

Business Analytics with Python

Trainer: Michael Li

Venue: NTUC Learning Hub

Date Conducted: 13 Sep 2018

Outline

- What is Business Analytics
- What is Python
- Business Analytics in Industry
- Data Analytics Examples
- About Data
- Data Tutorial
- Business Analytics Tutorial

About Your Trainer

Michael Li



Biomedical Engineering Degree, BioInformatics Specialist

Expansive Experience

Biomedical Devices and Diagnostics

Systems Engineering and Server Administration

Analytics Consultancy

Data Scientist

Innumerable strategies to solve **Data-Fusion** problems

Solution Architect

Recommendation Engines

Semantic Technologies

Fraud Analytics

Text Mining

www.linkedin.com/in/michaelqali

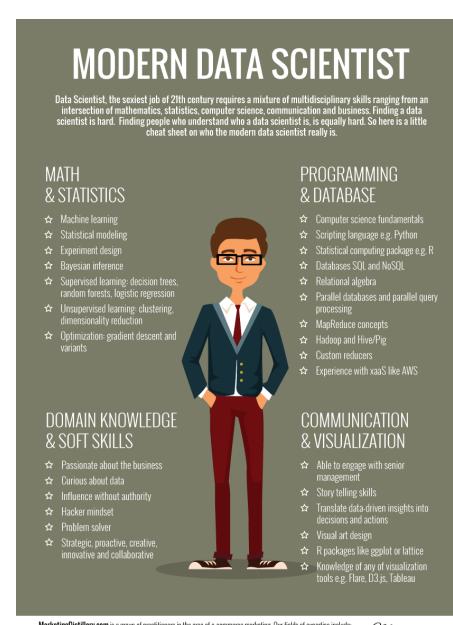
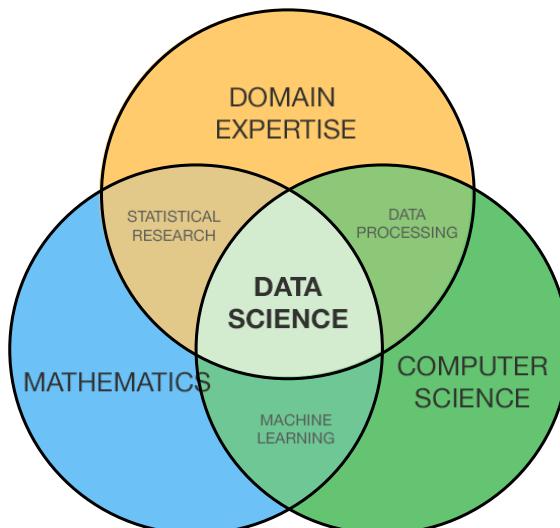
BUSINESS ANALYTICS WITH PYTHON

What is Business Analytics

What is Data Analytics

- The science of drawing insights from raw information sources
- Also commonly known as Data Mining, Data Science,
- Has multiple branches:
 - Big Data
 - Text Mining
 - Business Analytics
 - Machine Learning
 - Quantitative Analytics

What is Data Analytics



Definition of Business

- Investopedia: A business is an organization or enterprising entity engaged in commercial, industrial, or professional activities... to produce and sell goods and services for profit.
- A sustainable system of proving goods and service to the market to ensure the livelihood and longevity of the entity and its future.
- Some main components
 - People (Human Resource)
 - Operations
 - Sales and Marketing
 - Accounting
 - Finance
 - Strategy

What is Business Analytics

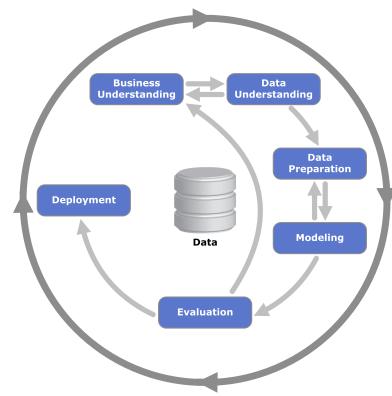
- Data Analytics is the drawing insights from raw information sources
- Business Analytics is the application of Data Analytics in order to aid and/or drive business performance
- The application mostly revolves around rapid prototyping and scaling of
- Business Analytics have been known to involve
 - Business Intelligence
 - Business Process Modeling
 - Six Sigma
 - Root Cause Analysis
 - Customer Profiling
 - Marketing Campaign Analysis

