



MIS 64036: Business Analytics

Lecture I

Rouzbeh Razavi, PhD

Let's Get Introduced & Break Some Ice!



Agenda

- Introduction to the Course
- Definition and History of Data Analytics
- Success Stories and Business Examples
- Data Analytics Lifecycle
- R Programming

Agenda

- **Introduction to the Course**
- Definition and History of Data Analytics
- Success Stories and Business Examples
- Data Analytics Lifecycle
- R Programming

Topics Covered

**Introduction
to Analytics**

**Analytics for
Managers**

**Descriptive
Analytics**

**Data
Preparation**

Predictive/Prescriptive Analytics

**Data Visualization
Communication**

**Formulating
Business
Problems**

**Models in
Operation**

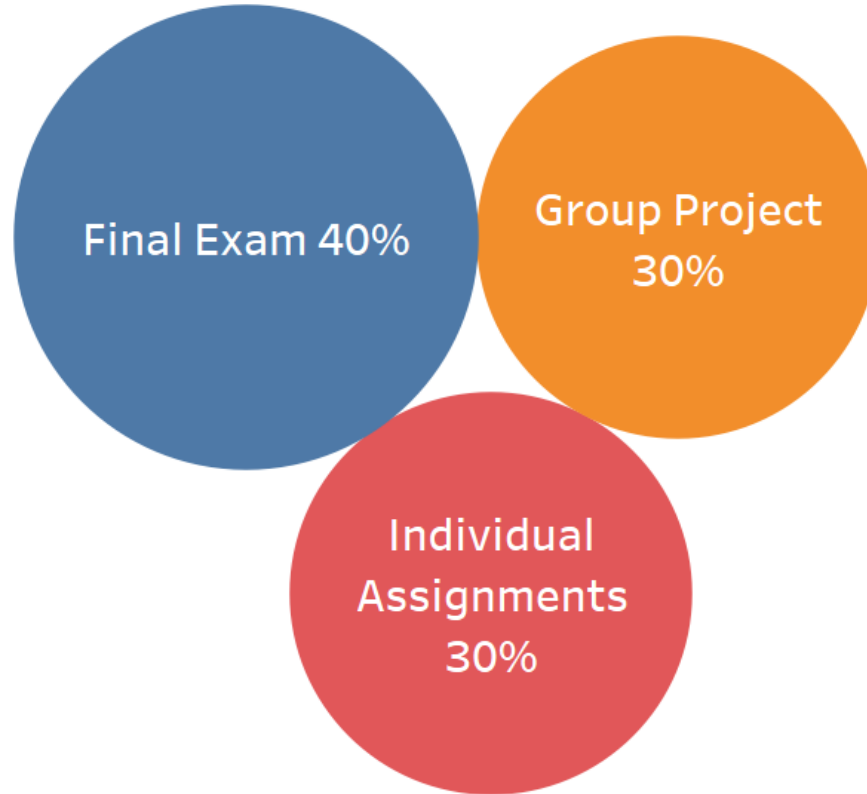
Course Timeline

Date	Topic Description
28 August	Introduction to Big Data and Business Analytics
4 September	Labor Day: No Class
11 September	Role of data analytics in strategic management (cont.)
18 September	Descriptive Analytics: key statistics and distribution
25 September	Descriptive Analytics (cont.) + Intro to data visualization
2 October	Preparing data for modelling
9 October	Predictive/Prescriptive Analytics: Segmentation
16 October	Predictive/Prescriptive Analytics: Recommendation Systems
⋮	⋮

Course Timeline II

Date	Topic Description
23 October	Predictive/Prescriptive Analytics: Classification
30 October	Predictive/Prescriptive Analytics: Regression
6 November	Predictive/Prescriptive Analytics: Time Series and Simulations
13 November	Data Visualization to Communicate Information
20 November	Formulating and solving business problems using analytics
27 November	Operationalizing analytical models
4 December	Wrap up, review / presentations

Assessment



Agenda

- Introduction to the Course
- **Definition and History of Data Analytics**
- Success Stories and Business Examples
- R Programming
- Data Analytics Lifecycle

What is Data Analytics? Few Definitions.

- **Data analytics** refers to qualitative and quantitative techniques and processes used to enhance productivity and business gain using data.
- **Data analytics** is the pursuit of extracting meaning from raw data using specialized computer systems. These systems transform, organize, and model the data to draw conclusions and identify patterns.
- **Data analytics** is an examination of data made to gain a better understanding of the data itself and the organization that produced it.

History of Data Analytics

1935

Ronald Fisher

Design of Experiment



1939

Edward Deming

Quality Control



1958

Peter Luhn

A Bussiness Intelegence System



1977

John W. Tukey

Exploratory Data Analysis



1989

Howard Dresner

Bussiness Intelegence



1997

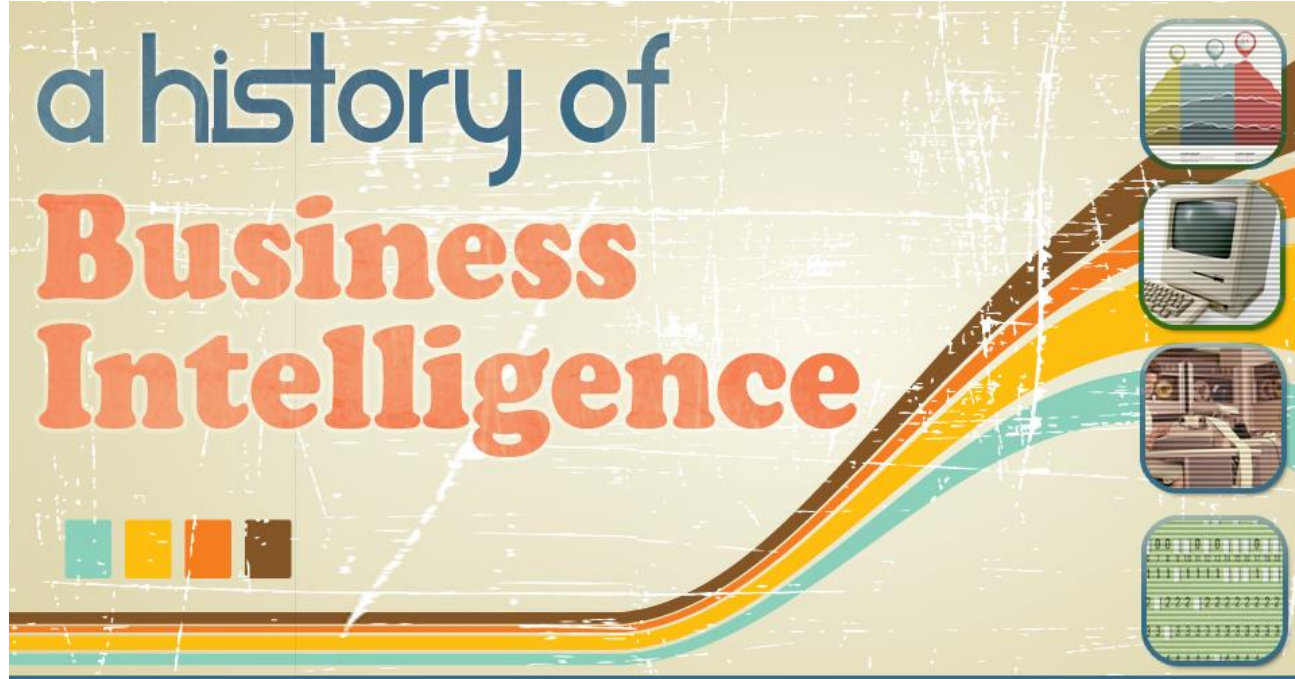
Tom M. Mitchell

Machine Learning



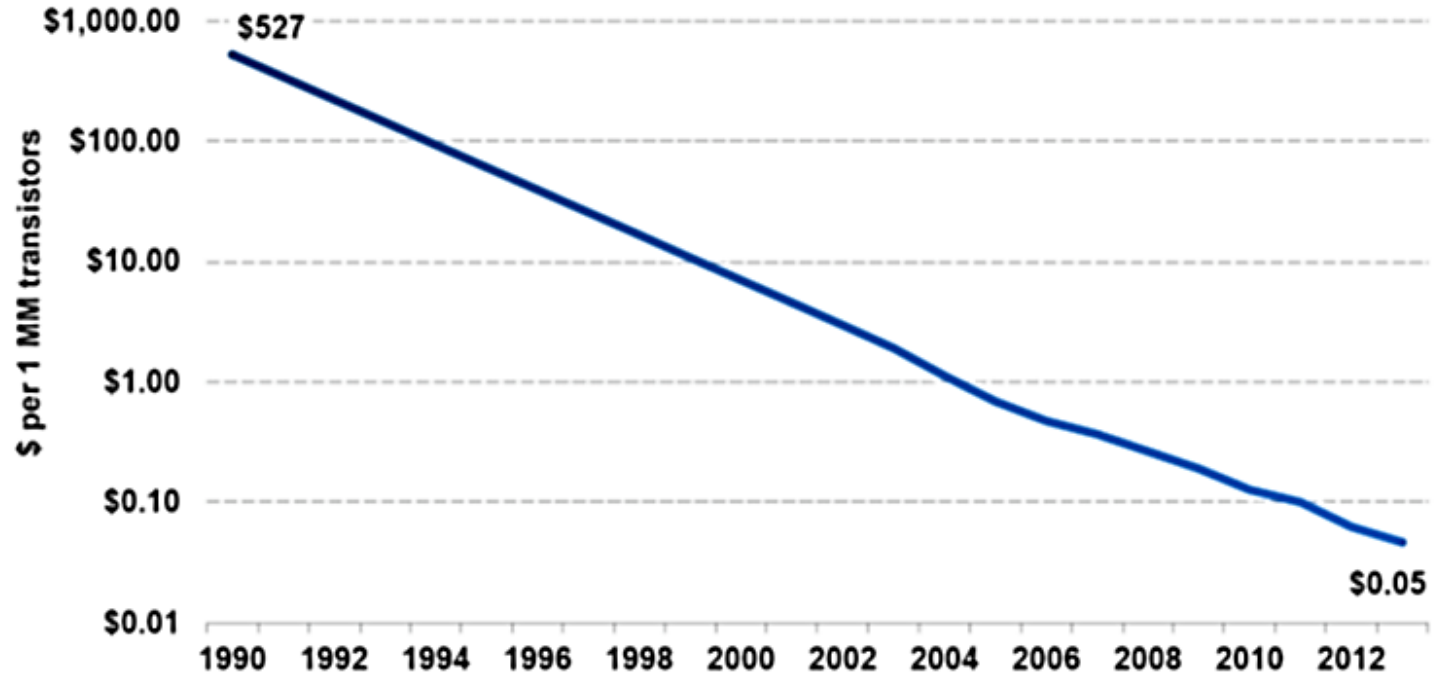
History of Data Analytics

[Click to Watch the Video](#)



