Business Intelligence and Process Modelling

F.W. Takes

Universiteit Leiden

Lecture 2: Business Intelligence & Visual Analytics
Business Intelligence
Recap

- **Business Intelligence**: anything that aims at providing actionable information that can be used to support business decision making
  - **Business Analysis**
  - Business Analytics
    - **Visual Analytics**
    - Predictive Analytics (next weeks)
- Data → Information → Knowledge (next week)
- Process Modelling (April and May)
Business Intelligence goals

- Operational intelligence
- Corporate governance
- Risk assessment
- Compliance
- Auditing
  - Sarbanes-Oxley (SOX) — role of IT in corporate governance
Management Approaches in BI

- **Continuous Process Improvement (CPI):** ongoing effort to improve products, services or processes
  - Incremental improvements vs. Breakthrough improvements
  - Evaluate based on efficiency, effectiveness and flexibility

- **Total Quality Management (TQM):** improve processes up to the microscopic level, focusing on meeting customer demands and realizing strategic company goals, e.g., **Six Sigma**
Six Sigma

- Originally developed by Motorola in the early 1980s
- Minimize Defective Parts per Million Opportunities (DPMO)
- Mean $\mu$ and standard deviation $\sigma$

<table>
<thead>
<tr>
<th>Quality level</th>
<th>DPMO</th>
<th>Percentage passed</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Sigma</td>
<td>690,000</td>
<td>31%</td>
</tr>
<tr>
<td>Two Sigma</td>
<td>308,000</td>
<td>69.2%</td>
</tr>
<tr>
<td>Three Sigma</td>
<td>66,800</td>
<td>93.32%</td>
</tr>
<tr>
<td>Four Sigma</td>
<td>6,210</td>
<td>99.379%</td>
</tr>
<tr>
<td>Five Sigma</td>
<td>230</td>
<td>99.977%</td>
</tr>
<tr>
<td>Six Sigma</td>
<td>3.4</td>
<td>99.9997%</td>
</tr>
</tbody>
</table>
Normal distribution
DMAIC approach

- **Define** the problem and set targets,
- **Measure** key performance indicators (KPI’s) and collect data,
- **Analyze** the data to investigate and verify cause-and-effect relationships,
- **Improve** the current process based on this analysis,
- **Control** the process to minimize deviations from the target.
Key Performance Indicators

- **KPI**: measure, variable or metric to analyze the performance of (part of) an organization
- Strategic goals → Measurable variables
- **SMART**
  - **Specific**
  - **Measurable**
  - **Acceptable**
  - **Realistic**
  - **Time-sensitive**
KPI examples

- Operational: increasing market share by 10%
- Financial: increase profit by 10%
- Sales: obtain 10 new customers
- Human resources: attract 10 new sales officers that are part of the world’s top 1% in the field
- Customer support: forward no more than 10% of the support calls to second line
BI in practice

- Codeless reporting
- Instant querying
- Rich visualization
- Dashboards ("Management cockpits")
- Scorecards
Balanced Scorecards

- Goal: align business activities to the vision and strategy of the organization
- Financial and nonfinancial goals
- Monitor a relatively small number of summative indicators
Balanced Scorecard

Financial
“To succeed financially, how should we appear to our shareholders?”

Customer
“To achieve our vision, how should we appear to our customers?”

Vision and Strategy

Internal Business Processes
“To satisfy our shareholders and customers, what business processes must we excel at?”

Learning and Growth
“To achieve our vision, how will we sustain our ability to change and improve?”
Balanced Scorecard

- Four perspectives:
  - Financial
  - Customer
  - Processes
  - Learning and Growth

- Four elements per perspective
  - Objectives
  - Measures
  - Targets
  - Initiatives
## Balanced Scorecard Perspectives

**Financial**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Measures</th>
<th>Targets</th>
<th>Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Customer**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Measures</th>
<th>Targets</th>
<th>Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Internal Business Process**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Measures</th>
<th>Targets</th>
<th>Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Learning and Growth**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Measures</th>
<th>Targets</th>
<th>Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Vision and Strategy**