Business Analytics Skills

Main Skills Category	Why	Examples
Business and Communication	Understand and extract Business Goals, Key Metrics. Define, formulate and deliver actionable targeted analyses. Work with stakeholders.	CRISP-DM, Key Performance Indicators, Change Management, Business Process Modelling, Stakeholder Management
Visualization and User Interface	Create easily digestible, usable dashboards and interactive charts.	Business Intelligence (Tableau, QlikSense), Visualization Libraries (D3.js, bokeh, dash)
Statistics	Pick out significant factors from data. Find outliers in data points. Understand the quality of the data provided. Quantify the accuracy or applicability of tests done.	Chi-square, Standard Deviation, Pearson Correlation
Software Engineering	Data Management e.g. Extract-Transform-Load. Quick prototyping.	Programming skills (Python, Java, Excel VBA), BA Tools (RapidMiner, TIBCO Spotfire, SAS), Enterprise Analysis Tools (OLAP, OLTP)
Machine Learning	Assess feasibility of running Predictive Modelling on data.	Precision-Recall, Confusion Matrix,
Science & Engineering	Formulate an analysis approach to a problem. Identify or create measurable performance indicators. Apply best practices and standard procedures for different domains. Create pipelines for analyses	AB-testing, Standards/Body-Of-Knowledge, Unified Modeling Language.
Social Skills	Keep in touch with current technology trends and techniques. Maintain network of contacts from various domains.	Knowledge Sharing, People Management

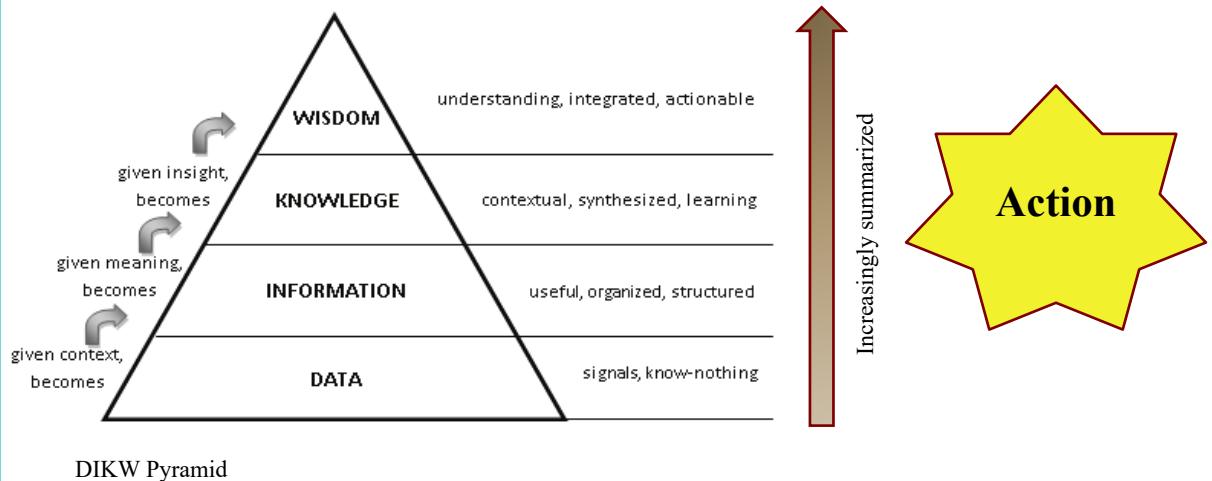
Methodology

- Cross-Industry Standard Process for Data Mining (CRISP-DM)
- CRISP-DM is a very general approach and widely used in analytics projects.
- Due to the highly varied nature of data from different entities, the length of an Analytics project may range from 1 week to 1 year (excluding deployment).
- Thus, it is now a recommended practice to conduct a small-scale feasibility study first.

