Fundamentals of Data

Data warehouse

- storage of summarized performance data to represent what is happening about the organization

Data cube

- Smallest unit of structure in a warehouse
- 2 elements:
    - Dimensions (and dimension value): minimum requirement to have 3 dimensions
    - Measures: what you monitor

Note: Performance data usually lasts for 5 years. Because every 5 years, strategic plan changes so different performance measures happen.

Types of Data

Type	Direction	Magnitude	Example
Nominal	NO	NO	Name, department, address, gender
Ordinal	YES	NO	Rank, year level
Scale	YES	YES	Temp, age, height

Note: All statistical treatments can be applied to all data, but some treatments are meaningless.

Statistics

- Process
    - Transformation of data
    - Frequency
    - Involves pre-computation
- Computation

Descriptive Statistics

1. Frequency and distribution
    a. Relative frequency (part of one to the whole)
    b. Cumulative frequency (cumulated from all values)
    c. Absolute frequency (exact value)
2. Central tendency
    a. Mean, median, mode
    b. Percentile
3. Variability
    a. Range (min/max)
    b. Variance (how far are the values from each other)
    c. Standard deviation (how far are the values from the mean)
4. Relationship of variables
    a. Correlation
    b. Linear regression
5. Forecast
    a. Moving average
    b. Weighted moving average
6. Shape
    a. Skewness
    b. Kurtosis
7. Total/Count (not something you can do for all data)

Measures in Data Warehouse

- 1st Degree measures (Based on balanced scorecard): ex. Total sales
- Referential Data (refers to 1st degree): ex. Sales target
- Historical measures - other measures you select (statistical treatment)

*Calculated data in excel
*additive: sum added together to form a bigger sum
*non-additive: min values added together to form a bigger min value

- Historical Measures
    - Pre
    - Post (what you compute later)
- Priority is to pre-compute
    - Post is heavy on processing
    - Post is heavily computed every time

Fundamental Form of Data where Descriptive Statistics are preformed

- Frequency
    - Frequency on a defined dimension
    - Frequency of a variable
    - Pivot table: pivot chart to visualize (excel)
- Anything outside the standard deviation range, they are outliers (not normal, extreme)
    - It depends on the population
- Linear Regression
    - Computed to see the behavior of data
    - Trend, going up or going down?

Frequency

- Bin: number of elements the data will be distributed into
- Bin Size: range of the bin
- Bin size = (Max-Min)/Number of bins

- Relative Frequency Answers the question __% of the time, what is my sales? (Values of sales at particular %)
- Cumulative Frequency can show how fast/slow sales is increasing (Based on given trend, it shows how fast/slow the variable grows)
- Cumulative Relative Frequency: the value that sales reaches ___% of the time

---

Fundamental Concepts of Business Intelligence

Performance Management and Business Intelligence:

Key points in BI

1. System of technologies and processes to support organization decision-makers in generating insights about the organization
    a. What is happening / what happened
        i. KPI
        ii. Measure
        iii. Metric
    b. What is going right or wrong / what went right or wrong
        i. Standard
2. Knowing what happened, what is going right or wrong requires
    a. Establishment of a system of measuring pre-defined set of measures and metrics to determine success (performance)
3. BI is situated under a continuous improvement agenda that involves performing a cycle of activities
    a. Gather -> measure -> monitor -> make decisions

***BI Insights:**

1. Data and info
2. Visual (statistical treatment) (ex. charts)
3. Dimensional navigation (ex. filtering)

BI Proposition

- "Decision-making" and "business insights" in a more specific context
- Continuous Improvement Loop
- Gather Data
- Measurement of pre-determined metrics and measures of success
- Monitor successes and failures
- Make evidence-based decisions to improve success
- Bi implementation is primarily driven by the need and requirement of the org to implement a continuous improvement loop
- One business app using data for analytics, no need to clean otherwise app is faulty due to "dirty" data
- Multiple business app using data for analytics, ok to clean but necessarily not "dirty"
- Non-business app giving data for analytics, need to "clean" because data is processed by humans.
- Data cleaning part of data science but not BI (BI uses "clean" data)
- One business app, no need for sampling but need data aggregation (to make data smaller)
- Sampling happens when you use external data