**To succeed financially, how should we appear to our shareholders?**

**To achieve our vision, how should we appear to our customers?**

**To satisfy our shareholders and customers, what business processes must we excel at?**

**To achieve our vision, how will we sustain our ability to change and improve?**
Balanced Scorecard Processes

**Balanced Scorecard**

**Translating the Vision**
- Clarifying the vision
- Gaining consensus

**Communicating and Linking**
- Communicating and educating
- Setting goals
- Linking rewards to performance measures

**Feedback and Learning**
- Articulating the shared vision
- Supplying strategic feedback
- Facilitating strategy review and learning

**Business Planning**
- Setting targets
- Aligning strategic initiatives
- Allocating resources
- Establishing milestones
Relations between Perspectives

Financial
- accounts receivable
- operating expense
- return on capital employed

Customer
- customer satisfaction

Internal Business Process
- rework

Learning and Growth
- employees’ morale
- employees’ suggestions

(+) or (-) arrows represent the direction of the relationship. 
Some terms . . .

- **Business Activity Monitoring (BAM)**: insight in operational status and events of a business
- **Complex Event Processing (CEP)**: monitor events and react immediately if a pattern occurs
- **Corporate Performance Management (CPM)**: measuring the (financial) performance of a process or organization
Some systems . . .

- Enterprise Resource Planning (ERP) Systems
- Enterprise Information Systems (EIS)
- Business Information Systems (BIS)
- Management Information System (MIS)
- Executive Information System (EIS)
ETL

- **Extract** data from source systems: generate dumps, exports, etc.
- **Transform** data: aggregating, linking, sorting, joining, etc.
- **Loading** data into target system into desired (reporting) format
OnLine Analytical Processing (OLAP)

Given a data table with $n$ attributes:
- Dimensions of an $(n - 1)$-dimensional cube represent $n - 1$ attributes of the data
- Value in a cell of the cube represents the remaining attribute

Use a slice or dice to get the desired information

Suitable for, e.g., star schema data
OLAP Example

Example: website visitor logs, storing:

1. Time
2. Web page
3. Action
4. Conversion
OLAP Cube Example

Time

Homepage
Catalogue page 1
Product page 1
Product page 2
Web page

Page view
Add to basket
Like
Signup
Action

http://snowplowanalytics.com
OLAP Cube Example Slice

2010 slice

Homepage
Catalogue page 1
Product page 1
Product page 2

Page view  Add to basket  Like  Signup

http://snowplowanalytics.com
OLAP Cube Dice

Product action dice

2011
2010
Product page 1
Product page 2
Product page 3
View
Add to basket

http://snowplowanalytics.com
OLAP Formalized

- **OnLine Analytical Processing (OLAP)**
- Given a data table $D$ with 4 attributes $W, X, Y$ and $Z$
- An OLAP cube can be characterized as a function $f : (X, Y, Z) \rightarrow W$
- An example of a slice is a function $g : (Y, Z) \rightarrow W$
- Given subsets $X' \subseteq X$ and $Z' \subseteq Z$ a dice is a function $h : (X', Y, Z') \rightarrow W$
Schemas

**Star Schema**
- Redundancy results in maintenance difficulty
- Dimension table is not connected to other dimension table
- Queries are easy to write and interpret
- Faster execution time due to simplicity

**Snowflake Schema**
- Relational aspect makes it easy to maintain
- Dimension table can be connected to other dimension table
- Queries are more complex and involve JOINs
- Slower execution time due to relational aspect
Star schema

http://prashanthobiee.blogspot.nl/2012/12/star-schema-and-snowflake-schema.html
Snowflake schema

http://prashanthobiee.blogspot.nl/2012/12/star-schema-and-snowflake-schema.html
Break?
Visual Analytics
What is Visualization?

- **Intuition**: data is more than its raw bits and bytes
- **Visualization**: making something visible to the eye (Oxford dictionary)
- **All visualizations share a common “DNA” — a set of mappings between data properties and visual attributes such as position, size, shape, and color — and customized species of visualization might always be constructed by varying these encodings.**


- **Visual Analytics**: knowledge discovery (DIKW) based on visualization
What is Visualization?