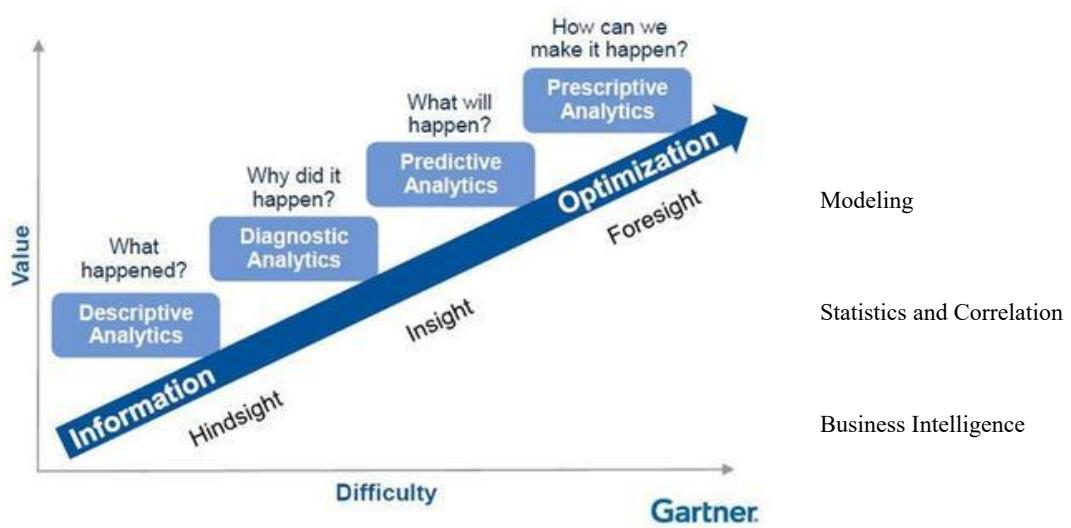


6 Phases of CRISP-DM

From Data to Insights



Outcomes from Analytics



BUSINESS ANALYTICS WITH PYTHON

What is Python?

What is Python?

Official Definition:

Python is an **interpreted, object-oriented, high-level programming language** with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for **Rapid Application Development**, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are **available in source or binary form without charge for all major platforms**, and can be freely distributed

Why/Why Not Python?

Why:

- Easy to pick up
- Community Support is very good
- Documentation of Libraries are mostly good
- Many libraries (especially the data science ones) are computationally efficient!
- Rapid Prototyping
- It can do many things!

Why not:

- Some people think its still a simple scripting language and not an Enterprise ready language
- Library not built yet
- It can do many things!

Libraries to note

Library	Why	Remarks
IPython Notebook	Provide a data science notebook to run scripts and and in line documentation.	Essential. IDE and notebook
Pandas	Data Frames in R was really cool. Python now has it too!	Essential. Rapid prototyping analysis.
SQLAlchemy	SQL toolkit and object relational mapper	Essential.
Scrapy	Web scraping framework	Essential.
Flask	Web microframework	Essential. Rapid prototyping and scalable to production.
Matplotlib	Data visualization	Essential. Basic plotting library
Scikit-learn (sklearn)	Machine Learning Library	Essential. Well documented with examples.
Sklearn_pandas	Bridge between pandas and sklearn	Good to know. Bridge between Pandas and sklearn
Networkx	Graph/Network Analytics Library	Good to know.
NumPy	Computationally efficient numerical calculation/storage package.	Learn-as-you-go
SciPy	Computationally efficient matrix calculation/storage package.	Learn-as-you-go

The Pandas Library

The pandas library is a Data Frame library (inspired from R) with a collection of data manipulation packages.

It is deemed essential in any Data Scientist's toolbox due to the ease of running descriptive analyses (like R Data Frames) and data manipulation.

The manipulation operations are computationally efficient as the “batteries” behind it are NumPy and SciPy.

The SKLEARN Library (Scikit-Learn)

The Scikit-learn library contains a wide collection of machine learning algorithms with extensive documentation on each family of algorithms and examples of the usage of each algorithm.

