

Introduction to Business Analytics

Overview

Objective

After completing this lesson you will be able to:

- Describe business analytics
- Explain the components of business analytics
- Explain the usage of business analytics in various domains



In God we trust, all other must bring data
- W Edward Deming



Corporate Decision Making–The HIPPO Algorithm



Highest Paid Person's Opinion

Business Analytics–Definition

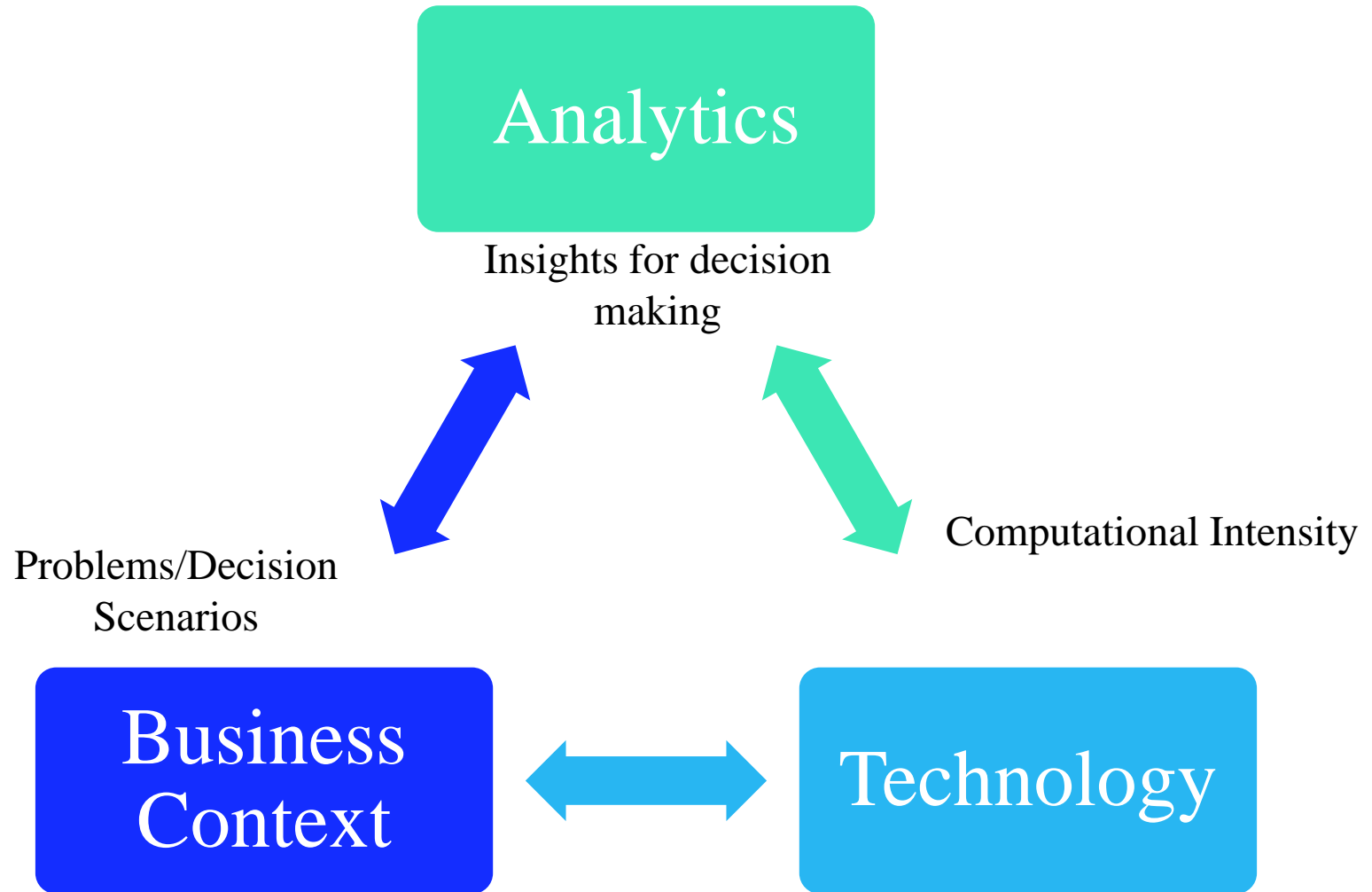
- Business analytics (BA) refers to the tools, techniques and processes for continuous exploration and investigation of past data to gain insights and help in decision making.
- Business Analytics is an integration between science, technology and business context that assist data driven decision making.

Data Explosion

- About 350 million photos are uploaded every day in the Facebook
- Amount of credit card debt in US: \$762.1 billion
- Amount of credit card debt in India: Rs. 45,383 crore (\$709 million)
- Loss due to global Credit card and debit card fraud \$21.84 billion during 2015
- Every day, Walmart processes \$36 million dollars an hour in sales
- BMTC with approx. 6000 buses plying in Bangalore sends 1 billion signals to the server updating its location every month

Interesting Stats: <http://expandedramblings.com/>

Analytics Trilogy



Why Analytics


- Analytics provides competitive advantage.
- Analytics removes inefficiency in the system/organization.
- Provides ability to make better decisions.

Analytics in Use—Flipkart

- Forecast demand for each SKU.
- Predict customer cancellations and returns.
- Predict customer contacts at the customer service.
- Predict what a customer is likely to purchase in the future?
- How to optimize the delivery system?



Analytics in Use—Big Basket



Search for more than 10,000 products...

Search

Your Basket
0 items
CHECK OUT

SHOP

OFFERS

NEW ARRIVALS

SHOP BY LIST

HOME > SMART BASKET

Smart Basket (125)
Collection of products that you spend on most or buy often.

SELECTED PRODUCTS (0)
☐ Select all


FOR SELECTED PRODUCTS:

ADD TO BASKET

COPY TO LIST

▼ Fruits & Vegetables (28)
☐ Select all

☐



FRESHO
Onion - Medium


1 kg

₹ 17.00

Qty

ADD

☐



FRESHO
Potato


1 kg

₹ 25.00

Qty

ADD

☐



FRESHO
Pomegranate - Kesar


1 kg

₹ 181.00

Qty

ADD

☐



FRESHO
Cauliflower (Medium) - Grade A


1 nos (approx. 500 ~)

₹ 19.00

Qty

ADD

☐



FRESHO
Banana - Robusta Semi Ripe (Grade A Super) (7...