- **Data properties**: attributes of (groups of) data objects
  - Name; Age; City

Visual attributes: e.g., position, size, shape, label, color, etc.

Visualization: mapping data properties to visual attributes

- Name → Label
- Age → Size
- City → Position

"Frank", log \(_2\) (28), (52.1603216, 4.4939262)
What is Visualization?

- **Data properties**: attributes of (groups of) data objects
  Name; Age; City
  *Frank; 28; "Niels Bohrweg 1, Leiden"*

- **Visual attributes**: e.g., position, size, shape, label, color, etc.

- **Visualization**: mapping data properties to visual attributes
  
  *Name* → *Label*
  *Age* → *Size*
  *City* → *Position*
What is Visualization?

- **Data properties**: attributes of (groups of) data objects
  - Name; Age; City
  - *Frank; 28; "Niels Bohrweg 1, Leiden"

- **Visual attributes**: e.g., position, size, shape, label, color, etc.
  - Label; Size; Position

- **Visualization**: mapping data properties to visual attributes
What is Visualization?

- **Data properties**: attributes of (groups of) data objects
  - Name; Age; City
  - *Frank; 28; ”Niels Bohrweg 1, Leiden”*

- **Visual attributes**: e.g., position, size, shape, label, color, etc.
  - Label; Size; Position

- **Visualization**: mapping data properties to visual attributes
  - Name $\rightarrow$ Label
  - Age $\rightarrow$ Size
  - City $\rightarrow$ Position
What is Visualization?

- **Data properties**: attributes of (groups of) data objects
  - Name; Age; City
  - *Frank; 28; "Niels Bohrweg 1, Leiden"

- **Visual attributes**: e.g., position, size, shape, label, color, etc.
  - Label; Size; Position

- **Visualization**: mapping data properties to visual attributes
  - Name $\rightarrow$ Label
  - Age $\rightarrow$ Size
  - City $\rightarrow$ Position
  - "Frank", $\log_2(28)$, (52.1603216, 4.4939262)
What is Visualization?
**Why visualization?**

Which of the following are the most important business benefits that your organization seeks to gain from deploying data visualization and visual analysis technologies? (Please select all that apply.)

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved operational efficiency</td>
<td>77%</td>
</tr>
<tr>
<td>Faster response to business change</td>
<td>62%</td>
</tr>
<tr>
<td>Ability to identify new business opportunities</td>
<td>59%</td>
</tr>
<tr>
<td>Higher employee and partner productivity</td>
<td>47%</td>
</tr>
<tr>
<td>Increased return on data assets</td>
<td>44%</td>
</tr>
<tr>
<td>Financial accountability and transparency</td>
<td>39%</td>
</tr>
<tr>
<td>Take advantage of technology change (e.g., mobile)</td>
<td>37%</td>
</tr>
<tr>
<td>Better regulatory compliance and governance</td>
<td>27%</td>
</tr>
</tbody>
</table>

SAS, Data Visualization: Making Big Data Approachable and Valuable, 2014
Why visualization?

**Top Benefits of Data Visualization Tools**

- Improved decision-making: 77%
- Better ad-hoc data analysis: 43%
- Improved collaboration/information sharing: 41%
- Provide self-service capabilities to users: 36%
- Increased ROI: 34%
- Time savings: 20%
- Reduced burden on IT: 15%

SAS, Data Visualization: Making Big Data Approachable and Valuable, 2014
Visualization theory

- Discrete vs. continuous data
- Categorical vs. quantitative data
- Mean or median?
- Variance?
- Correlations? Regression?
- Normal distribution or power law?
- The correct visualization depends on the data itself!
Listen to the data to . . .