See the user guide http://scikit-learn.org/stable/user_guide.html

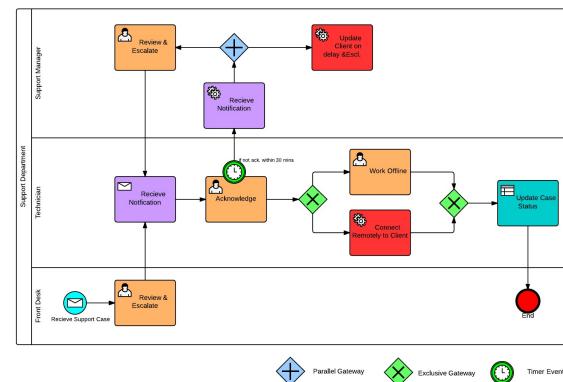
- Please check that you have a working copy of Python on your machine
- <https://www.anaconda.com/download> for those that have not yet installed Python on their system.
- There are slight differences between Python 2 and Python 3
- For our exercises; we will be using Python 3.
 - Those using Python 2 may want to set up an environment with Python 3 using <https://conda.io/docs/user-guide/tasks/manage-python.html>
 - Otherwise, there will be a few lines of code that need to be changed e.g. print statements

BUSINESS ANALYTICS IN PYTHON

Business Analytics in Practice

Business Process Modeling

- The reflection of current business operation processes into diagrams. Such diagrams are used to identify bottlenecks and key points of intervention for process optimization
- Involves:
 - Unified Modeling Language
 - Business Process Optimization
 - Root Cause Analytics
 - Data Flow Diagram



<https://www.mcftech.com/use-bpmn-flowcharts/>

Sales Funnel Analysis

- Funnel Analysis is the examination of a pipeline of events that lead to a specified goal or target.
- Sales Funnel Analysis looks at the conversion rates of the general public into customers.
- Funnel Analysis has also been applied to Digital Marketing and Web Analytics



Source: <https://www.brafton.com/blog/content-marketing/the-four-sales-funnel-metrics-you-should-care-about/>

Cost-Benefit Analysis

- The study of how a change to a system will impact the business and how to best affect change management in order to optimize resources
- Also applied to evaluation of projects and
- Involves
 - Fermi Estimation (Back-of-envelope calculation)
 - Pareto Principle (Law of the vital few. 80/20 rule)
 - Accounting

CAPEX		Quantity	Solution1		Solution2	
No.	Items		Cost per unit	Total Cost	Cost per unit	Total Cost
1	Platform	8	\$22,000	\$176,000	\$22,000	\$176,000
2	Sensors	8	\$10,000	\$80,000	\$10,000	\$80,000
3	Actuators	8	\$20,000	\$160,000	\$20,000	\$160,000
4	Services	8	\$77,000	\$616,000	\$150,000	\$1,200,000
5	Development	1	\$80,000	\$80,000	\$160,000	\$160,000
Total				\$1,112,000		\$1,776,000

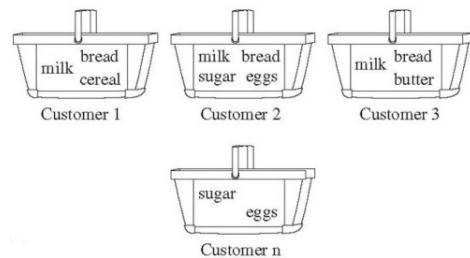
OPEX		Quantity	Solution1		Solution2	
No.	Items		Cost per unit	Total Cost	Cost per unit	Total Cost
1	Servicing	1	\$60,000	\$60,000	\$50,000	\$50,000
2	OOE	1	\$1,000	\$1,000	\$1,000	\$1,000
Total				\$61,000		\$51,000

Benefit		Quantity	Cost per unit		Total Cost	
No.	Items		Cost per unit	Total Cost	Cost per unit	Total Cost
1	Manpower Savings	20	\$60,000	\$1,200,000	\$60,000	\$1,200,000
Total				\$1,200,000		\$1,200,000

Return on Investment		Year				
			1	2	3	4
			\$27,000			
		1		\$1,166,000		
		2			\$2,305,000	
		3				\$1,671,000
		4				\$2,820,000
		5				\$4,583,000
						\$3,969,000