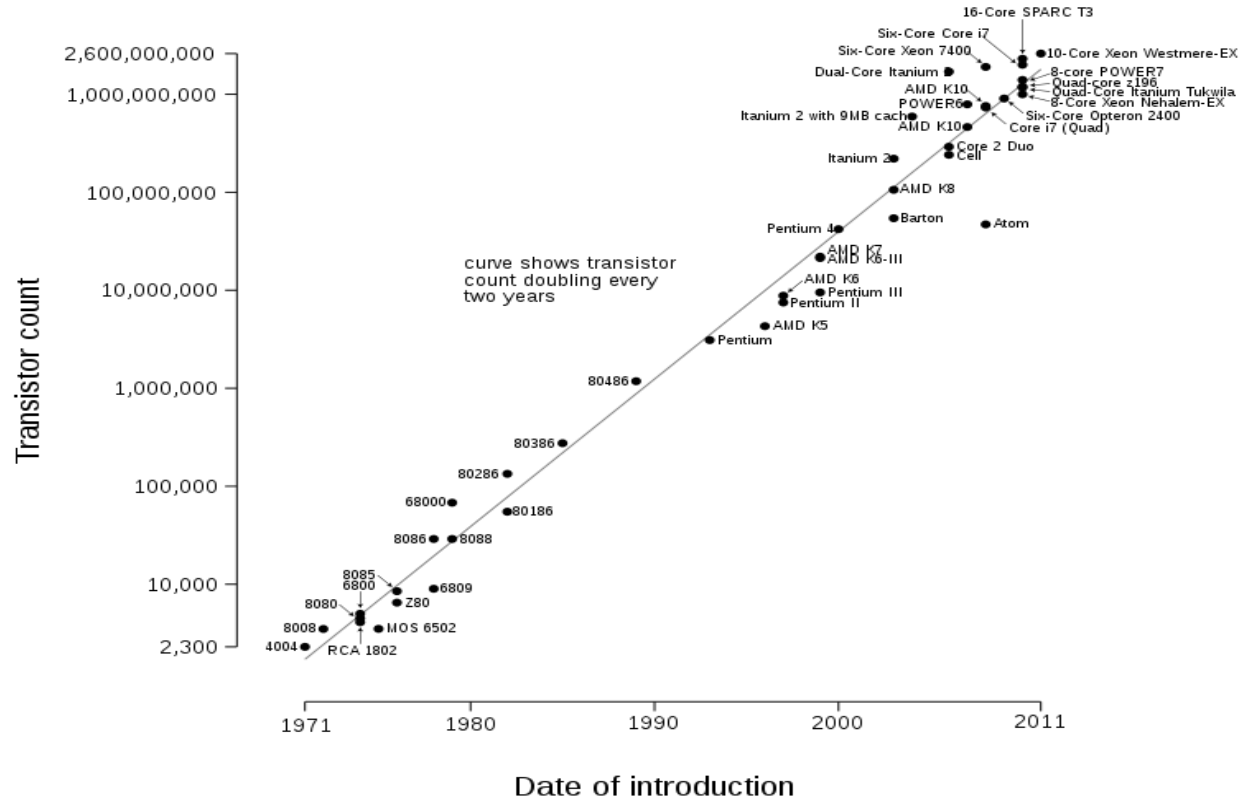
What Has Changed Now? Computing Cost

Global Compute Cost Trends

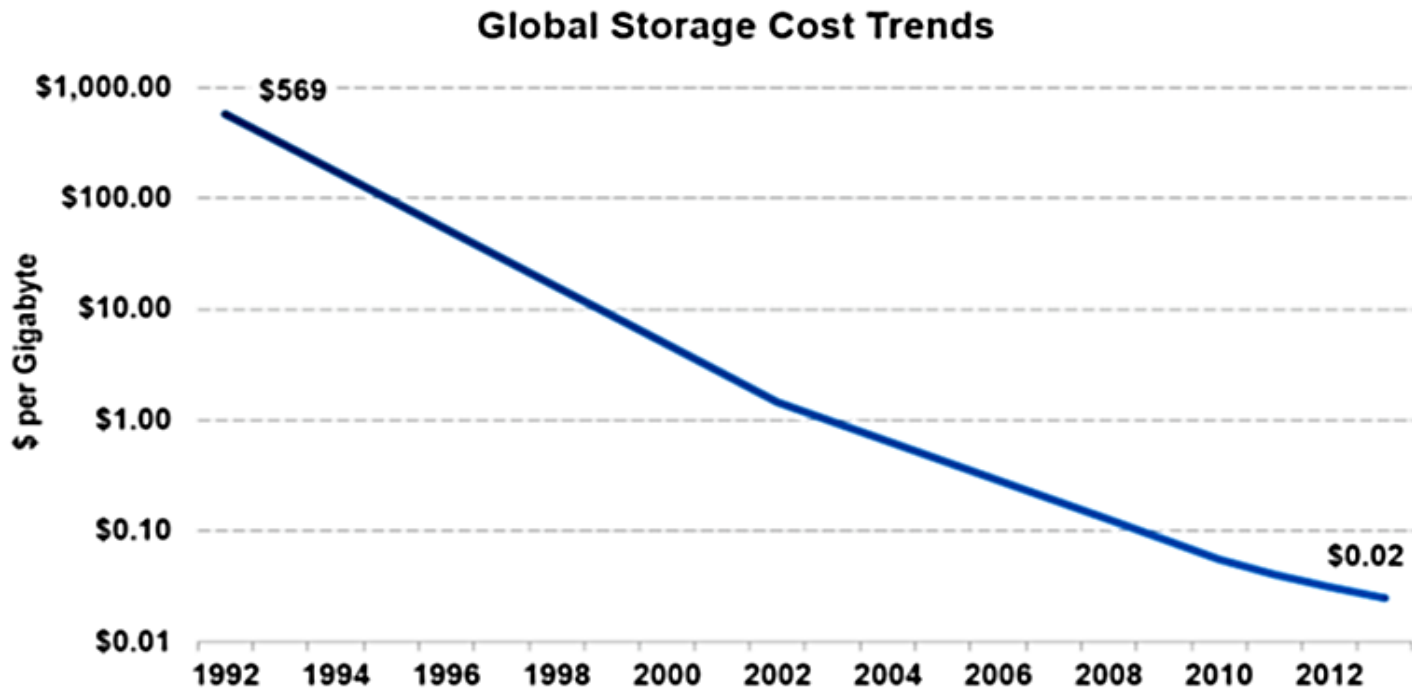


*Source: KPBC report 2013

What Has Changed Now? Computing Capability



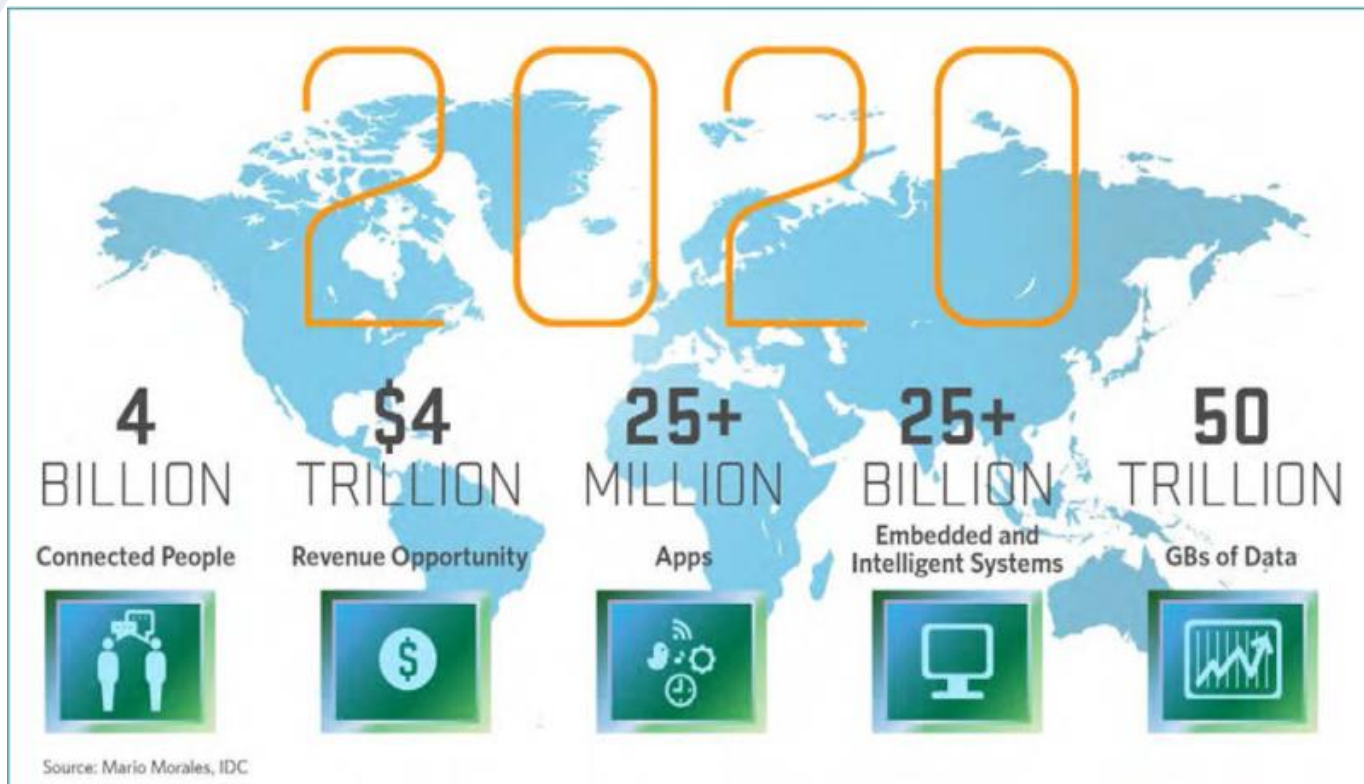
What Has Changed Now? Storage Cost



...Storage Costs Declining = 38% Annually, 1992-2013...