- Catch mistakes
- See patterns
- Find violations of statistical assumptions
- Generate hypotheses
- Do outlier detection
Visualization Quality

- When is a certain visualization “good”?
Visualization Quality

- When is a certain visualization “good”?
- “Proper mapping of data properties to visual attributes”?
Visualization Quality

- When is a certain visualization “good”?
- “Proper mapping of data properties to visual attributes”?
- The number of data properties (variables) that is visualized?
Visualization Quality

- When is a certain visualization “good”?
- “Proper mapping of data properties to visual attributes”?
- The number of data properties (variables) that is visualized?
- The number of visual attributes that is utilized?
Visualization Quality

- When is a certain visualization “good”?
- “Proper mapping of data properties to visual attributes”?
- The number of data properties (variables) that is visualized?
- The number of visual attributes that is utilized?
- Aesthetics?
- ...
When is a certain visualization “good”? 
“Proper mapping of data properties to visual attributes”? 
The number of data properties (variables) that is visualized? 
The number of visual attributes that is utilized? 
Aesthetics? 
... 
Hard to answer objectively!
Infographic of infographics

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

**Design**

*Chart Style*
Percentage of infographics with the following charts:

- Pie Chart: 22%
- Pictorial Chart: 24%
- Line Chart: 24%
- Bar Chart: 32%

*Font*
- Sans Serif: 85%
- Serif: 15%

**Content**

*Countries Featured*

- United States: 88%
- China: 22%
- United Kingdom: 12%
- Australia: 12%
- Canada: 10%
- India: 10%
- France: 10%
- Mexico: 8%

*Theme*
Relative popularity of different infographic themes:

- Business
- Technology
- Health
- Culture
- Politics
- Environment

**Key Info**
Percentage of infographics with keys:
- 33%

Average number of symbols per key: 9.1

*Base Color*

- Pie Chart: 13%
- Bar Chart: 29%

*Navigational Iconography*
Frequency of axes & connecting lines in infographics:

- Labeled axes: 13%
- Labeled lines: 18%
- Both: 29%

**Sections**

Average number of sources per infographic: 2.29

**Credits Sources**

Average number of words per infographic title: 4.36

“RICHEST AND POOREST AMERICAN NEIGH...”
Visualization Metaphors

- Important is Big
- Happy is Up
- More is Up
- Categories Are Containers
- Organization is Physical Structure
- Similarity is Closeness
- Control is Up
Visualization Metaphors

- Important is Big
- Happy is Up
- More is Up
- Categories Are Containers
- Organization is Physical Structure
- Similarity is Closeness
- Control is Up

Visualization Metaphors

- **Important is Big**
- Happy is Up
- More is Up
- Categories Are Containers
- Organization is Physical Structure
- **Similarity is Closeness**
- Control is Up

Visualization Metaphors

- Important is Big
- Happy is Up
- More is Up
- Categories Are Containers
- Organization is Physical Structure
- Similarity is Closeness
- Control is Up

Visualization Metaphors

- Important is Big
- Happy is Up
- More is Up
- Categories Are Containers
- Organization is Physical Structure
- Similarity is Closeness
- Control is Up

Examples

- Time-series data
- Statistical data
- Geographical data
- Hierarchical data
- Network data

Index chart

Time-Series Data: Figure 1a. Index chart of selected technology stocks, 2000–2010.

Stacked graph

Time-Series Data: Figure 1b. Stacked graph of unemployed U.S. workers by industry, 2000–2010.

Small multiples

Time-Series Data: Figure 1c. Small multiples of unemployed U.S. workers, normalized by industry, 2000–2010.

Horizon graphs

Time-Series Data: Figure 1d. Horizon graphs of U.S. unemployment rate, 2000–2010.

Examples

- Time-series data
- **Statistical data**
- Geographical data
- Hierarchical data
- Network data
Scatter plot matrix

Statistical Distributions: Figure 2c. Scatter plot matrix of automobile data.

Parallel coordinates

Statistical Distributions: Figure 2d. Parallel coordinates of automobile data.

Examples

- Time-series data
- Statistical data
- **Geographical data**
- Hierarchical data
- Network data
Flow maps

Maps: Figure 3a. Flow map of Napoleon’s March on Moscow, based on the work of Charles Minard.

http://holstanford.edu/heer/files/zo0/ex/maps/napoleon.html
Choropleth maps

Maps: Figure 3b. Choropleth map of obesity in the U.S., 2008.

Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/choropleth.html
Graduated symbol maps

Maps: Figure 3c. Graduated symbol map of obesity in the U.S., 2008.

Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/symbol.html
Maps: Figure 3d. Dorling cartogram of obesity in the U.S., 2008.

Source: National Center for Chronic Disease Prevention and Health Promotion; http://hci.stanford.edu/jheer/files/zoo/ex/maps/cartogram.html
Examples

- Time-series data
- Statistical data
- Geographical data
- Hierarchical data
- Network data
Node-link diagram

Hierarchies: Figure 4a. Radial node-link diagram of the Flare package hierarchy.

Circular dendogram
Examples

- Time-series data
- Statistical data
- Geographical data
- Hierarchical data
- Network data
Force-directed layout

Networks: Figure 5a. Force-directed layout of Les Misérables character co-occurrences.

Arc diagram

Networks: Figure 5b. Arc diagram of *Les Misérables* character co-occurrences.

Annotated matrix

Networks: Figure 5c. Matrix view of Les Misérables character co-occurrences.
Dashboards

- Multiple **widgets** on one page
- A widget can contain:
  - OLAP slice
  - KPI metric
  - Data mining results
  - ...
- Codeless reporting
- BI in the blink of an eye!
Dashboard (4)

Executive KPI

Revenue 2012

Net Profit Margin | 2012 vs 2011

Debt-to-Equity


Return on Equity | 2010-2012

http://www.klipfolio.com/
The Music Industry Dashboard

POPULARITY OF MUSIC GENRES PER AGE GROUP IN THE US

<table>
<thead>
<tr>
<th>Music Genre</th>
<th>Age Group</th>
<th>Followers</th>
</tr>
</thead>
<tbody>
<tr>
<td>50’s - Dance</td>
<td>13-15 Female</td>
<td>22K</td>
</tr>
<tr>
<td>&amp; Electronic</td>
<td>13-15 Male</td>
<td>173</td>
</tr>
<tr>
<td>50’s - Pop</td>
<td>16-24 Female</td>
<td>22K</td>
</tr>
<tr>
<td></td>
<td>16-24 Male</td>
<td>173</td>
</tr>
<tr>
<td>50’s - Rock</td>
<td>25-34 Female</td>
<td>22K</td>
</tr>
<tr>
<td></td>
<td>25-34 Male</td>
<td>173</td>
</tr>
<tr>
<td>60’s - Dance</td>
<td>35-44 Female</td>
<td>22K</td>
</tr>
<tr>
<td>&amp; Electronic</td>
<td>35-44 Male</td>
<td>173</td>
</tr>
<tr>
<td>60’s - Pop</td>
<td>45-54 Female</td>
<td>22K</td>
</tr>
<tr>
<td></td>
<td>45-54 Male</td>
<td>173</td>
</tr>
<tr>
<td>60’s - Rock</td>
<td>55-64 Female</td>
<td>22K</td>
</tr>
<tr>
<td></td>
<td>55-64 Male</td>
<td>173</td>
</tr>
<tr>
<td>70’s - Dance</td>
<td>65+ Female</td>
<td>22K</td>
</tr>
<tr>
<td>&amp; Electronic</td>
<td>65+ Male</td>
<td>173</td>
</tr>
</tbody>
</table>

TOP 20 TWITTER FOLLOWERS & TWEETS PER ARTIST

TWITTER FOLLOWERS & ILLEGAL DOWNLOADS PER ARTIST

http://insideanalysis.com/
Assignment 1

- Gaming industry context
- Sales log spanning 4 years of sales
- Apply and compare BI techniques
- Inspect, visualize, aggregate, segment, score . . .
- Deliverables:
  1. Web-based BI Dashboard
  2. Short assignment report, preferably in \LaTeX
Lab session February 12

- Read and understand Assignment 1
- Mount the webserver on your workstation
- Obtain the data from the shared folder
- Inspect the data
- Setup a framework for your dashboard
- Load some data into your framework
- Investigate visualization options