Market Basket Analysis

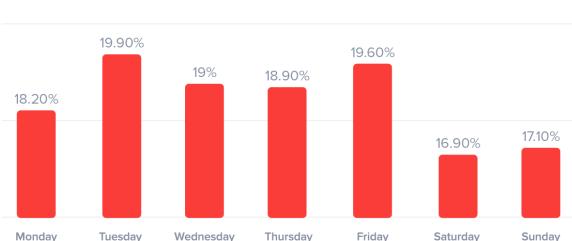
- Technique to find associations from purchasing/sales data.
- The principle behind this is that customers tend to buy certain products together
- Related to:
 - Affinity Analysis
 - Collaborative Filtering



Customer Profiling

- Customer profiling is about understanding your customer's preferences and tendencies. Usually used to give better and optimum recommendations, and to improve on customer's experience
- Involves
 - User behavior analysis
 - Pareto Principle (Law of vital few)

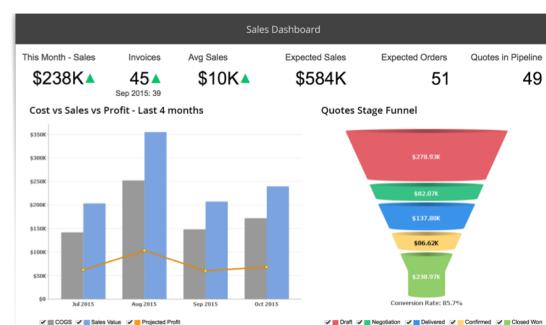
Open rates by day of week



Example: Email open rates of Electronic Direct Mail

Business Intelligence

- Business Intelligence is the summarizing of Enterprise Data into a dashboard to enable a higher level of tracking key metrics.
- Involves:
 - Online Analytical Processing (OLAP)
 - Decision Support Systems



Competitive Intelligence

- Strategic analysis of competitors in the enterprise's industry. Key metrics mostly involve market share and market penetration rate.
- Involves
 - Corporate Intelligence
 - Public Relation Intelligence (Media Scanning)
 - Patent watch
 - Competitor profiling

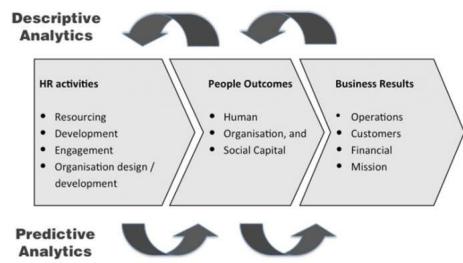


BUSINESS ANALYTICS IN PYTHON

Data Analytics Examples

Human Resource Analytics

HR Value Chain

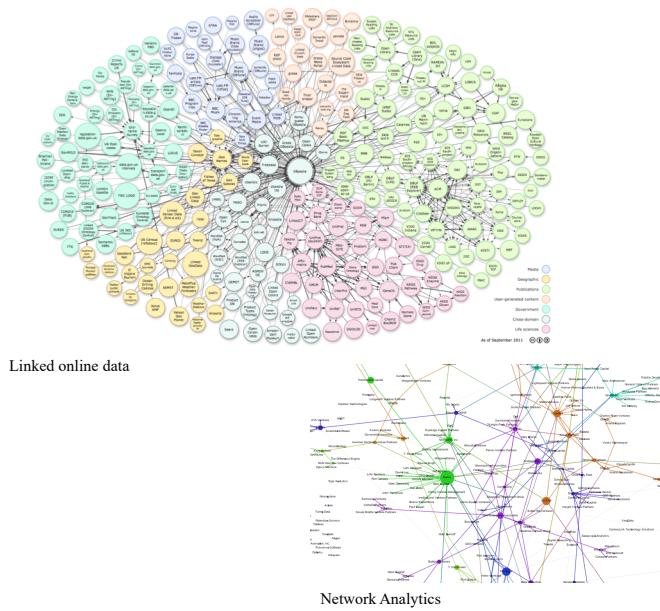


Human Resource creates a variety of data during the operation of the business. These datum can be in the form of exit surveys, employee feedback, claims, salaries, etc.

Involved in:

- Workforce/Talent Profiling
- Turnover Prediction
- Performance Tracking
- Skills Audit

Graph Analytics



The study of interactions between entities is known as graph analytics or network analytics.