1 kg

₹ 32.00

Qty

ADD

© Copyright 2015 All rights reserved.

How would you solve this?

6/10/2015

Flipkart delivers 2 stones instead of iPod to a user! What if...

Trak.in

Business of Tech, Mobile & Startups in India

it's gone. [Linkin](#)

What was wrong with this ad?

Repetitive

Inappropriate

Irrelevant

Google

HOME

BUSINESS

TECHNOLOGY

INTERNET

TELECOM

MOBILE

STARTUP

OTHERS

ABOUT

Home / Internet / Ecommerce / Flipkart delivers 2 stones instead of iPod to a user! What if...

Flipkart delivers 2 stones instead of iPod to a user! What if...

Posted by: Arun Prabhudesai | In Ecommerce, India | March 11, 2013 | 46 Comments

Now, this is outrageous – A twitter user tweeted today that his sister had ordered an iPod, but to her surprise she was delivered 2 stones inside the box.

Twitter user with handle @nikhilsekhar tweeted with accompanying picture of the said delivery.

“

@flipkart My sister got two stones instead of the INR 20K iPod that she ordered from Flipkart. What is wrong with you twitter.com/nikhilsekhar/s...

— Nikhil (@nikhilsekhar) March 10, 2013

Here is the photo accompanying that tweet:

The good part is Flipkart promptly replied the user with following tweets:

“

We are very sorry about the incident with our customer's iPod purchase, we're taking up this issue extremely seriously. 1/3

— Flipkart (@Flipkart) March 10, 2013

“

We have spoken to our customer and will make sure a replacement is sent over right away. 3/3

— Flipkart (@Flipkart) March 10, 2013

But what if...

Now, this kind of a thing happening is outrageous – and it is good to see that Flipkart has owned the responsibility and is going to offer an immediate replacement of the product.

However, I have a question in mind – What if the customer is lying?

Please bear in mind, I am not talking about this incidence, but putting across a theoretical situation.

What if a buyer actually replaces the delivery (after he has received the correct product) and then alleges that Flipkart has delivered him with stones (or whatever) inside it? What happens in that case. If you think that such kind of thing will not happen...you are wrong. There are many out there waiting to take advantage of the system..

Flipkart is in no position to actually contend that they have delivered it correctly, neither can they challenge the customer (and if they do, it will be one big social media mess).

<http://trak.in/tags/business/2013/03/11/flipkart-delivers-stones-instead-ipod/>

OH...IS IT?

FOLLOW US

45.6k Follows

RSS Feed

23.8k Followers

Facebook

16.4k Followers

Twitter

3.4k Followers

LinkedIn

0 Followers

Pinterest

775 Followers

Google+

1k Followers

FIND US ON FACEBOOK

How would you solve this?

Man orders Oppo phone on Amazon, gets dummy iPhone instead



Decision Making

Dosa King

- Established in 1992 by R. Narayanan under Indian Food and Fermentations Ltd. (IFFL)
- They have 250 Franchises across India, and a franchise in Bangkok



- IFFL1992: Dosa King Fast food Restaurant by R Narayanan
- 20 million Dosa consumed everyday (IFFL Stats 1992)
- Automated Dosa delivery system which will dispense a Dosa every 30 seconds
- VC funding to start the business of fast food Dosa chain

TIME BUT NO DATA



- Johnson & Johnson 1982: Certain people take Tylenol and die
- 37% market share with a revenue of \$1.2 million.
- James Burke has to take a call to recall the drug from market
- 31 million bottles with a loss of \$100 million

LITTLE TIME AND SOME DATA



- Jan 2008: British Airways flight BA038 from Beijing to London (Heathrow Airport)
- Just Before landing engine malfunctions
- Peter Burkill crash lands the flight and saves 152 lives

NO TIME BUT ACCURATE DATA

Decision Making–The Monty Hall Problem



After having seen “What lies besides door 1”, Would you like to switch?

The Game Changers

- Google
 - Used Markov chains to rank pages.
- Proctor and Gamble
 - Analytics as competitive strategy.
- Target
 - Predicts customer pregnancy.
- Capital One
 - Identifies the most profitable customer.
- Hewlett Packard
 - Developed “flight risk score” for 3,30,000 employees.
- Obama’s 2012 presidential campaign.
 - Persuasion Modelling.

