Lecture 1

Overview of Business Analytics Project Lifecycle

BT4301 – Business Analytics Solutions Development and Deployment

AY 2023/24 Semester 2

Lecturer: A/P TAN Wee Kek

Email: tanwk@comp.nus.edu.sg :: **Tel**: 6516 6731 :: **Office**: COM3-02-35

Consultation: Wednesday, 9:30 pm to 10:30 pm. Additional consultations by appointment are welcome.

Learning Objectives

- At the end of this lecture, you should understand:
  - The business analytics project lifecycle.
  - Traditional lifecycle frameworks.
  - More modern lifecycle frameworks.
  - Agile lifecycle frameworks.
  - Issues beyond agility that are pertinent to the business analytics project lifecycle.

Readings

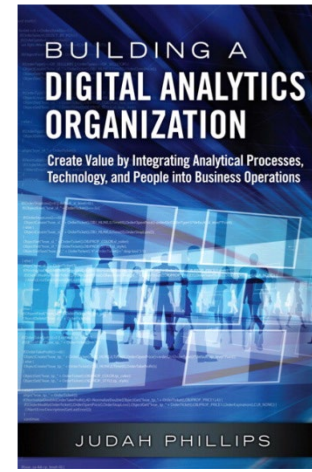
- Reference Book 1:

  Building a Digital Analytics Organization: Create Value by Integrating Analytical Processes, Technology, and People into Business Operations

  Author: Judah Phillips

  1st Edition (2013), Pearson

  https://linc.nus.edu.sg/record=b3796367



Readings (cont.)

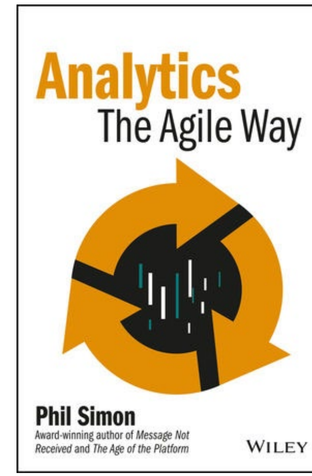
- Reference Book 2:

  Analytics: The Agile Way

  Author: Phil Simon

  1st Edition (2017), Wiley

  https://linc.nus.edu.sg/record=b3751448

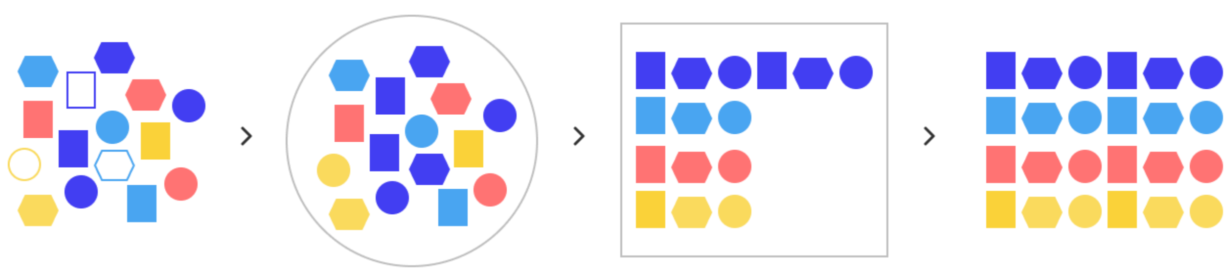


Readings

- Required readings:
  - Reference Book 1 – Chapter 2
  - Reference Book 2 – Chapter 4
- Suggested readings:
  - None.

What is Business Analytics?

- **Business analytics** refers to the process of using <u>raw data</u> to derive <u>valuable insights</u> and increase understanding of a business phenomenon:
  - Analyse historical and current events to identify potential trends.
  - Make optimal or at least better-informed decisions.
- More specifically, business analytics attempts to discover <u>meaningful patterns</u> in data:
  - The patterns are then applied to solve real-world problems.



What is Business Analytics? (cont.)

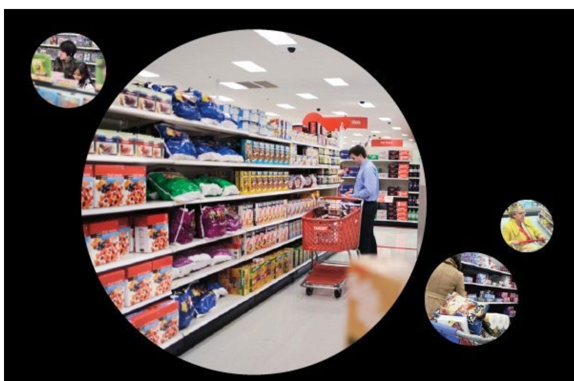
- Example of meaningful <u>patterns</u> that can be obtained from traditional Point of Sale checkout <u>data</u>:

S/N	Pattern	Solvable Problems
1	Popularity and demand of products.	• What products to promote? • How much inventory to stock?
2	Product bundles that customers typically buy together.	• Product positioning. • Cross-sell (e.g., PwP) and bundle promotion.
3	Peak shopping periods.	• Off-peak/seasonal promotion. • Staffing level.