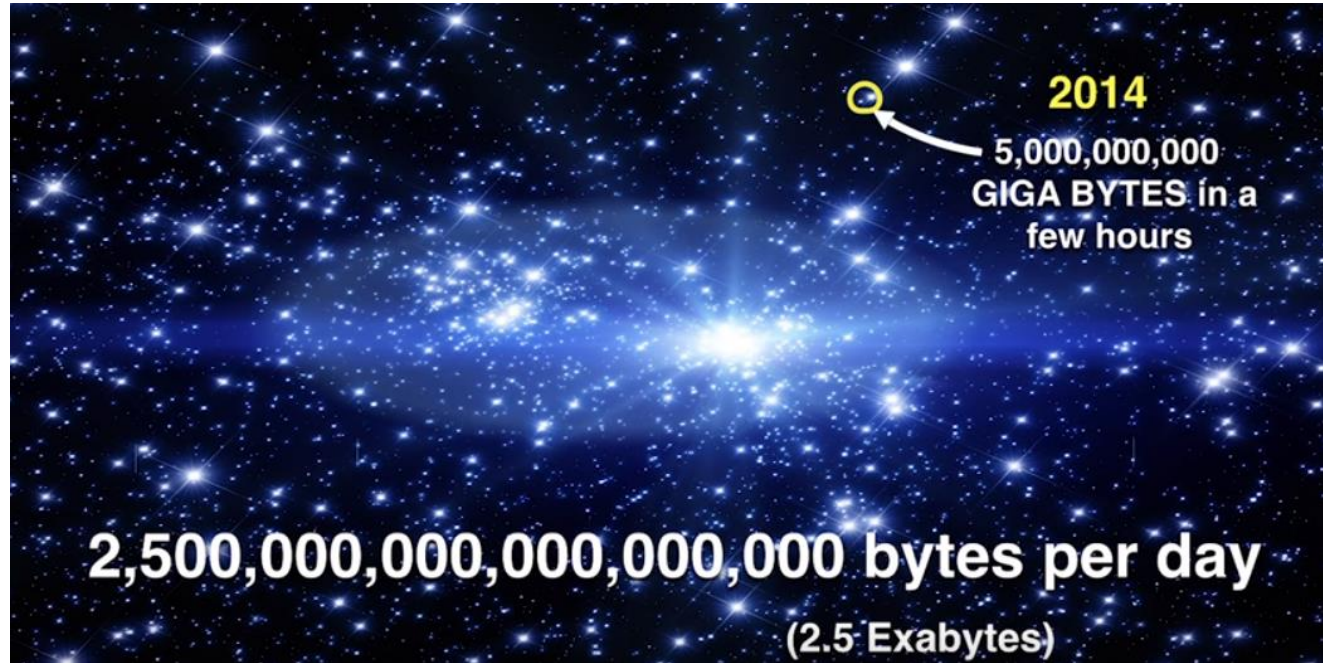
*Source: KPBC report 2013

What Has Changed Now? Available Data Volume

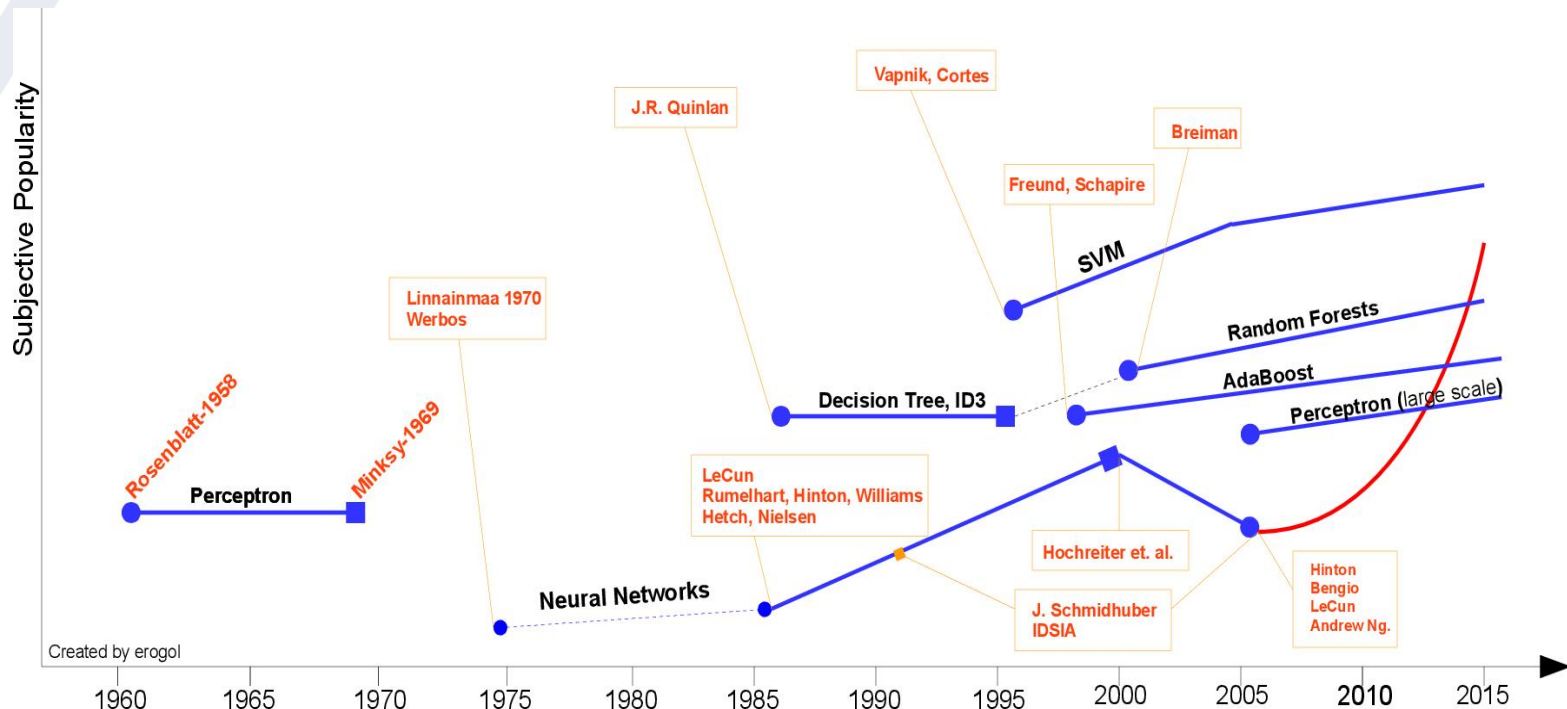


What Has Changed Now? Available Data Volume

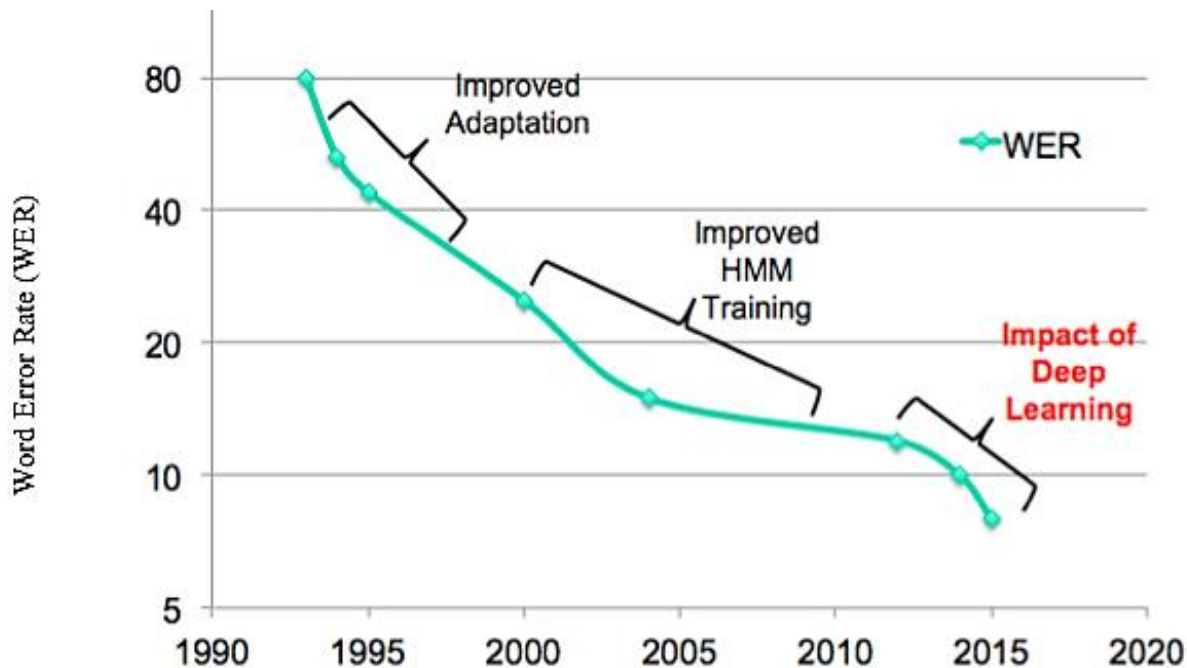
[Click to Watch the Video](#)



What Has Changed Now? New Algorithms



What Has Changed Now? Algorithms' Performance



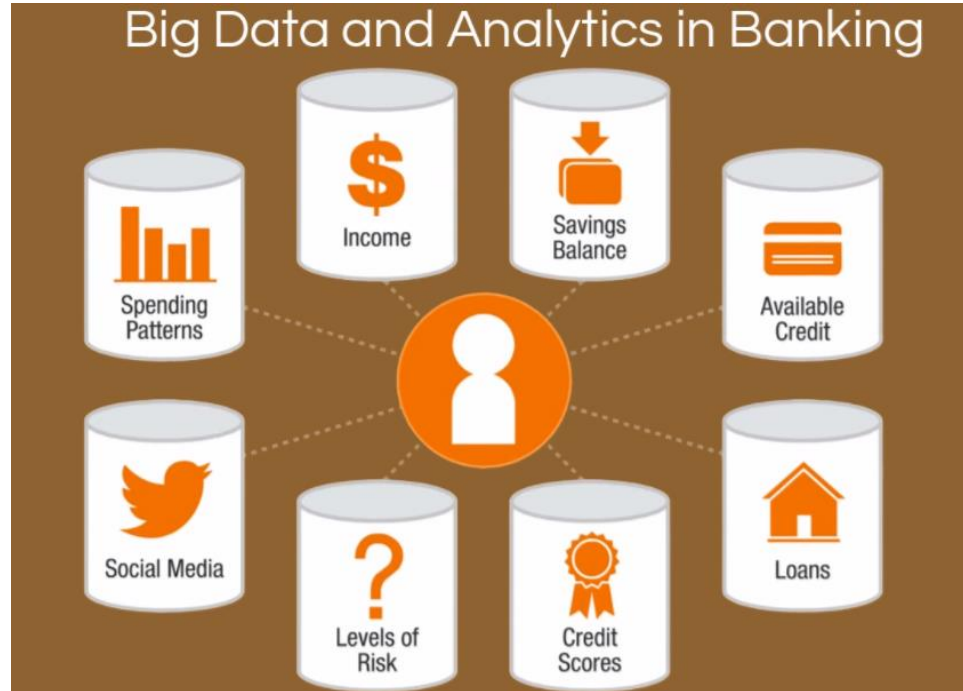
Example: Word Error Rate in Automatic Speech Recognition (ASR)

Agenda

- Introduction to the Course
- Definition and History of Data Analytics
- **Success Stories and Business Examples**
- Data Analytics Lifecycle
- R Programming

Data Analytics: Use Cases in Banking

[Click to Watch the Video](#)



Data Analytics: Mastercard Story

[Click to Watch the Video](#)

MasterCard Big Data



2.2B+

Global CARDS

160MM

TRANSACTIONS
PER HOUR

52Billion

(annual transactions)



Card Swipes

Amount, spent, time, merchant & location.