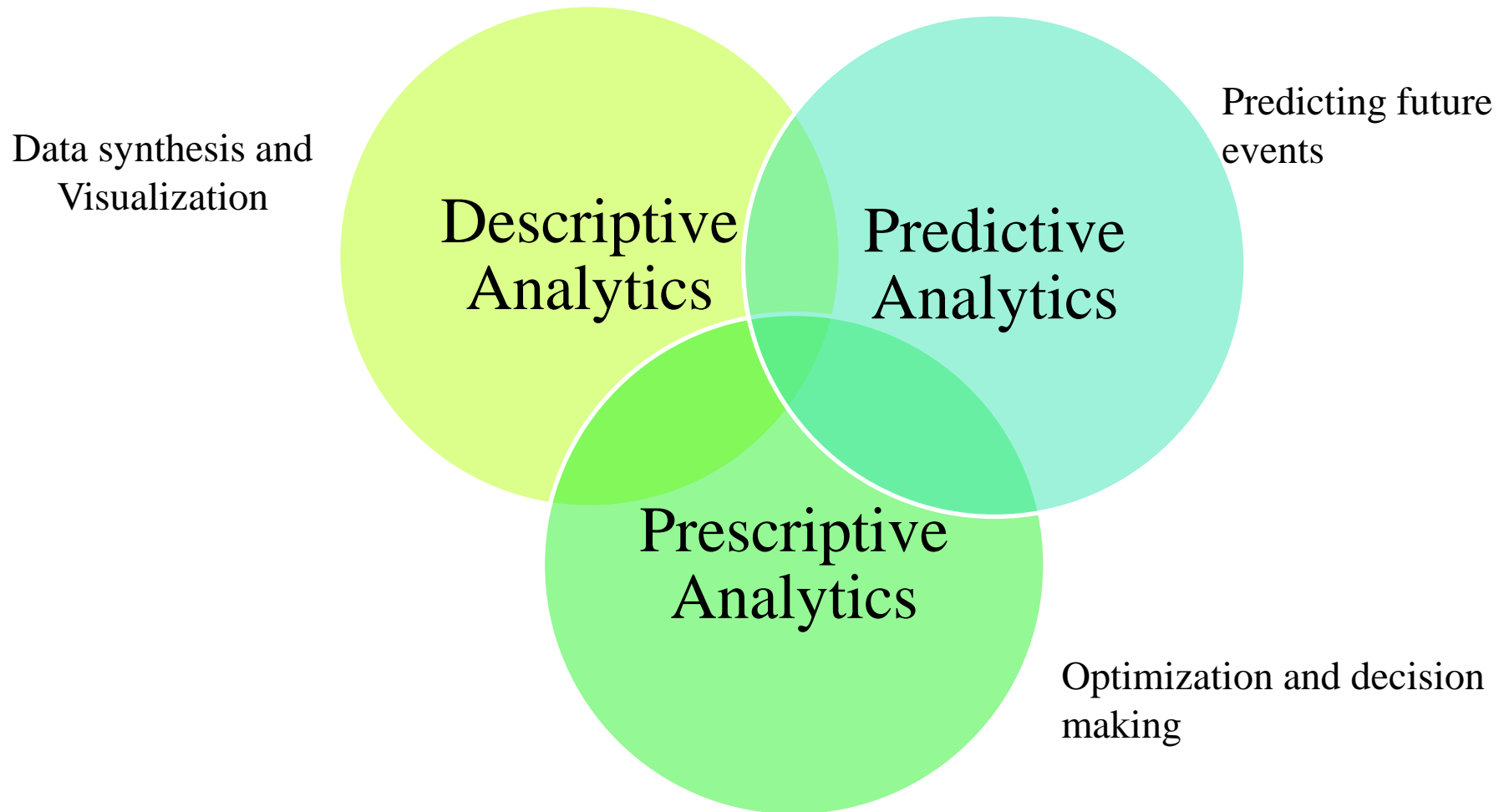
The Innovators

- OKCupid: Predicts which online dating messages is most likely to get a response!
- Polyphonic HMI: Uses “hit song science” to predict commercial success of a song.
- Netflix: Predicts movie ratings by customers (RMSE is 1%).
- Amazon.com: 35% of sales come from product recommendations.
- Citizens Bank: Predicted fraudulent cheques.
- Divorce360.com: Predicting success of a marriage!

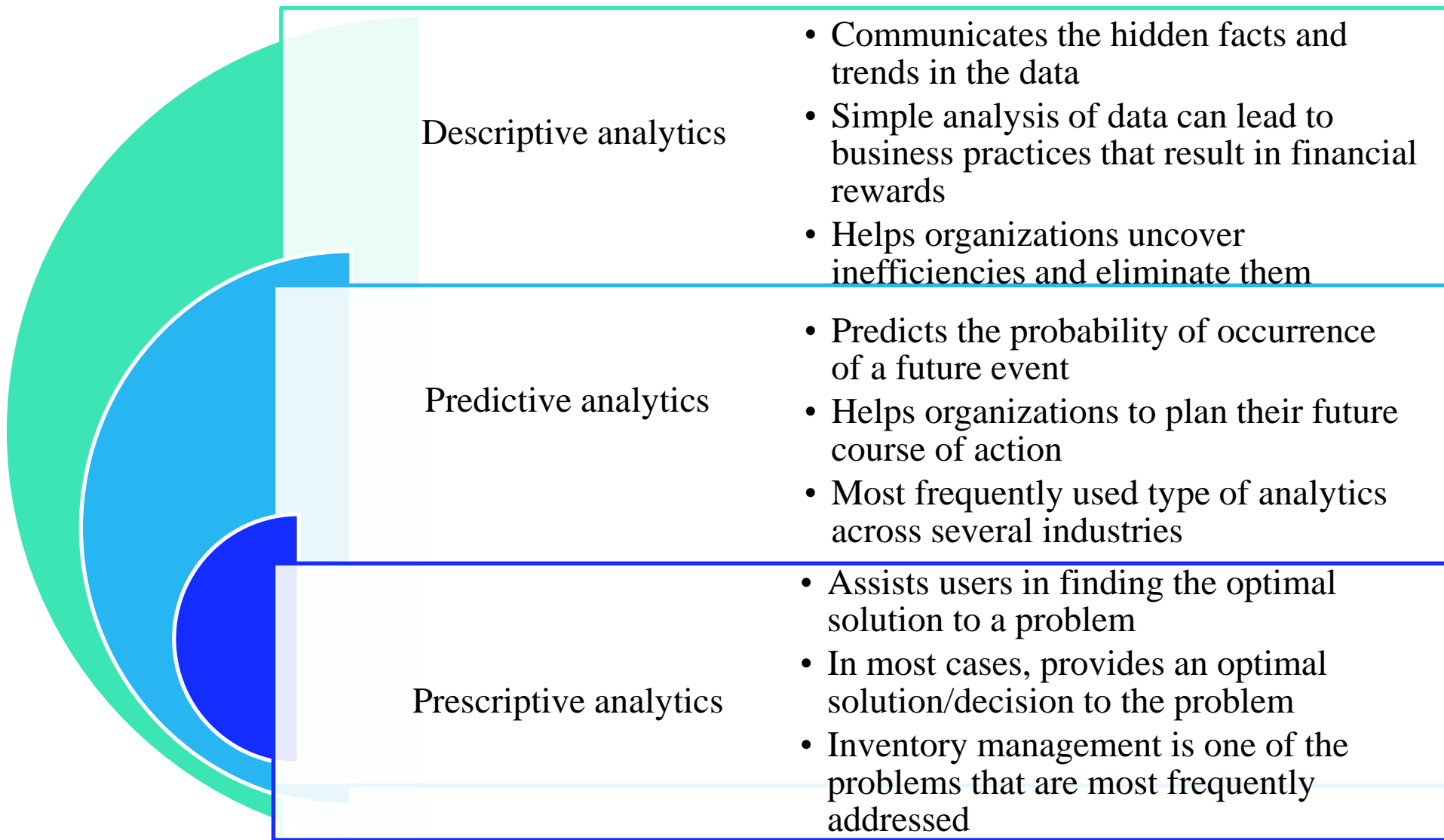
Data Scientists will be the sexiest job of 21st century