Pre-BI systems that exist related to decision-making:

- Decision Support System
    - Use of mathematical models that will allow "what-if" analysis of managers
    - Mathematical models represent behaviors of an element of the business
    - Tech limitations restricted actual implementation of DSS to department-level or subject-level DSS
- Executive IS
- Management IS

Elements of BUSINTL

1. Technology
2. Decisions
3. Data Dimension Navigation
4. Information Visualization
5. Statistics

* ETL and Data Warehousing = Prerequisite knowledge or knowledge to self-review

BUSINTL common elements in all varied definitions:

1. System of technology and processes to support decision-making
2. System of technology and processes to allow the human-led creation of timely, accurate, high-value, and actionable business insights
    - Clarity
    - Discovery
    - Not available to old-fashion decision making

Old fashion decision-making

- Urgency and estimation vs. accuracy and completeness
- Tidbits of information

- Grapevine information (“chismis”)
- Recollection of the past
- Advice from selected individuals in the company
- Gut instinct
    - Brining discovery (what is going right/wrong)

Technological Components of Business Intelligence

1. Data Source
    - Involves various forms of stored data.
    - Taking raw data and using software applications to create meaningful data sources that each division can use to positively impact business.
2. ETL Processor
3. Data Warehouse
    - Aggregate all structured data from one or more resources so that it can be compared and analyzed for business intelligence whenever required.
4. OLAP (Online Analytical Processing) (Analytical Engine)
    - Allows executives to sort and select aggregates of data for strategic monitoring.
5. Dashboard (Interface)
    - A visual display of the most important and relevant information that can be viewed in one summarized glance.
    - Key characteristics BI dashboard
        - All the data and visualizations fit on a single computer screen so there is no need to scroll them.
        - Displays most important performance indicators which need to be monitored on a regular basis.

Dashboard components

- Charts and graphs
- Underlying statistical processing
- Dimensional navigation
    - Selection
    - Ranging
    - filtering/slicing
    - Drill-down
    - Drill-through

Technological Components of BI

- Data Warehouse: Collection of performance data
- Online Analytical Processor (OLAP): use data for statistical treatment
- Dashboard: Graphs & Visualizations, you make quick decisions based on the dashboard.

BI Modes		
Age of Data	**Traditional BI**	Historical data, Data is at most one month old
	Real-time BI	Running data, advanced technologies that quickly performs integrations & calculations at real-time speed
Scope	**Enterprise BI**	Dashboards for whole company
	Departmental BI	Certain segment in a company
Power vs. Usability	**Usability BI / Predefined application-based**	App development, tool is pre-developed
	User self-service BI / Power BI	User builds BI, tool development, users choose what elements/insights they want

Scope of BI	
Descriptive Analysis	What is happening / what happened
Historical Analysis	What happened
Predictive Analysis	What is to happen
Prescriptive Analysis	What is to be done

Data Warehouse	**Big Data**
Internal Data	External Data
Structured data	Unstructured and structured
	Velocity, volume, variety
Known data	**Unknown data**
Captured and collected data	Mined data

Balanced Scorecard

What is the basis for identifying correct dimensions and measures expected in a data warehouse?

- Doing a balanced scorecard to determine Key Performance Indicators
    - Developed by Kaplan and Norton (1990)
    - Strategic Management Approach in order to measure the health of the organization
    - Translate mission and vision of the organization into tangible and measurable elements that can be monitored and analyzed
    - Performance cannot be viewed in a single perspective

How is a balanced scorecard done?

- Top-level effort, where scorecard elements are defined first at the highest level of the organization (Organizational scorecard)
    - Each element of the scorecard is further defined (specifics) for each department (Departmental scorecard)
    - In some cases, the scorecard trickles down to individual level (Employee scorecard)

What are the elements being defined in a balance scorecard?

- These elements are in a hierarchy
    - An element is defined first before the lower elements are defined

Elements of a balanced scorecard

- Vision
    - What you want to become
    - Aspirations as an organization
    - World problems you aim to solve or contribute to solve
    - Societal change you wanted to contribute to
- Mission
    - What you want to do
    - Who do you want to do it for
- Strategic priorities
- Strategic results
- Business Strategy Map
- Performance Measurement System

Elements of a Business Strategy Map

Note: Strategies are defined based on a hierarchy of perspectives of health of an organization. Normally, three strategies are defined per perspective.