Data Anonymized

Intelligence



Analysis | Risk Detection | Customer 360 | Location selection | Customer Engagement | Economic Indicators



Data Analytics: More Examples

[Click to Watch the Video](#)



Agenda

- Introduction to the Course
- Definition and History of Data Analytics
- Success Stories and Business Examples
- **Data Analytics Lifecycle**
- R Programming

Key Roles in Data Analytics Projects



Project Sponsor: Responsible for the genesis of the project, provides impetus for the project and usually provides funding as well.

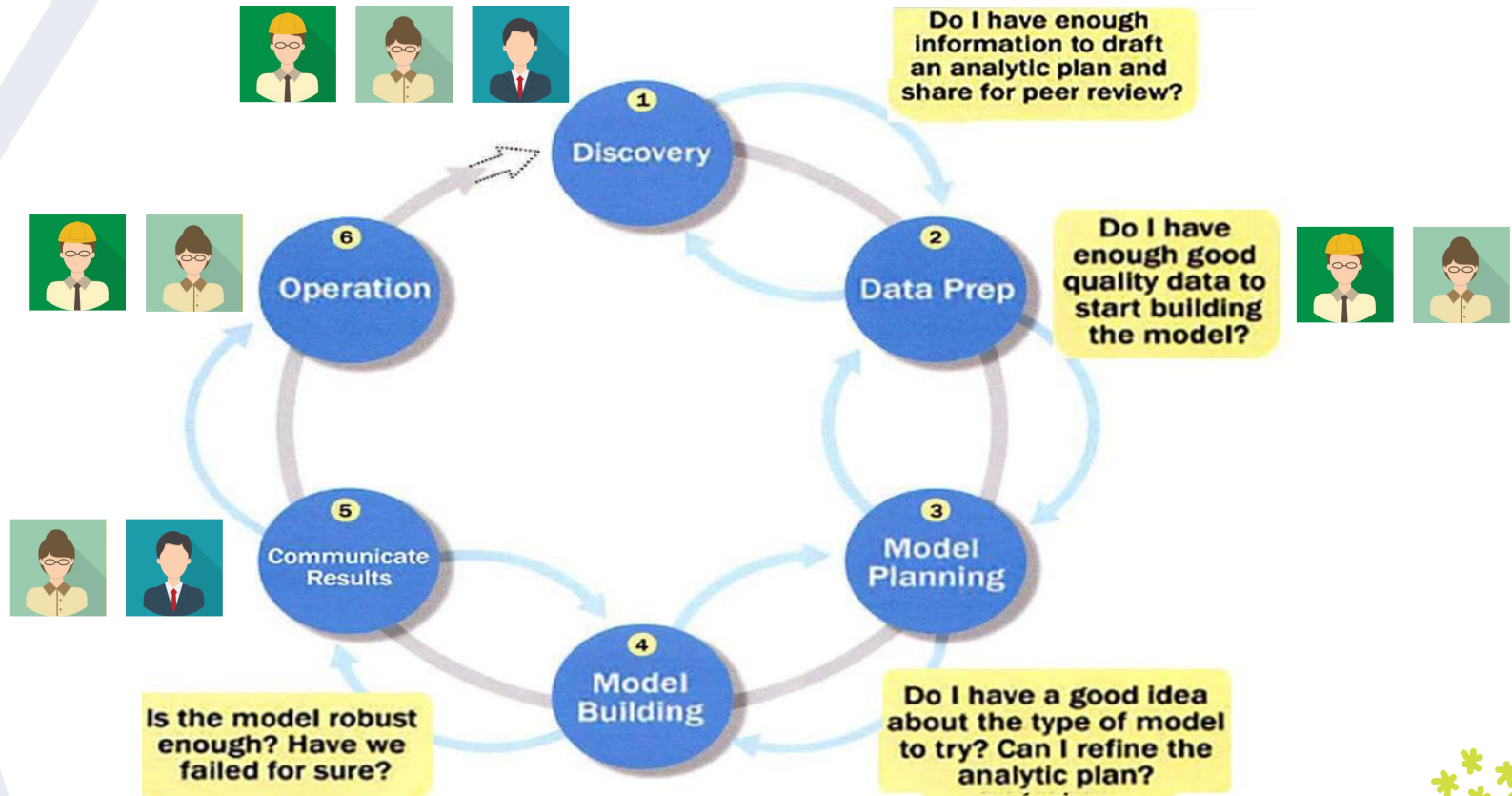


Business User/Analyst: Has a deep understanding of the business domain and is likely to use the end result of the project once implemented.

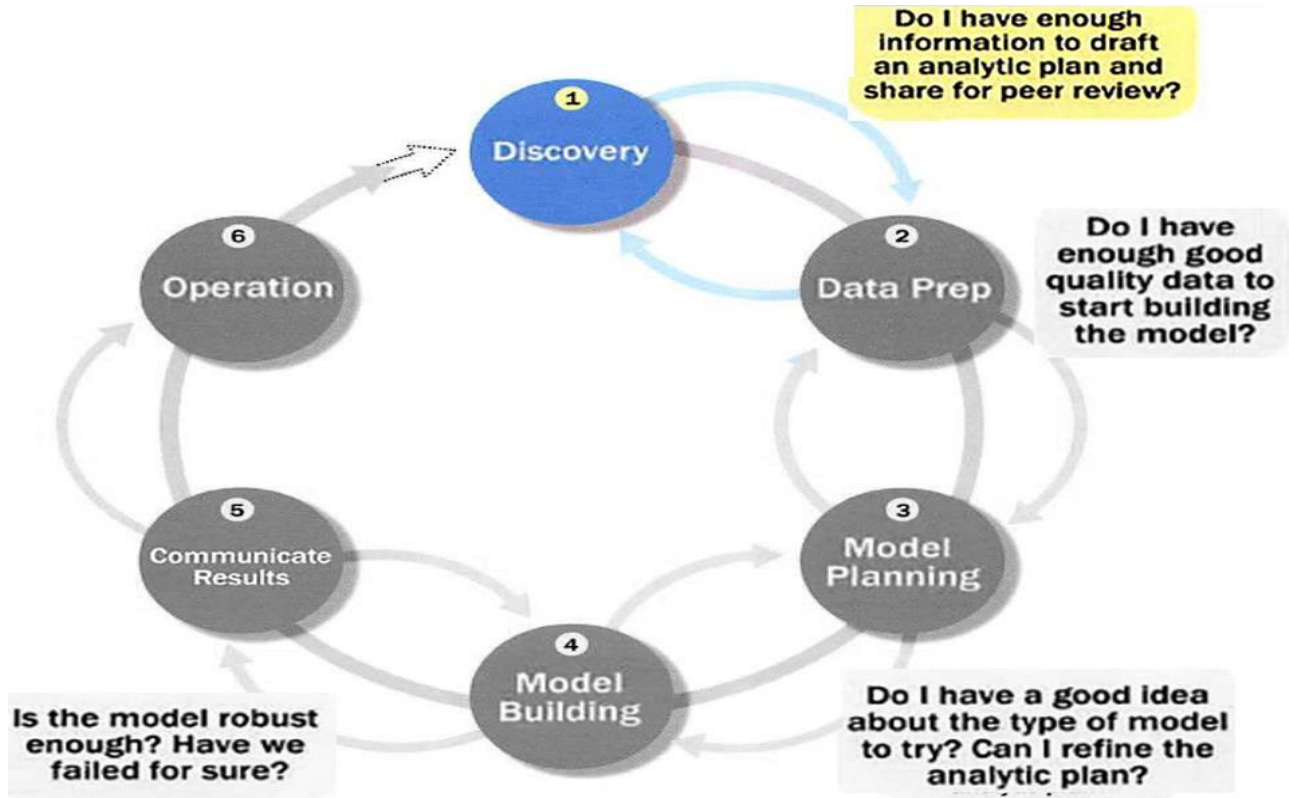


Data Engineer: Deep technical skills to work with databases and can assist providing access to right data sources.