Harvard Business Review 2012

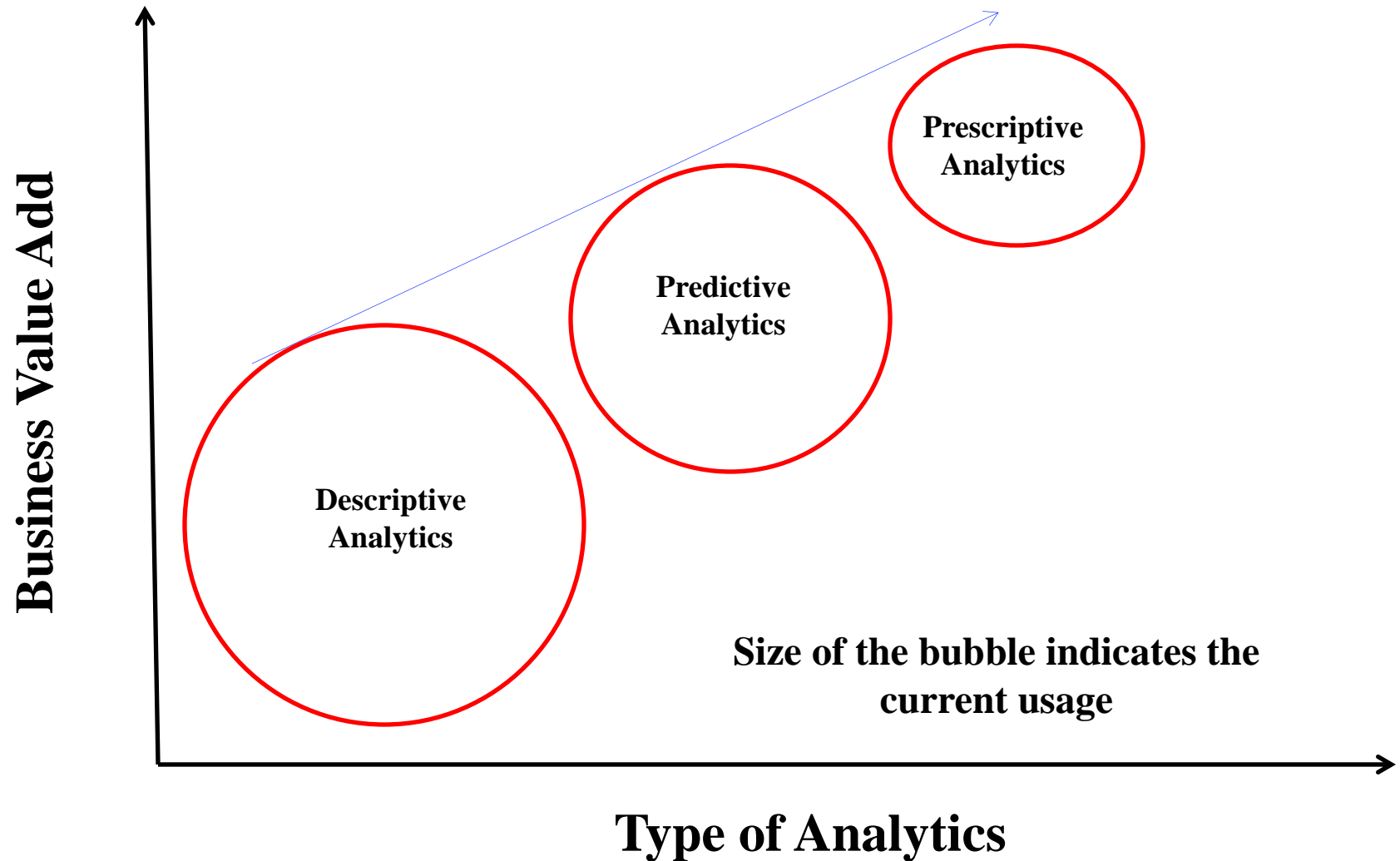
Components of Business Analytics



Components of Business Analytics



Business Analytics & Intelligence



Power of Descriptive Analytics

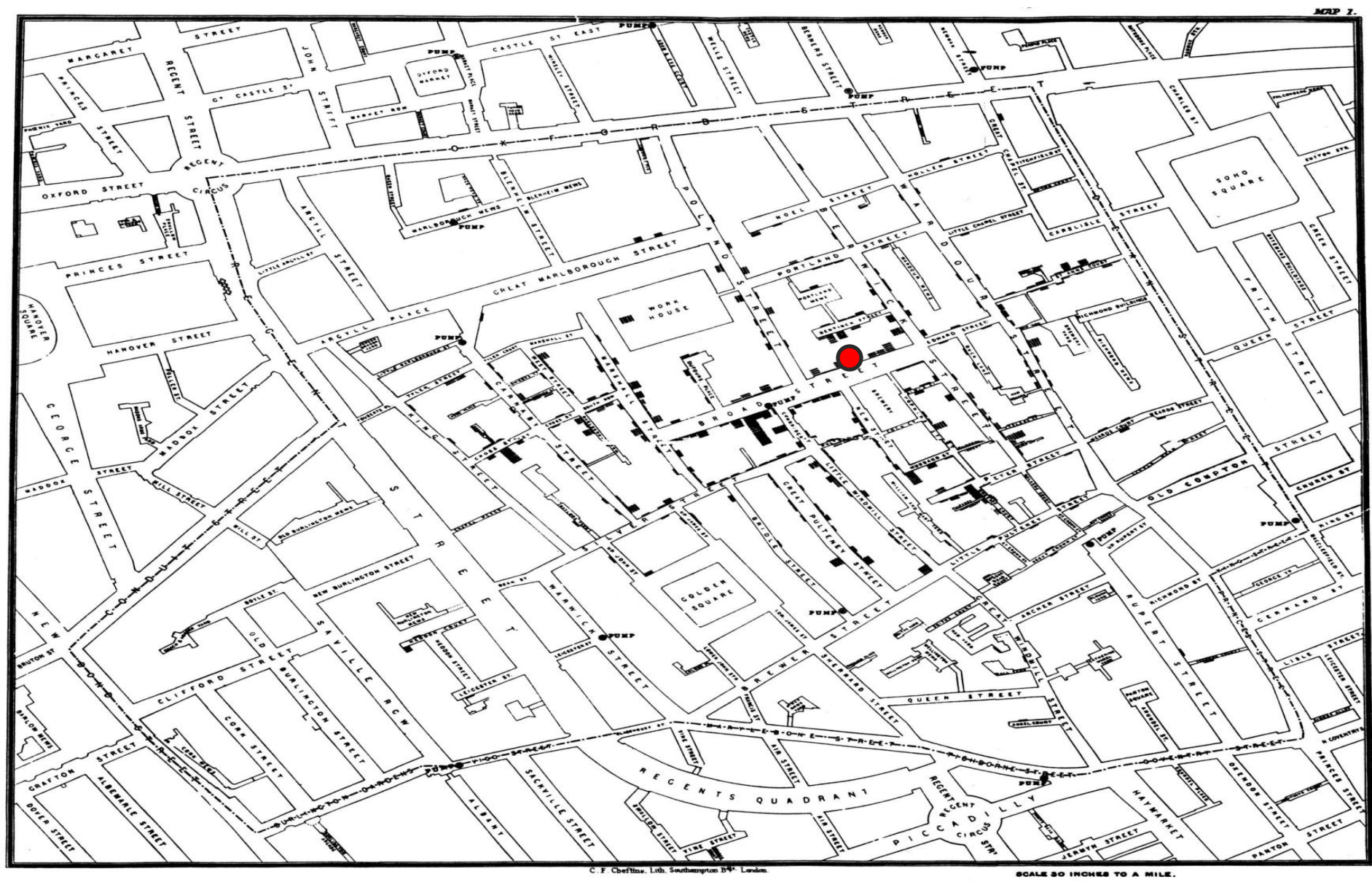
London Cholera Outbreak–1854

- Severe outbreak of cholera that occurred near Broad Street (now Broadwick street) in Soho district of London in 1854.
