques
• Practice in entire data pipeline
• “How” to do things using Python, SQL, R
• Learning-by-doing
• Bridge “hard” and “soft” disciplines
• Think creatively

Learning Approach
• Instructor leads and guides learning and ideas
• Learning from peers in class and through group projects
• Learning to learn and use references
• Technical tools as a means to a goal
• What is the goal?
  o Creative and out-of-the-box thinking backed by technical proficiency
• Discussion-based
• Class is time-consuming!
Topics

• Introduction to BA, math refresher
• Data, preparation, and webmining
• Databases & SQL
• Data exploration & Visualization
• Marketing Models
• Clustering & Segmentation
• Predictive Modeling: decision tree, cross sectional regressions, panel data
• What makes a good prediction?
• Variable transformation & reduction
• Social Network Analysis

Introduction to BA

• Learning Objectives:
  o What is BA?
  o Where and how is BA used?
  o BA in Organizations
  o BA Roles
  o BA Process and tools
  o BA as Multidisciplinary
  o BA Objectives
  o Math basics and interpretation of data
  o What can go wrong?

The train problem

<table>
<thead>
<tr>
<th>Eastbound trains</th>
<th>Westbound trains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Describe which trains are east/westbound?
• Attributes of a train:
  ✓ if a short car is rectangular then it is also double sided
  ✓ a short closed rectangular car can have either a flat or peaked roof

Answers:
✓ if a train has a short closed car, then it is eastbound, otherwise westbound
✓ if a short car is rectangular, or has a car with a jagged roof, then it is westbound, otherwise eastbound.
✓ and many others...

What is Business Analytics?

• Using data to support business decision making process
  o Data Mining, Business Intelligence, Pattern Recognition, Machine Learning, etc. ...
  o Actionable
• Data is ubiquitous because of rapid progress in IT
  o Data warehousing
  o Data production: Consumer data
• Data is BIG – GB? TB? PB?
• Data and analytics as a product?
• Is data valuable?
BA in an Organization

**DATA FLOW**

Real World → Collection → Raw Data → Import → Data Warehouse → Analyze → Report

**TRADITIONAL ORGANIZATIONAL ROLES**

- Ops
- Developers & Engineers
- DBA
- ?
- Managers
- Business Analyst

BA Process & Tools

**DATA FLOW**

Real World → Collection → Raw Data → Import → Data Warehouse → Analyze → Report

- Web crawling
- Server logs
- Applications
- Mobile
- Embedded Devices
- Server-side tools
- ETL
- Statistical tools
- R
- Data
- SAS
- SPSS

BA as Multidisciplinary

**COMPUTER SCIENCE**

- Software development
- Algorithms
- Databases
- Economics
- Business
- Psychology
- Sociology

**STATISTICS**

- Software development
- Regression

Break
Moneyball

- Show videos

BA Objectives

- Fact-based decision making
- Optimization vs. Prediction vs. Causation
- Types of analytics:
  - Unsupervised learning: use data to uncover underlying structure.
  - Supervised learning: use inputs to predict training set or outcomes (dependent variable)
  - Aggregation is sometimes necessary
- What are we analyzing?
- Probability of drawing two "bins" assuming equal probability for each item:
- Draw a letter from each "bin" 
- What are we analyzing?

Unit of Analysis

- What are we analyzing?
- Aggregation is sometimes necessary

Probability and Sets

- What is the probability of drawing two "bins"?
- Probability and interpretations

The Fallacy of the Masses

- Hospital A has a death rate of 50%, Hospital B has a death rate of 20%. If you are very sick, which would you choose?
- “Unit of Analysis” is critical to conclusions and insights => rows of data
- Let set $A = \{ \alpha, \beta, \gamma \}, |A| = 3$
- if set $B = \{ \beta, \gamma, \delta, \chi \}, |B| = 4$
- Intersection $A \cap B = \{ \beta, \gamma \} = C$
- Union $A \cup B = \{ \alpha, \beta, \gamma, \delta, \chi \} = D$
- Subtraction $A \setminus B = \{ \alpha \}, B \setminus A = \{ \delta, \chi \}$

Probabilities and Interpretations

- Objective is to find true "insight" or prove "hypothesis"
- Skeptic’s view: Can data tell us the "truth"?
- Maybe our data is only set B, but truth is set A?
- Bayes’ interpretation:
What can go Wrong?