Data Analytics Lifecycle: Overview



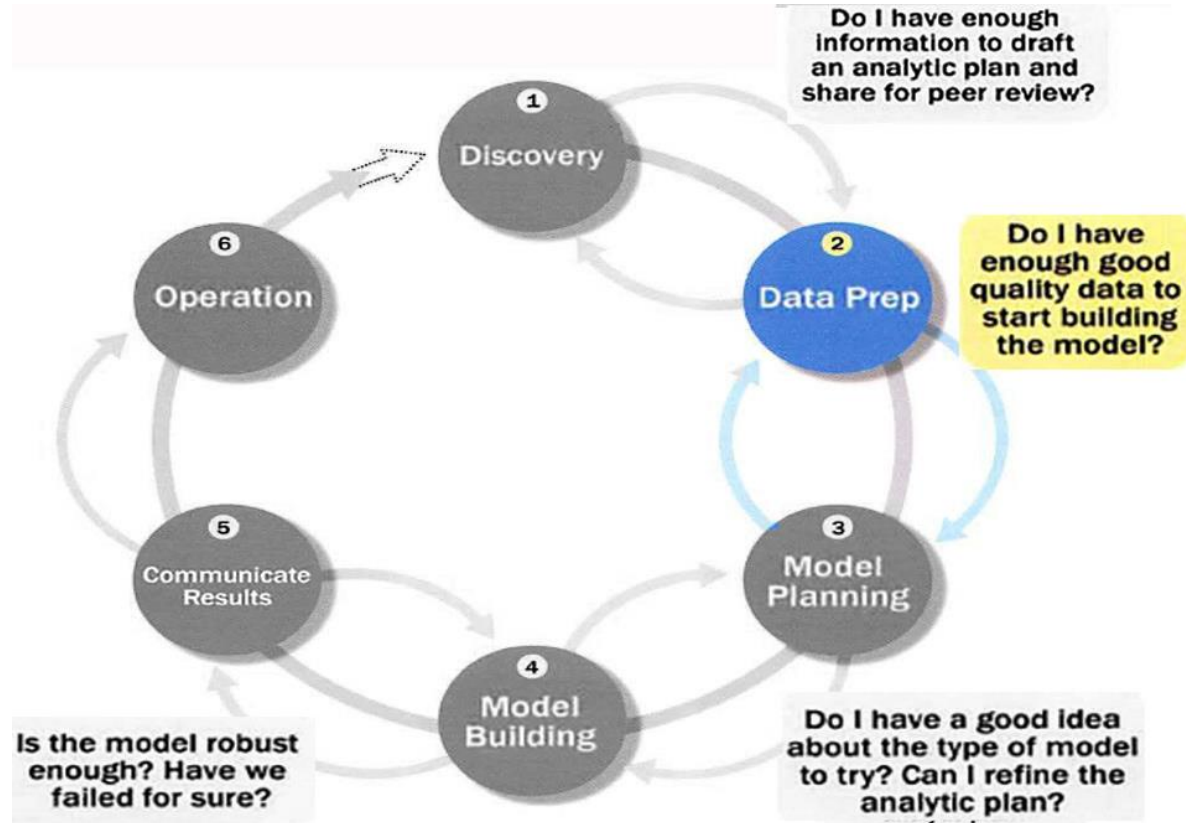
Data Analytics Lifecycle: Discovery



Data Analytics Lifecycle: Discovery

- Understand and formulate the business problem
- Articulate the pain points as clearly as possible so you can make sure to address them
- Define and formalize the scope of the project
- Identify the kinds of data you will need to solve the problem
- Determine the resources that you may need during the project (Infrastructure, Software, Data access, Business experts, etc.)

Data Analytics Lifecycle: Data Preparation



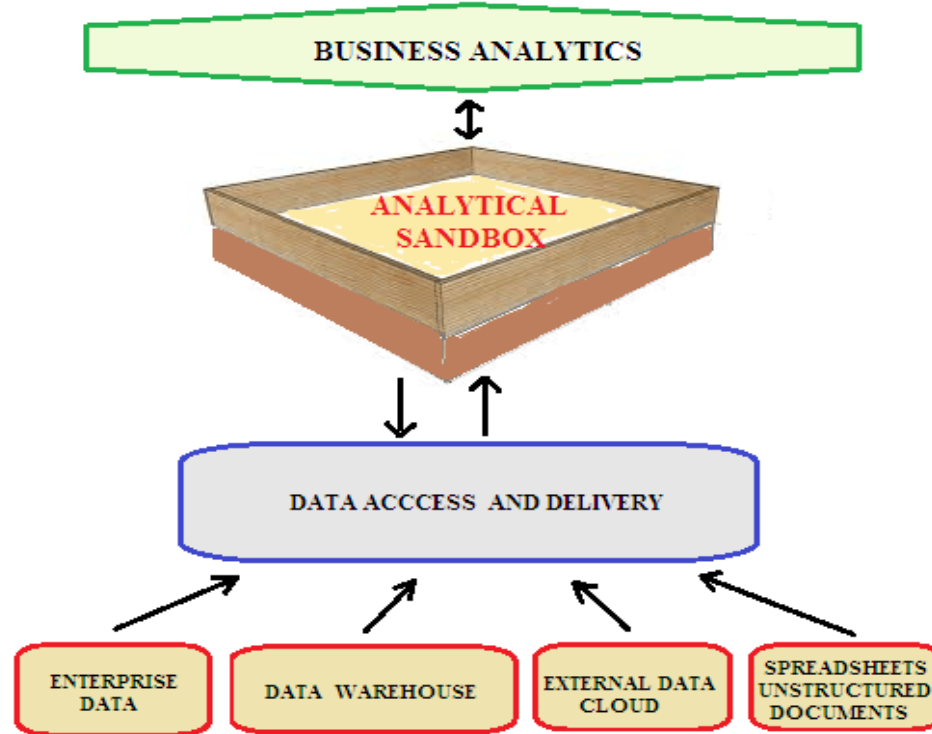
Data Analytics Lifecycle: Data Preparation

“By failing to prepare, you are preparing to fail.”

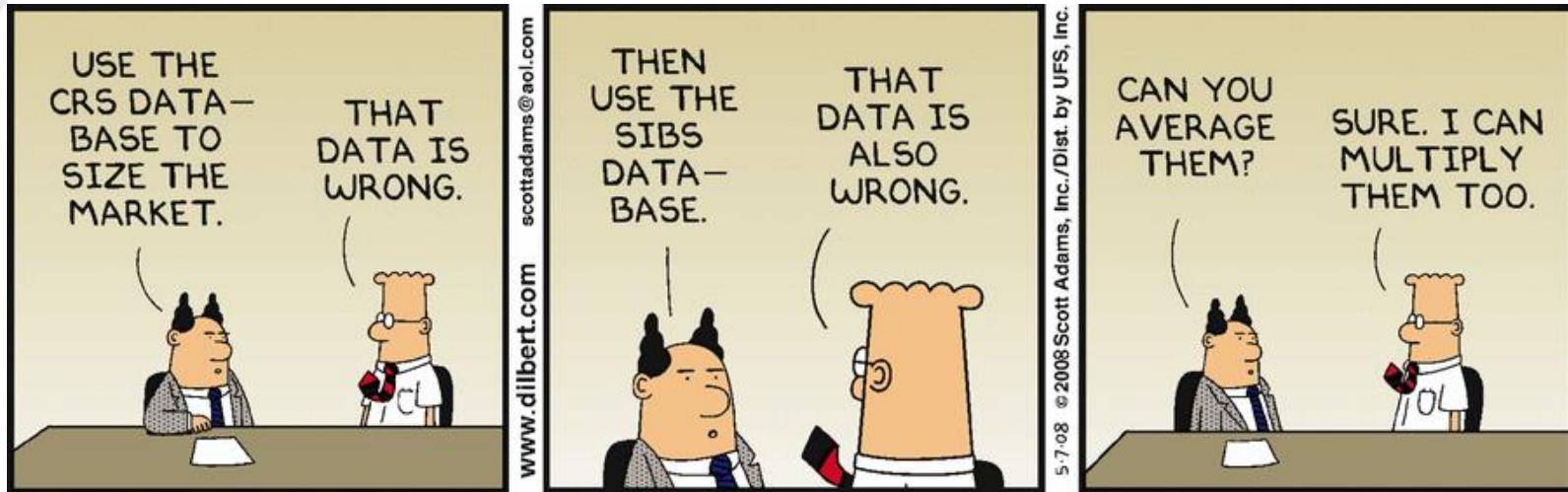
Benjamin Franklin

- Of all of the phases, the step of data preparation is generally the most iterative and time intensive.
- Prepare an analytic sandbox, in which you can work for the duration of the project.
- Extract data from different platforms into your analytics sandbox. You should be collecting all kinds of data that is available.
- Clean the data and apply relevant transformation

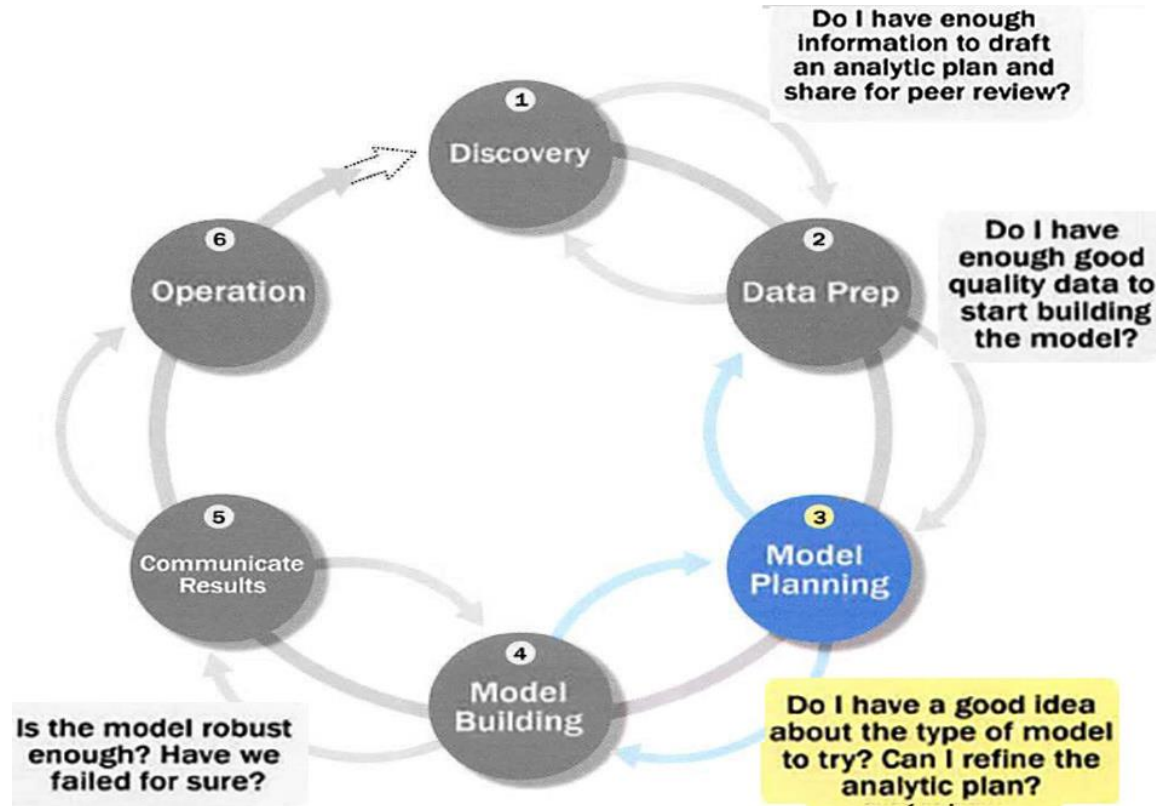
Data Analytics Lifecycle: Analytic Sandbox