- More than 500 people died within 10 days of the outbreak, the mortality rate in some parts of the city was as high as 12.8%.



https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

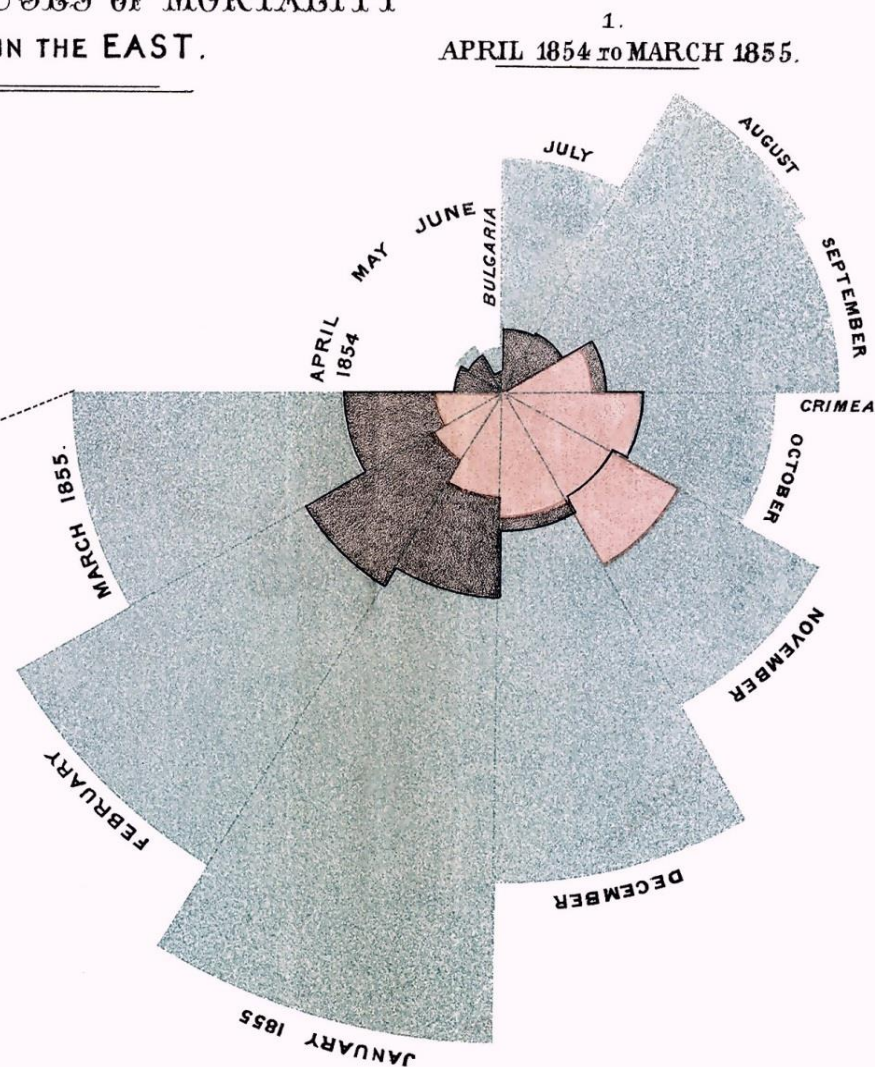
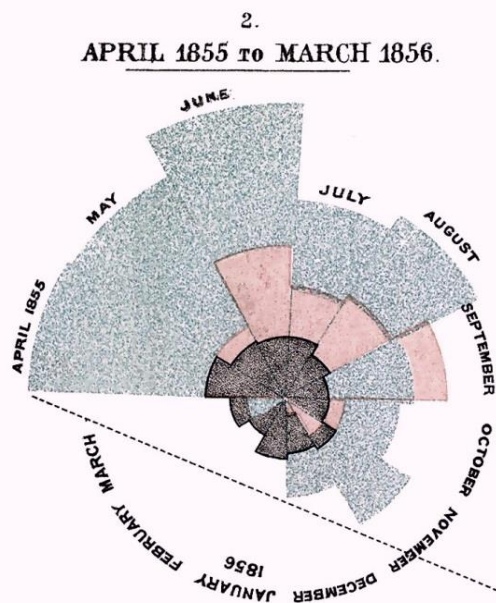
Visual Map of the Area