From the interactions between each entity, we are able to discover insights via cliques and gateway nodes.

Involved in:

- Fraud Detection
- Clustering
- Entities-of-interest
- Semantic Technologies

Text Analytics

I was **really happy this morning** walking into your store until when I needed to ask a question and **the staff were incredibly rude**. Normally you have **super fast service**. I wanted an item I saw online and tried to call the store but the **phone was always engaged**. Over experience was **not very nice**, the **staff didn't care**. Usually they are really **helpful**, I'm not sure what happened today.

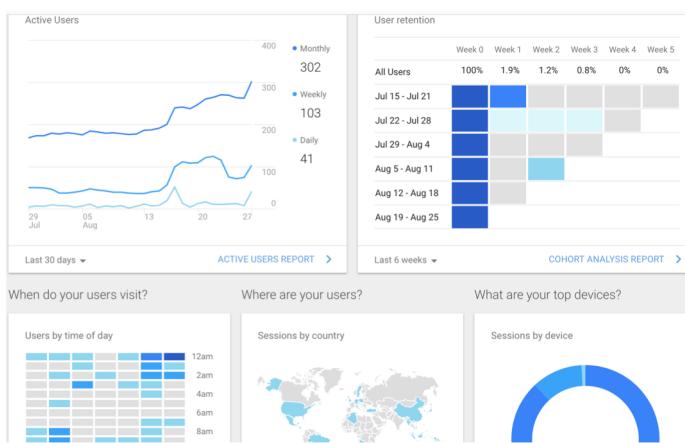
Text Extraction

Text Analytics can be domain by itself due to the complexity involved in working with text data. The majority of effort is in the information extraction from documents.

Involves:

- Document Clustering
- Sentiment Analyses
- Natural Language Processing (Linguistics)
- Named Entity Recognition

Web Analytics



Google Analytics (Web Tracking)

Web Analytics is the collection and summarization of web data to your website. The insights are mainly used to track traffic and can be used in conjunction with search engine marketing (SEM)

Involves:

- Site Traffic Monitoring
- User Profiling
- Click Analytics

Social Media Analytics



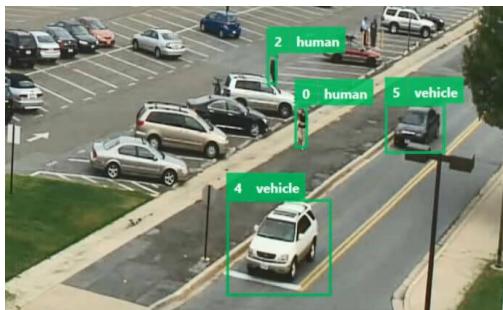
Social Media Sensing

Social Media Analytics is about tapping into different social media channels, and summarizing them into a coherent picture for the user.

Involves:

- Campaign Tracking
- Brand Tracking
- Sentiment Analysis

Video Analytics

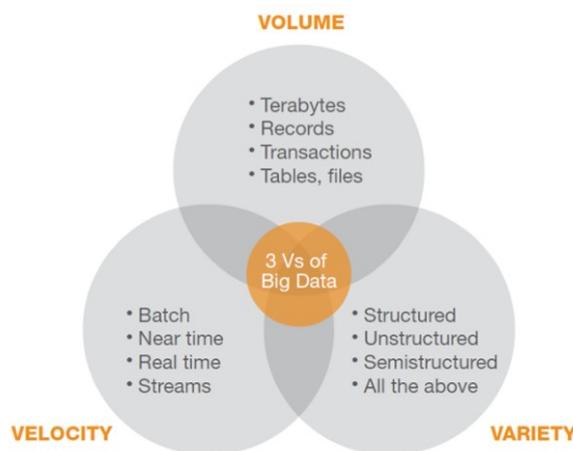


Video and Image Analytics mostly revolve around the extraction of features from the image or video data. It can be a highly complex area of analytics especially if it has to be in real-time (see autonomous vehicles)

Involved in:

- OCR (Optical Character Recognition)
- Traffic Analysis
- Security Monitoring

Big Data



Big Data is not just about Volume. It can be any or a combination of any of the 3 Vs. Most Big Data are transactional data.