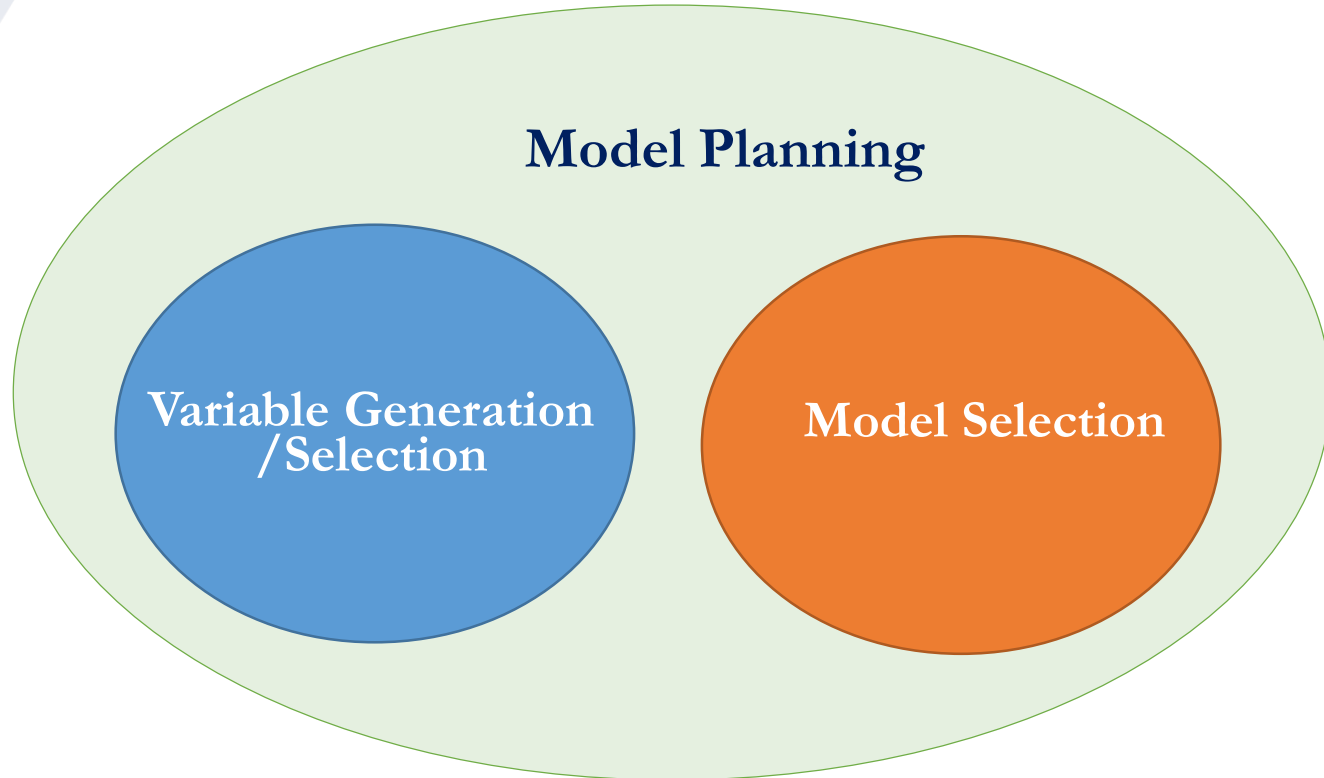
Data Analytics Lifecycle: Data Quality is Key!



Data Analytics Lifecycle: Model Planning



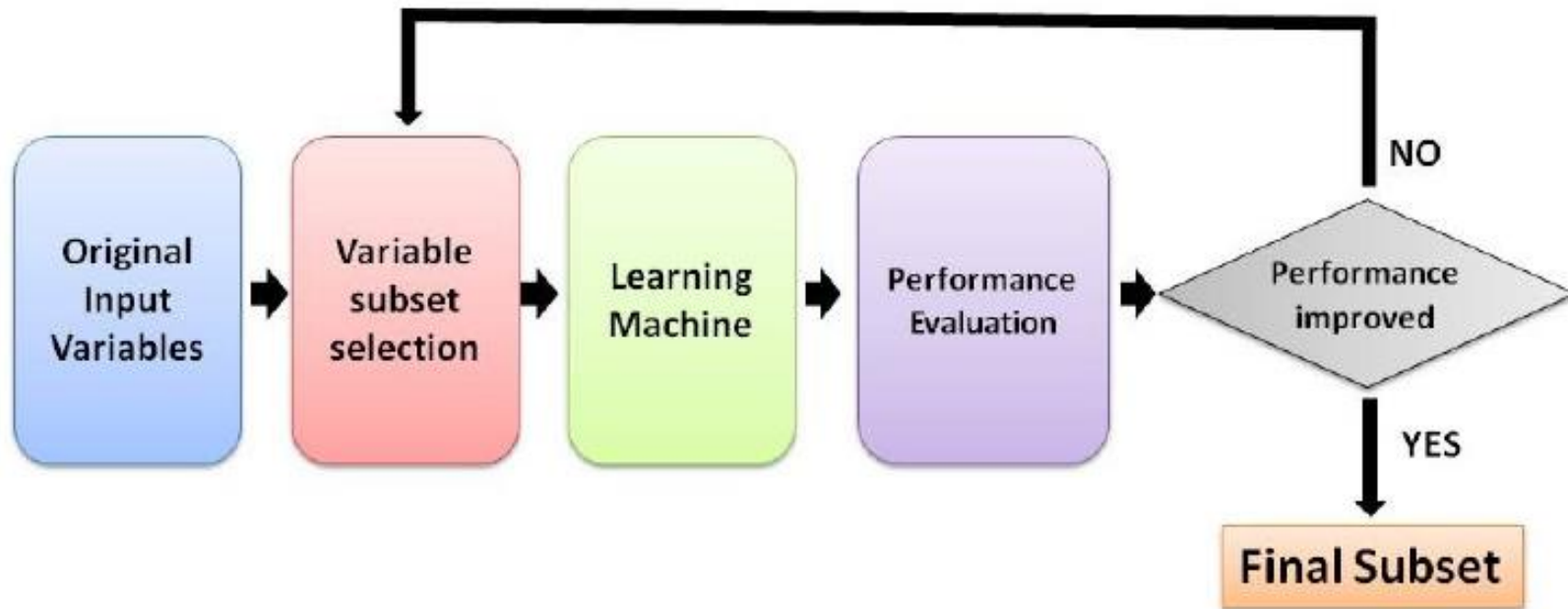
Data Analytics Lifecycle: Model Planning



Model Planning: Variable Generation/Selection

- What is it? Generate new variables by modifying/combining existing variables, as needed, and then select a subset of relevant variables for model building.
- Why Variable Selection/Filtering?
 - Improved model building process and computation speed
 - Improved model's explainability
 - Cost considerations (e.g. may need to pay for external data sources)
 - Legal and policy considerations (e.g. using demographic information for credit scoring)
 - Improved model accuracy (in some cases)

Model Planning: Variable Selection



Data Analytics Lifecycle: Data Selection!

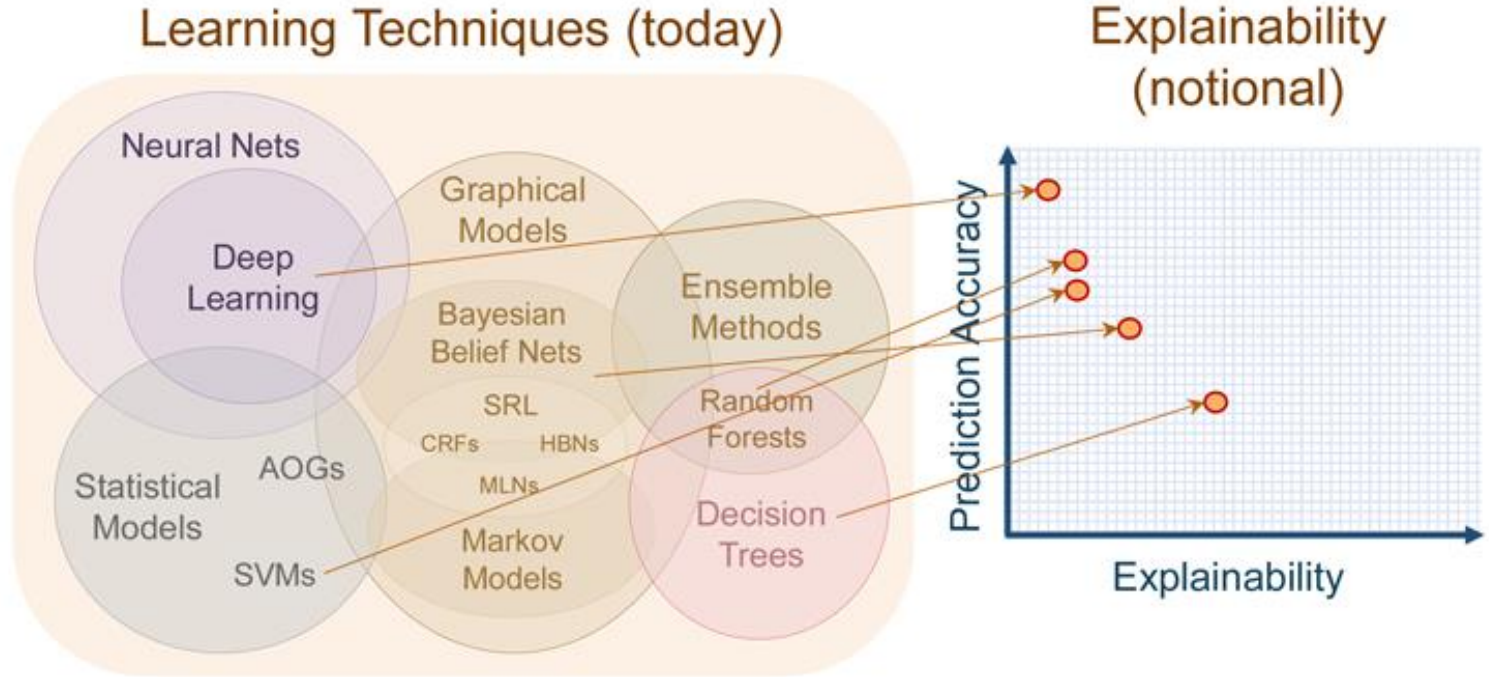


"You can't keep adjusting the data to prove that you would be the best Valentine's date for Scarlett Johansson."