A historical map of a street grid in London, likely from the 19th century. The map shows a network of streets including Broad Street, Pump Street, and others. A red dot is placed on Pump Street, and a red rectangle highlights a building on Broad Street. The map is oriented with North at the top. Various buildings and structures are labeled, such as 'WORK HOUSE' and 'PUMP'. The streets are labeled with their names, and the map shows the layout of the urban area.

Florence Nightingale's Pie Chart– Crimean War Data (1853-1856)

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Novr 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

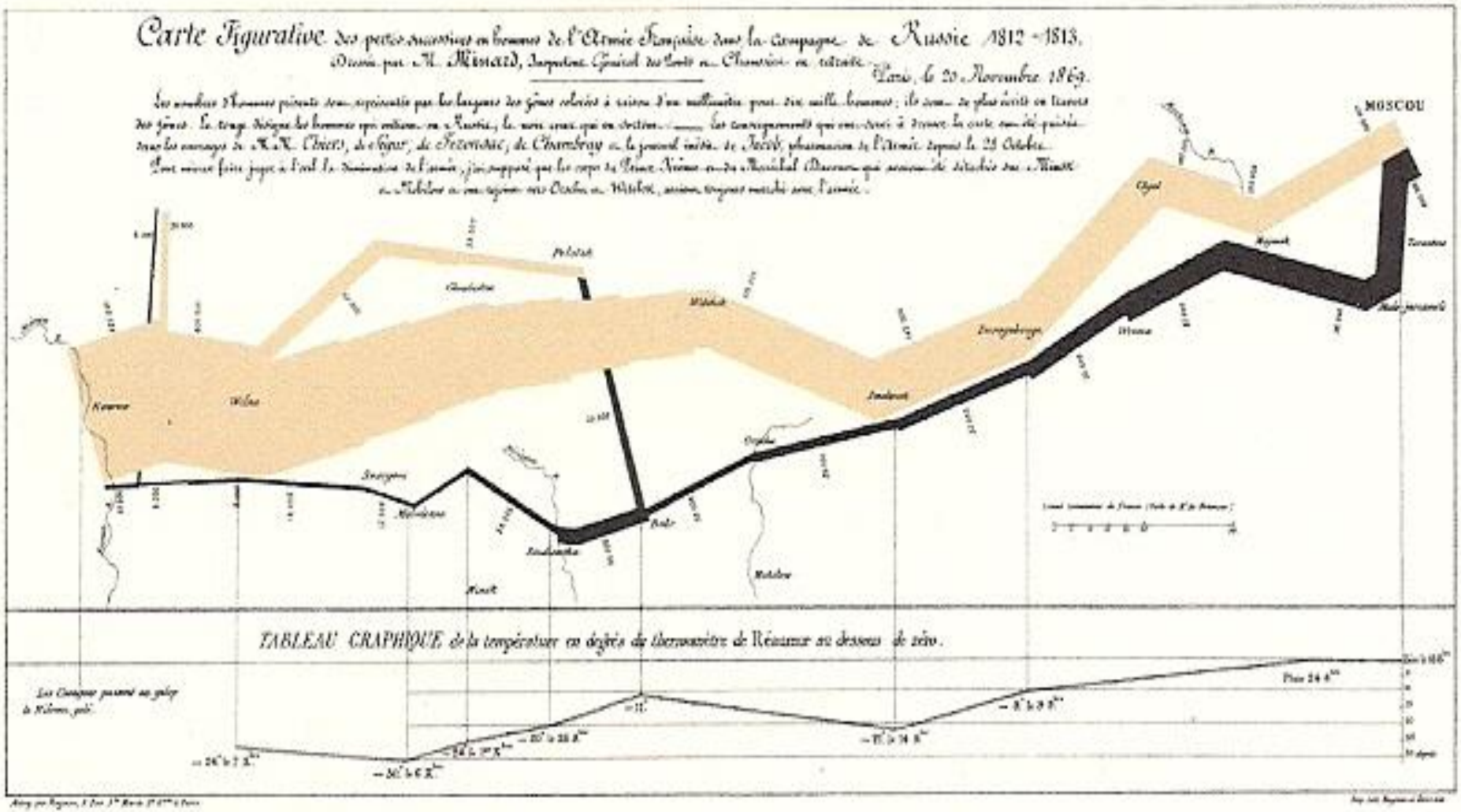
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