1. Organizational capacity
    - Human and technological element of strategies
    - Human capability
    - Technological advancement
2. Internal Process
    - Effect of improved Organizational Capacity is the improvement in the implementation of process
    - Process improvements expected
    - Quality of processes expected
3. Customer
    - Effect of improved Internal Processes and Organizational Capacity is positive results in the eyes of the customers
4. Financial
    - Positive results in the perspective of customers will result to improve financial situation of the organization

Performance Measurement System

- Objective
- Measure
- Metric and formula
- Weight
- Targets for the next 3 years
- Initiatives or projects to achieve targets

Parts of BSC:

Perspective	Strategy	Measure	Metric/ Formula	Target (Years)
Financial		“Per dimension”	“Per sub-dim”	
Customer				
Internal Processes				
Organizational Capacity				

Dimension	**Sub-dimension**
time	Year, quarter, month, week
department	Division, department, unit, team
product/service	Category, type, product, location
Customer location	Country, region, city
Supplier location	Country, region, city

Note:
*measures in BSC are NOT data
*metric = formula

*dimensions = primary key
*measures = not primary key

*1st degree measures - those recorded in the scorecard
*summary data
- Cannot be dissected
- Representation of summaries
- Sum, ave, total, range, count, central tendency, variability

*extensions
- 1st degree measures
- Extension summary
- Referential data (non-additive)
- Non additive: ordinals; additive: scales

Charts/Visuals

Chart	Primary Purpose	Secondary Purpose
Bar	Communicate achieved values on a dimension	Compare values across dimensions
Clustered bar	Communicate achieved values on several measures on a dimension	Compare values on several measures across dimension
Stacked bar (Level)	Communicate the participation of a measure to the total value of all measures	Compare participation of a measure across dimensions
Stacked bar (100%)	Communicate the participation of a measure in a percentage	Compare participation of a measure in percentage across dimensions
Line	Trend on a continuous dimension (eg. Time)	
Clustered Line	Trend on a several measures on a continuous dimension	Compare the trend of several measures
Area Chart	Focus, concentration of a measure across a continuous dimension	
Clustered Area	Opposing measures Exclusive are occupied by measures	
Stacked Area	Complementing measures	
Pie	Communicating the participation of the dimensions	
Donut	Communicating the participation and intensity of the participation of the dimensions	
Aster	Communicating the participation and intensity of the participation of the dimensions	
Gauge	Communicating how far or near a measure is from its target	
Map	Compare the value to geographical area	

	Compare the value to geographical proximity	
Card	Communicate a summary value	
Scatter Plot	Communicate the relationship of two continuous measures	

Special Charts	
Chart	**Description**
Radar Chart	- Levels attained by different elements - To show area (impact of measure to another measure) - To show bias
Tornado Chart	- Only used for opposing values - To avoid making a new column (for negative values) - To show the relationship between opposing values - Has a limit (up to 7 categories)
Waffle	- Only one to show values exceed > 100% - Show achievement % of a value that may exceed 100% - Cannot drill down

What are the Different Kinds of Values Shown in Charts?

- Total
- Average
- Min/Max
- Count
- Percentage (relative value)
- Cumulative Value
- Median

Row Value

- Total
- Average
- Min/Max
- Count

Set Value (something computed with other values) Computed Using DAX

- Percentage
- Median
- Cumulative Value
- Rank

DAX

- Computed per row -> simple DAX (New Column)
- Computed as a whole -> complicated DAX (New Measure) cannot access individual column

✔ you CAN use a measure to create new measure
✔ you CAN use a column to create new measure
✘ you CANNOT use a measure to create new column

Ranking

- Prominent to correlation
    - Attempting to determine/show if one measure affects the other measure
    - Used in scatter plot
- Easier to correlate variables if value rank is plotted

R

- In Power BI, concerned on what will be shown in advanced charts
- Will not activate without dataset
- Correlation coefficient with scatter plot to show the correlation clearly
- All functions need library
- Mathematical language