Model Planning: Model Selection

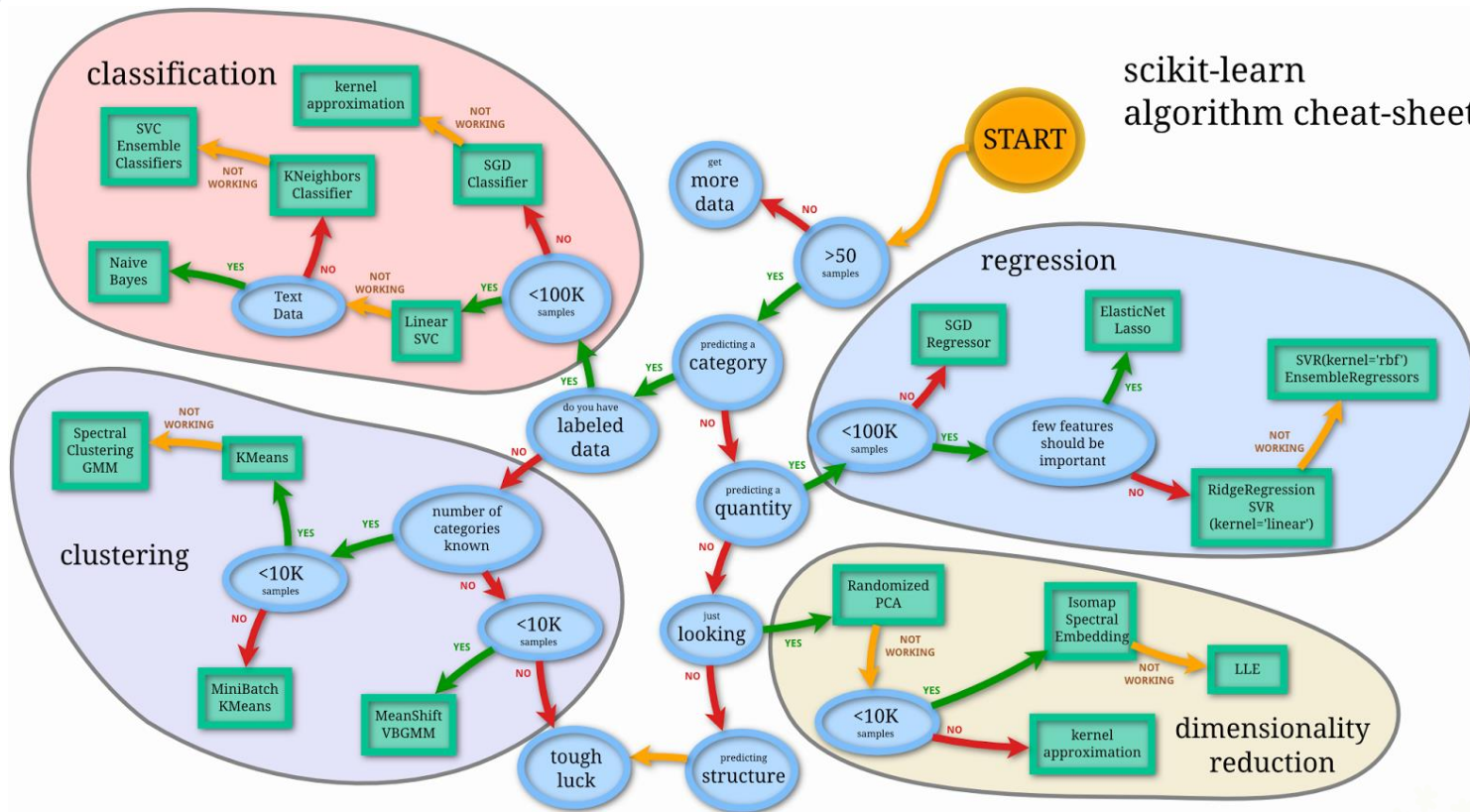
- What it is? There exist a large number of analytics algorithm. Choosing the right algorithm for the given problem is key.
- What to Consider?
 - Feasibility of applying the algorithm
 - Accuracy of the algorithm
 - Explainability of the algorithm
 - Computational complexity of the algorithm
- Recommendation
 - Always start with simpler algorithms (e.g. linear models)

Model Planning: Model Selection

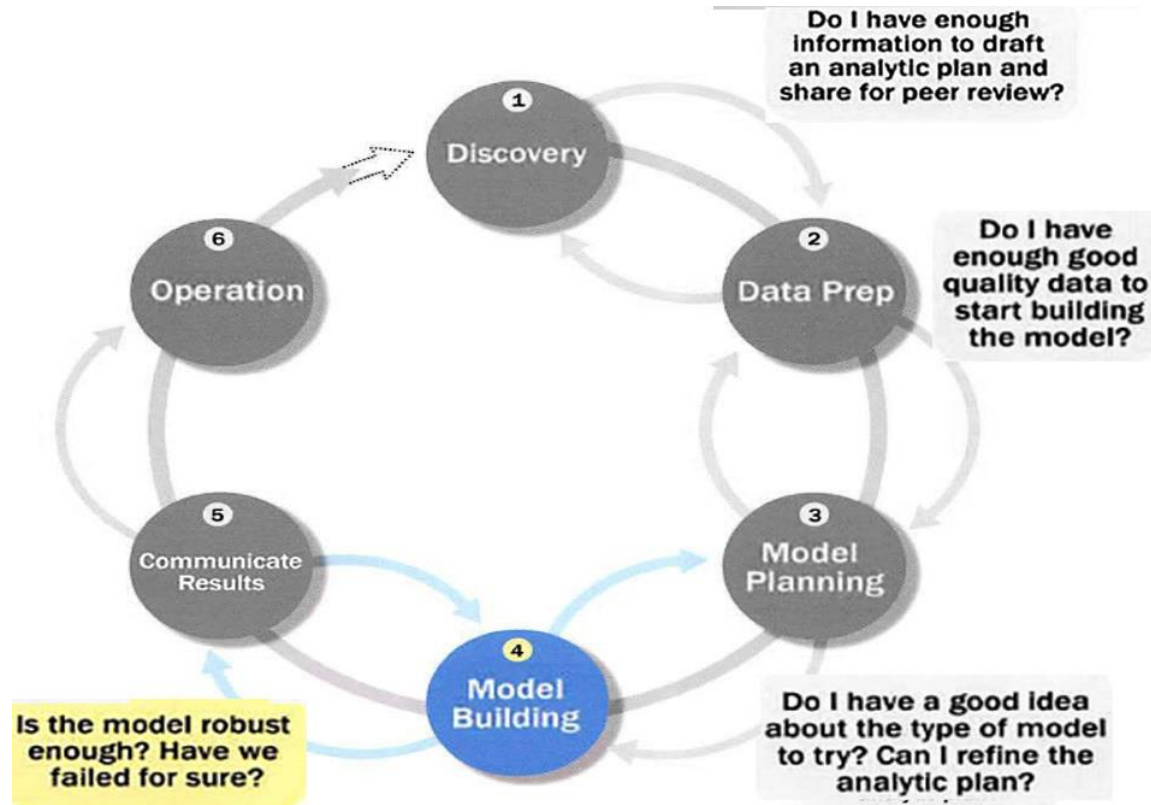


Model Planning: Model Selection

scikit-learn
algorithm cheat-sheet



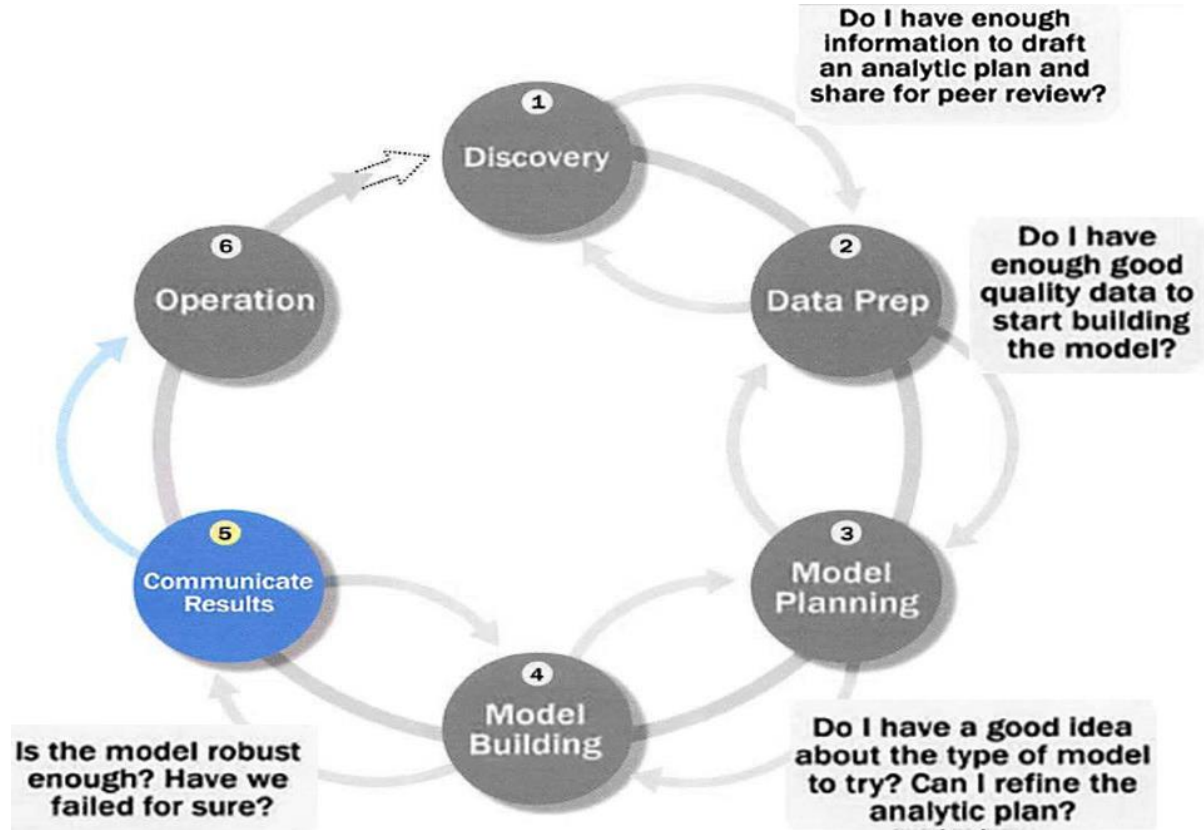
Data Analytics Lifecycle: Model Building