https://en.wikipedia.org/wiki/Florence_Nightingale

Napoleon's Invasion of Russia

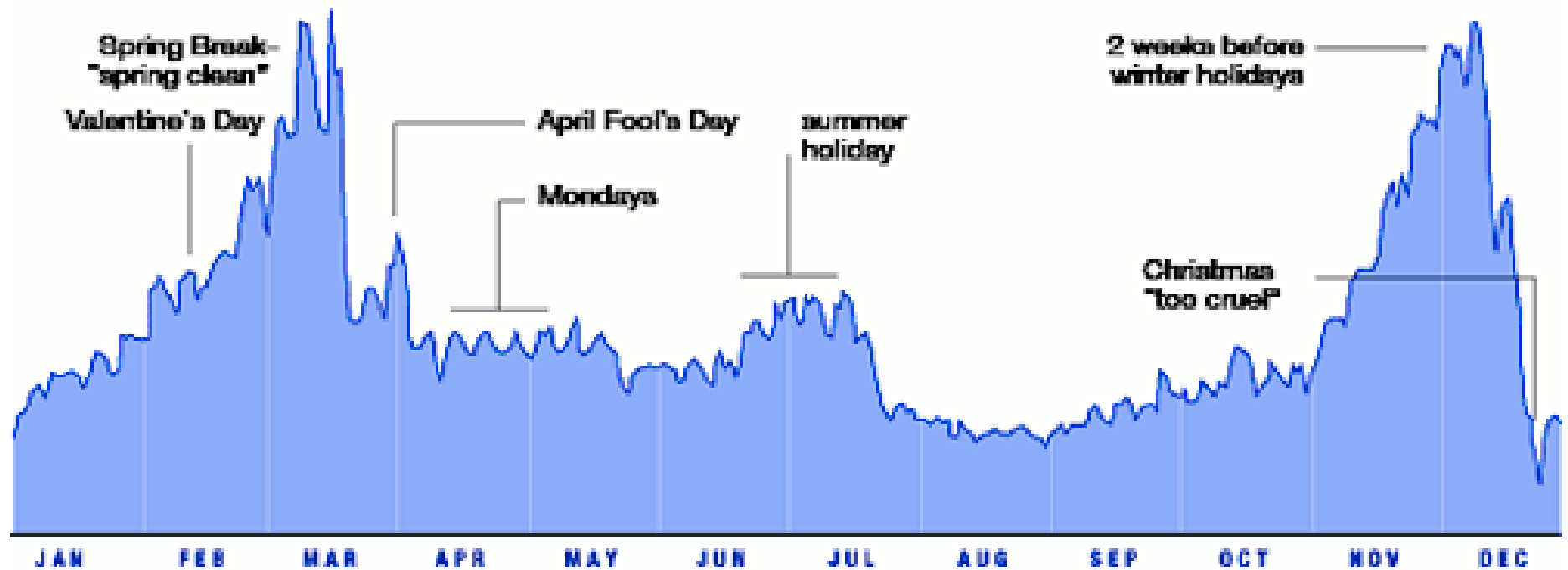
Losses of the French Army in the Russian Campaign 1812-1813, by Charles Joseph Minard

FRENCH ORIGINAL, 1869



https://en.wikipedia.org/wiki/French_invasion_of_Russia

Relationship Breakups Status Update on Facebook



Facebook Relationship Breakups

Infographics—Price of Love

Price of Love

The Valentine's Day tradition of chocolates, cards, flowers and dinner continues as lovebirds prepare the day for that special someone.

GoFigure!

SOURCE: YumaAZ.gov



\$1.7 billion
spent on flowers on
Valentine's Day
in the U.S.



Gifts

What single women prefer for Valentine's Day:

Gift certificate that can be shared together **38%**

Flowers **30%**

Jewelry **29%**

Low on the gift list:

Chocolate **5.2%**

Lingerie **1.6%**

It's the thought that counts



Source: Sears

\$3.4 billion
spent on Valentine's
dinner in 2011



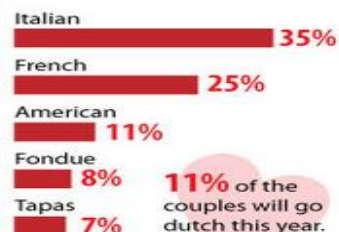
Romantic dinner



\$116.21
average spent on
Valentine's Day dinner
in 2011



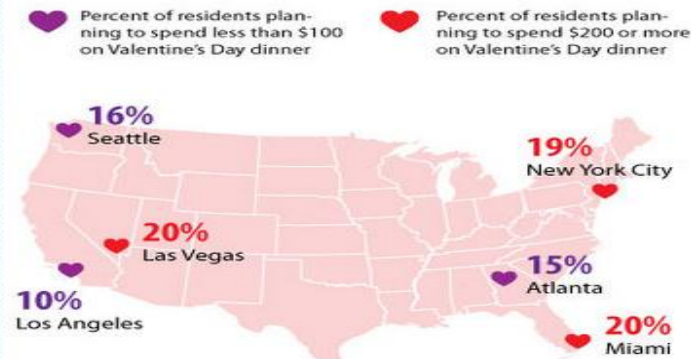
Cuisine



11% of the couples will go dutch this year.

Spending

More than half of the survey respondents expect to spend between \$100 to \$200 for the special meal.



Source: OpenTable (OPEN)

SOURCES: ITSJUSTLUNCH.COM, SEARS, MOTLEYFOOL.COM, GANNETT.COM, NATIONAL RETAIL FEDERATION, OPENTABLE

R. TORO / © LiveScience.com

SOURCES: ITSJUSTLUNCH.COM, SEARS, MOTLEYFOOL.COM, GANNETT.COM, NATIONAL RETAIL FEDERATION, OPENTABLE

R. TORO / © LiveScience.com

Descriptive Analytics Applications

- Most shoppers turn towards right when they enter the a retail store.
- Conversion rate of women shoppers is higher than male shoppers among electronic gadgets purchasers (Radio Shack).
- Strawberry pop-tarts sell 7 times more during hurricane compared to regular period (Wal Mart).
- Women car buyers prefer women sales person.

Predictive Analytics Application

- Which product the customer is likely to buy in his next purchase (recommender system).
- Which customer is likely to default in his/her loan payment.
- Who is likely to cancel the product that was ordered through e-commerce portal.

Prescriptive Analytics Application

- What is the optimal route for a delivery truck.
- Whether a company should introduce a new product?
- What is the optimal product mix?
- How to manage the fleet of vehicles owned by a company for employee drop and pick up?

Framework For Decision Making

Opportunity Identification

- Domain knowledge is very important at this stage of the analytics project. This will be a major challenge for many companies who do not know the capabilities of analytics.

Collection of relevant data

- Once the problem is defined clearly, the project team should identify and collect the relevant data. This may be an interactive process since "relevant data" may not be known in advance in many analytics projects. The existence of ERP systems will be very useful at this stage.