- Isn’t everything in the data?
- Data quality
  - Selection bias
  - Operational bias
  - Methodological bias
- Modeling errors: our perception of the world
  - Technical errors
  - Missing variables
  - Insufficient data
- Type 1, 2 error

Data Quality

- Selection Bias: results are driven by how individuals or groups choose a treatment
  - Example: Using Facebook data, you find that 50% of Singaporeans have university degrees
  - Solution: collect data which supports and does not support your results
- Operational Bias: results are driven by how you collect the data
  - Example: webcrawling the top 100 bestsellers on Amazon reveals that most book sellers have low margins.
  - Solution: use better methods to collect data
- Methodological Bias: results are driven by how you transform the data
  - Example 1: using only the first 100 credit card customers for a bank over 10 years, you find that they are rich and older
  - Solution: Use better sampling methodology
  - Example 2: all else equal, sales in February are lower than in January.
  - Solution: pick appropriate aggregation level

Modeling Errors

- Technical errors: predictors are correlated. More on this later
  - Example: Using Recency, Frequency, Monetary to predict response rate to an advertising campaign.
  - Solution: various, perform cross-correlation analysis

- Missing variables: conclusions cannot be concluded from the available data
  - Example: Education increases your productivity, hence a higher salary.
  - Solution: try to collect better or 3rd party data, triangulate missing variables if possible, otherwise interpretations are limited

- Reverse Causality: more on this later

- Insufficient data: a conclusion cannot be drawn based on the available number of observations
  - Big but sparse data!

Logical Errors

Type 1, 2 errors

<table>
<thead>
<tr>
<th>Observed</th>
<th>False</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td>☹</td>
<td>☺</td>
</tr>
</tbody>
</table>

- Type 1: false positive
  - Example: Study falsely shows new medicine is effective.
- Type 2: false negative
  - Example: Study falsely concludes that new drug is ineffective.

Course Grading & Deliverables

- Two guided mini projects (2x20%): 40%
- Final Project: 50%
- One case write ups: 5%
- Class participation: 5%

Deliverable Details

- Form teams up to 4 for the semester
- Should be a balance of students with technical and business backgrounds
- Week 6, Feb. 16: Project Pitch
- Week 7, Mar. 2: 2 page project proposal
- Week 12, Apr 6: Project presentations
- Reading week, Apr 20: Term Showcase (STePS)
- Apr 24: Project write-ups due
- Choose 1 of 4 cases to write-up about analytic opportunities
- 2 page write up
  - Present 3-5 minutes in class
- Class participation
  - Write name on index card and contribution
STePS 2014

Find My Doctor

Alicament:
Food network and medicine

Whisky Recommendation

Who do you fly with next?

Projects to Avoid
- Stock prediction or finance related
- Demand and media sales
- Real-estate/property price forecasting
- Any time series with a single metric over time
Misc. Issues
• IVLE
• No late submissions
• Plagiarism is not tolerated
• Hand-phones – switch off/vibrate mode

Introduction to Case Studies

General Guidelines
• Trains you to think clearly in real-world ambiguous and confusing situations.
  • Who is the protagonist?
  • Who are other players with a vested interest?
  • What are the incentives of the players?
  • What is the current environment?
  • History, market structure, laws, regulations, etc…
  • What are the decisions to be made?
  • What is YOUR recommendation?
  • Support with facts and data from the case
  • What would be the competitor’s reaction?
  • How sustainable is your recommendation?
• More tips on IVLE

Is it Art or Technical?
• Art, “softer”: business, psychology, sociology, economics
  • Technical, “harder”: Math, Programming, Logic, Computer systems
  • BA: must be technical to make it an art! (and add value!)

* Pick up syllabus and schedule
* Purchase HBS Case from http://hbsp.harvard.edu
* Data.gov, #9-610-075
* Sign up team of 4 on IVLE by Jan. 30
  • Use IVLE forums to find team mates