Data Analytics Lifecycle: Model Building

- Implement the model using the appropriate programming language/tool (R, Python, SAS, SPSS, etc.)
- Fine tune the models to optimize the results
- Construct a list of key variables (variable importance scores)
- Assess the validity of the model and its results
- Record the results, and logic of the model

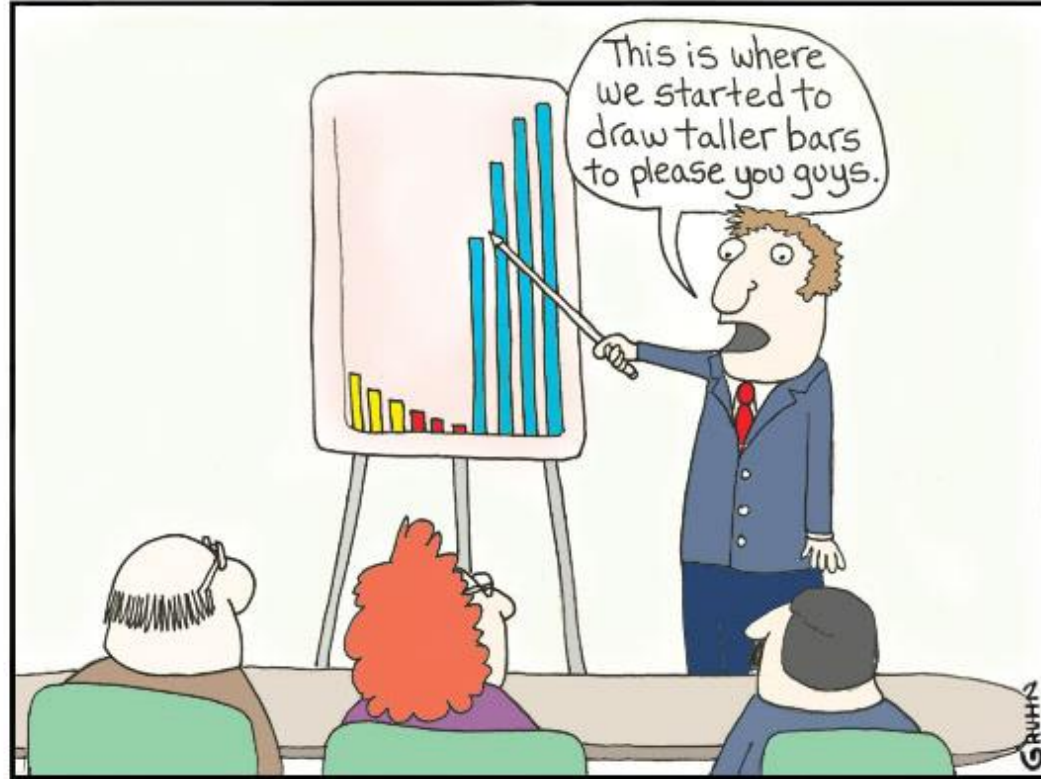
Data Analytics Lifecycle: Communication



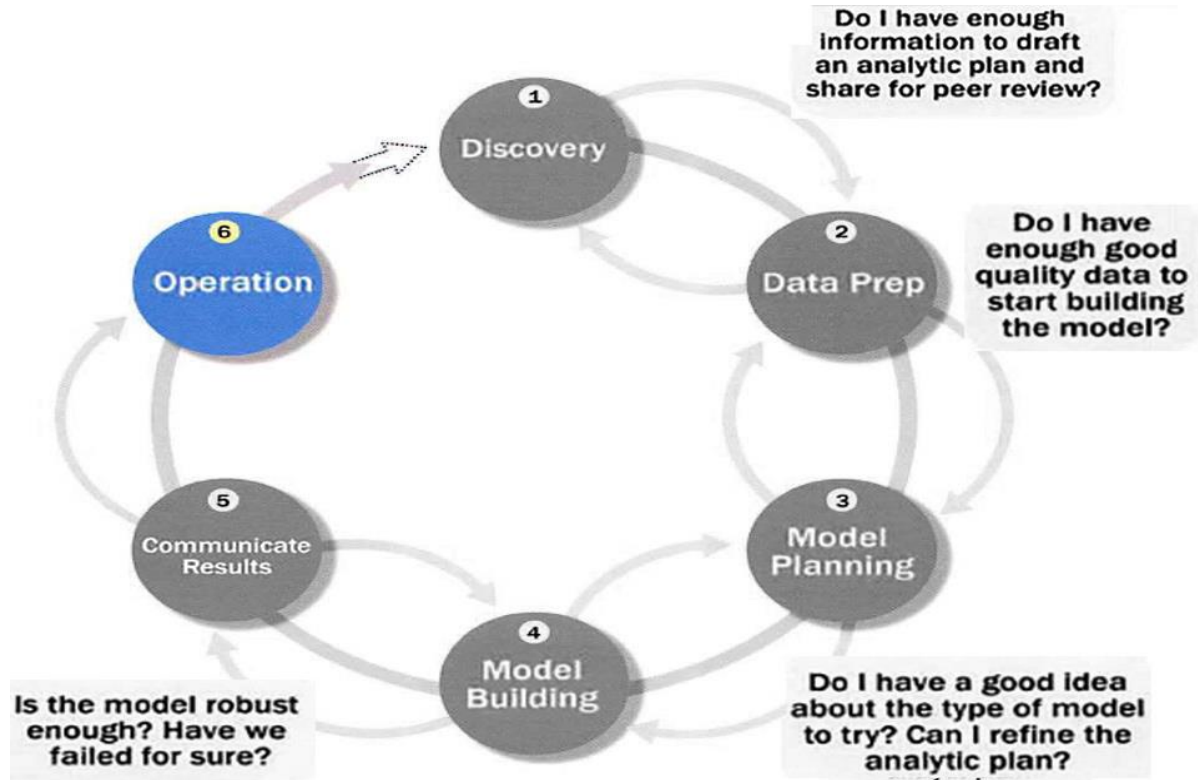
Data Analytics Lifecycle: Communicating Results

- Clearly articulate the methodology at an appropriate technical level according to the audience
- Interpret results
- Explain key findings
- Quantify business value
- Make recommendations for future work or improvements to existing processes

Data Analytics Lifecycle: Communicating Results



Data Analytics Lifecycle: Operation



Data Analytics Lifecycle: Operationalize the model

- Deliver technical documentation of the model and data
- Provide consultancy and support for best way to operationalize the model
- Define test cases and assure correct procedures are followed for operationalizing the model
- Define the processes to retain and update the models.

Agenda

- Introduction to the Course
- Definition and History of Data Analytics
- Success Stories and Business Examples
- Data Analytics Lifecycle
- **R Programming**

R Programming

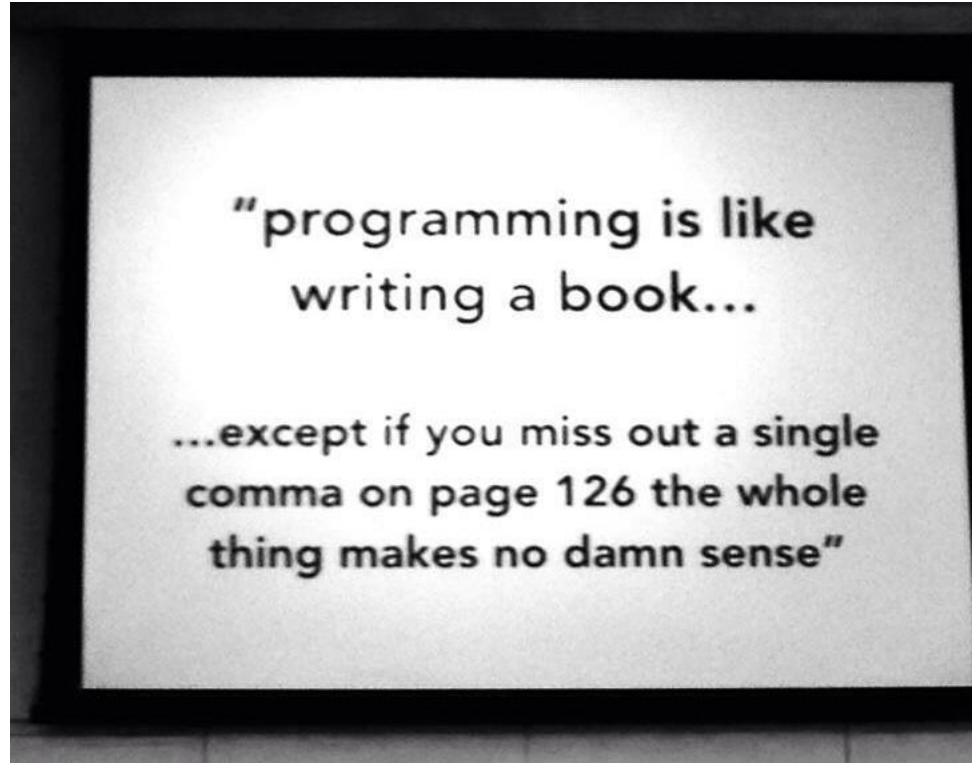
- Will use R throughout this course.
- R is an open source programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing.
- The R language is rich, powerful and widely used among statisticians and data miners for developing statistical software and data analysis.



R or Python?



R: Easy Programming Language But Still ...



R: Where to Start?

- Instructions to install R and R Studio:
- <https://www.andrewheiss.com/blog/2012/04/17/install-r-rstudio-r-commander-windows-osx/> (Windows and Mac) or
- <https://medium.com/@GalarnykMichael/install-r-and-rstudio-on-windows-5f503f708027> (Windows Only)
- Few tips and commands to get you started
<http://web.cs.ucla.edu/~gulzar/rstudio/basic-tutorial.html>
- I have uploaded three mini-guides to R on the course website under Course Material > Learning R
- Make sure you have R and R Studio installed on your laptops for the next lecture.

What We Have Covered Today

- Introduction to the Course
- Definition and History of Data Analytics
- Success Stories and Business Examples
- Data Analytics Lifecycle
- R Programming

ANY
QUESTIONS
?