Data Pre-processing

- This would include data imputation and the creation of additional variables such as interaction variables and dummy variables in the case of predictive analytics projects.

Model Building

- Analytics model building is an iterative process that aims to find the best model. Several analytical tools and solution procedures will be used to find the best analytical model in this stage.

Communication of the data analysis

- The communication of the analytics output to the top management and clients plays a crucial role. Innovative data visualization techniques may be used in this stage.

Industry Wide Application of Analytics

Manufacturing

Supply chain analytics

Quality and Process improvement

Revenue and Cost Management

Retail

Assortment Planning

Promotion Planning

Demand Forecasting

Market Basket Analysis

Customer Segmentation

Healthcare

Clinical Care

Hospitality related data

Service

Demand Forecasting

Service Quality Analysis

Customer Segmentation

Promotion

Banking and Finance

Service Demand Analysis

Customer Transaction Analysis

Credit Scoring

IT and ITES (IT enabled services)

Demand for Analytics Services

Software Development Cycle Time

Statistical learning and Machine learning

Useful read: 50 years of Data Science

Statistical Learning vs. Machine Learning

Breiman's 'Two Cultures', 2001

“... Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables x (independent variables) go in one side, and on the other side the response variables y come out. Inside the black box, nature functions to associate the independent variables with the response variables ...”

There are two goals in analyzing the data:

- Inference: to infer how nature is associating the response variable to the input variable
 - Inference based modelling (Generative Modelling), tries to develop model which maximizes the chance of observing the data.
 - Trying to understand how age, gender, past medical history effects the cost of treatment of a disease.
 - By how much cost of treatment will go up for a 35 year old person compared to a 36 year old person.

Statistical Learning vs. Machine Learning

- Prediction: to predict (discriminative modelling) what the responses are going to be to future input variables.
 - Silent about the underlying mechanism generating the data, and allows for many different predictive algorithms, preferring to discuss only accuracy of prediction made by different algorithm on various datasets
 - Using age, gender, past medical history to predict the cost of treatment of a disease. Even if the feature weights does not necessarily help in the right inference.

Users of the data are split into one of the two cultures based on interest and objective they try to achieve with data

What is Machine learning

Common Task Framework is a crucial but unappreciated methodology driving predictive modeling's success

CTF has these ingredients:

- Training data set
- Models whose task is to do class prediction using training data
- Scoring (test set) on which prediction accuracy is reported.

Common Task Framework is the single idea from machine learning and data science that is most lacking attention in statistical training.



Combination of a Predictive Modeling culture together with CTF is the `secret sauce' of machine learning.

Statistics vs. Machine Learning – Debate continues

Machine Learning is a glorified statistics but if all we care about is prediction, why bother using a probability model at all?

Glossary

Machine Learning	Statistics
Network, Graphs	Models
Weights	Parameters, Coefficients
Features	Attributes, Variables
Learning	Fitting
Generalization (Bootstrapping, Cross validation)	Test of Performance (Sampling, Hypothesis testing, p-value)
Supervised Learning	Regression/Classification
Unsupervised Learning	Density Estimation, Clustering
ML sounds like it's young, vibrant, interesting to learn, and growing; Stats does not.	

Blame statistics for not marketing its ideas well enough, or blame CS for ignoring statistics.

Broad Classification of Machine Learning Algorithms

- Supervised Learning
 - Input (X's) and Output (Y) both are known features
- Unsupervised Learning
 - Input (X's) is known but Output (Y) is unknown
- Reinforcement Learning
 - Input (X's) is unknown but Output (Y) is unknown
 - Misspell “avaible” in doc
- Evolutionary Learning
 - evolutionary algorithm (EA) is a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm. An EA uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection.

Data Science and Data Scientists

What is data science

Data Science has six main divisions:

1. Data Exploration and Preparation
2. Data Representation and Transformation
3. Computing with Data
4. **Data Modeling (Statistical learning or Machine learning sits here).**
5. Data Visualization and Presentation
6. Science about Data Science



<http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf>

Data Exploration and Preparation

- Exploration of data for sanity checks
 - histograms, scatterplots, time series plots
- Datasets have anomalies:
 - Reformatting, recoding, grouping, imputations

Exploration and preparation where 80% of the effort goes

Data Representation and Transformation

- Varied sources of data:
 - text files, spreadsheets to SQL and noSQL databases, distributed databases and live data streams
- Different formats:
 - Acoustic, image, sensor, video data

Need to know the structures, transformations, and algorithms involved in dealing with the above representations.

Computing with Data

- Several languages for data analysis and processing
 - R, Python, specific languages for text manipulation
- Understanding the computational efficiency to effectively use the languages.
- Develop packages which automates commonly used workflows for data analysis in future.

Data Visualization and Presentation

- Dashboards for monitoring data processing pipelines that access streaming or widely distributed data.
- Developing visualizations to present conclusions from a modeling exercise or CTF challenge.

Presentation data which is visually appealing

Data Modelling

- Uses tools and viewpoints from both of Leo Breiman's modeling cultures:
 - Generative modeling, in which one proposes inference based modelling
 - Predictive modeling, in which one constructs methods which predict well over some given data universe i.e. some very specific concrete dataset.

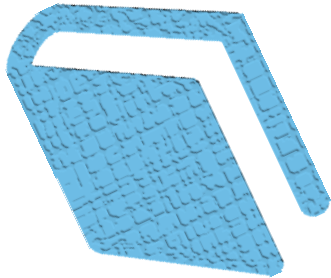
Science of data science

- Applying the tools in conjunction with proper business context and judge the outcomes in terms of effectiveness in cutting human time, computing resources, other performance KPIs.
- Identifying commonly occurring analysis workflows in specific domains.

True effectiveness of a tool is related to the probability of deployment times the probability of effective results once deployed.

Summary

Summary of the topics covered in this lesson:



- With the data explosion across industry, the usage of analytics in decision making will become the most critical factor for being competitive in business.
- Descriptive analytics becomes the stepping stone to all the complex problems which can be solved using analytics.

End of Lesson–Introduction to Business Analytics