- Volume -> Size of data, quantity of data files
- Velocity -> Amount of data to process per time unit
- Variety -> How do I integrate many different sources of data

BUSINESS ANALYTICS WITH PYTHON

About Data

What is Data

Structured vs Unstructured Data:

Structured data is typically well organized e.g. data tables.

Unstructured data typically refers to text data where there is no well defined form e.g. images, videos, text documents

Raw vs Processed vs Summarized Data:

Raw data or processed data is preferable over summarized data, although some information may be lost after processing data.

Different Types of Data

Identifiers:

Data fields that are used to **uniquely identify entities** of interest e.g. IC number, company registration number, purchase order number



Categorical:

Data fields that allow the data set to be grouped into **categories**. E.g. gender, occupation, department



Measures:

Data fields that hold numbers that quantify the **value** of an entity. E.g. Age, price, quantity



Tutorial

- IPython Refresher
- Identify different types of data

BUSINESS ANALYTICS WITH PYTHON

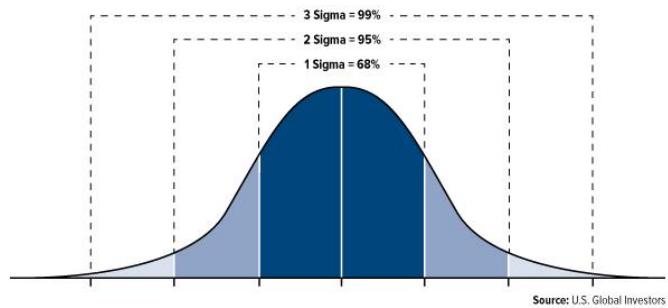
Descriptive Analytics, Affinity Analysis

Descriptive Analytics

- **Descriptive Analytics** is the examination of data or content, usually manually performed, to answer the question “What happened?” (or What is happening?), characterized by traditional business intelligence (BI) and visualizations such as pie charts, bar charts, line graphs, tables, or generated narratives.

- What we do
 - Data Distribution
 - Data Visualization

Standard Deviation (Sigma) Measures Degree of Variance from Average



Affinity Analytics

- **Affinity Analytics** examines data or content to answer the question “Are there items or relations that co-occur?”
- What we do
 - Market Basket Analysis (Apriori Algorithm)
 - Customer Profiling

Predictive Analytics

- **Predictive Analytics** describe the use of statistics and modeling to determine future performance based on current and historical data. Predictive analytics look at patterns in data to determine if those patterns are likely to emerge again, which allows businesses and investors to adjust where they use their resources in order to take advantage of possible future events.
- What we do
 - Linear Regression
 - Classification

Sentiment Analysis

- **Sentiment Analysis** is a Text Mining technique applied to social media data in order to summarize users' feelings over a set of documents.
- Involved in:
 - Social Media Sensing
 - Case Prioritization
 - Brand/Product Line/Feature tracking



Tutorial

- Conduct Data Understanding
- Descriptive Statistics
 - Anomaly Detection
 - Customer Segmentation and Profiling
- Affinity Analysis
- Predictive Modeling
- Sentiment Analysis with nltk
- Hands-on Rapid Prototyping
 - Data Wrangling
 - Data Scraping
 - APIs and Webservices

BUSINESS ANALYTICS WITH PYTHON

APPENDIX

Additional Resources

- Business Analytics
 - Business Analysis Body of Knowledge
 - <https://insights.principa.co.za/logical-data-fallacies-to-avoid-in-data-analysis>
https://en.wikipedia.org/wiki/List_of_fallacies
- Ipython/Jupyter Notebooks
 - <https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks>
- Business Analytics with Python
 - <https://www.kaggle.com/learn/overview>
 - <http://pbpython.com/>

Community

Extra Resources:

<http://pugs.org.sg/pages/learning.html>

Local Community:

Python User Group (Singapore) <http://pugs.org.sg>

PyData-SG <https://www.facebook.com/groups/pydatasg/>

PyLadies-SG <https://www.facebook.com/PyLadiesSG/>

Events:

PyCon Singapore – usually held in June/July

Business Analytics with Python

Trainer: Michael Li

Venue: NTUC Learning Hub

Date Conducted: 13 Sep 2018