- Through the process, **business value** is created!

Case Study – Creating Business Value with Analytics

- Target Corporation is the second-largest discount store retailer in the United States, behind Walmart.
- Target's creation of a "pregnancy predictor" is a good illustration of the insights that can be gained through leveraging big data in an effective business analytics practice.

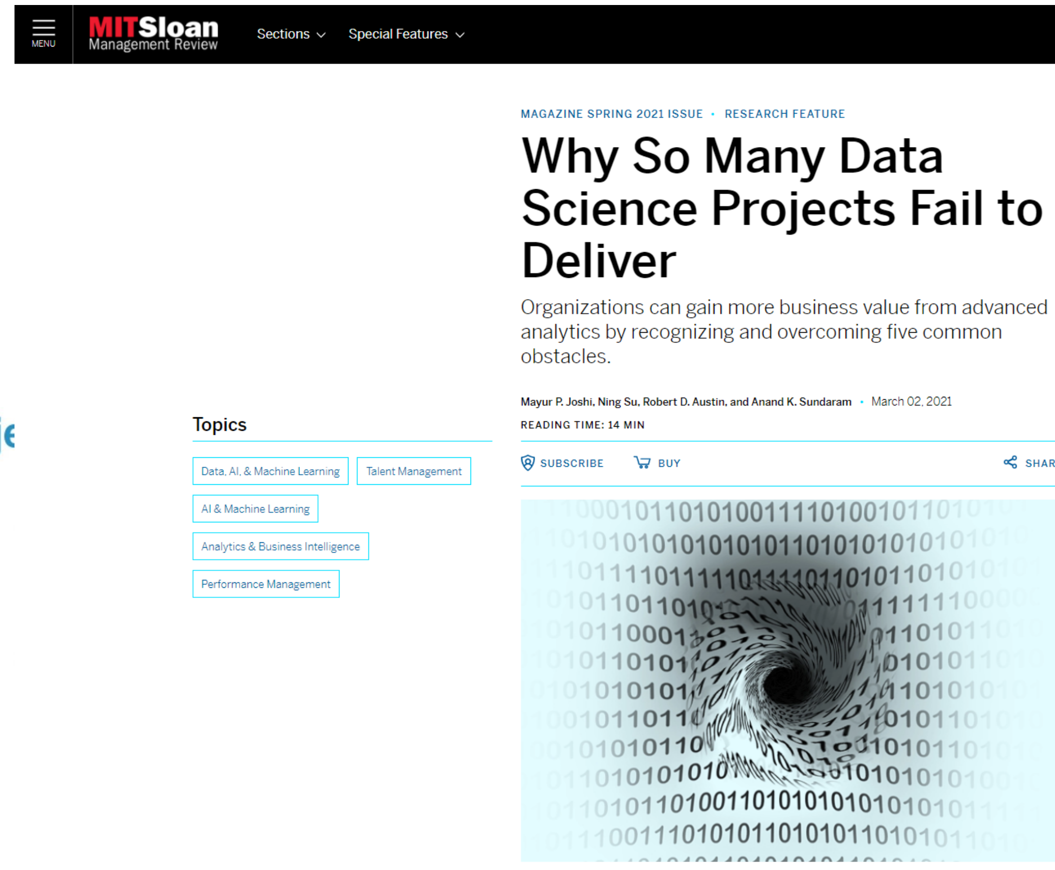


Case Study – Creating Business Value with Analytics (cont.)

- "*Andrew Pole had just started working as a statistician for Target in 2002, when two colleagues from the marketing department stopped by his desk to ask an odd question: 'If we wanted to figure out if a customer is pregnant, even if she didn't want us to know, can you do that?.........'*"
- http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Business Analytics Projects Suffer from High Failure

- But business analytics is a <u>complex</u> process and analytics projects suffer from many challenges and high <u>failure</u> rate:


MIT Sloan Management Review
Sections
Special Features
MAGAZINE SPRING 2021 ISSUE • RESEARCH FEATURE

Why So Many Data Science Projects Fail to Deliver

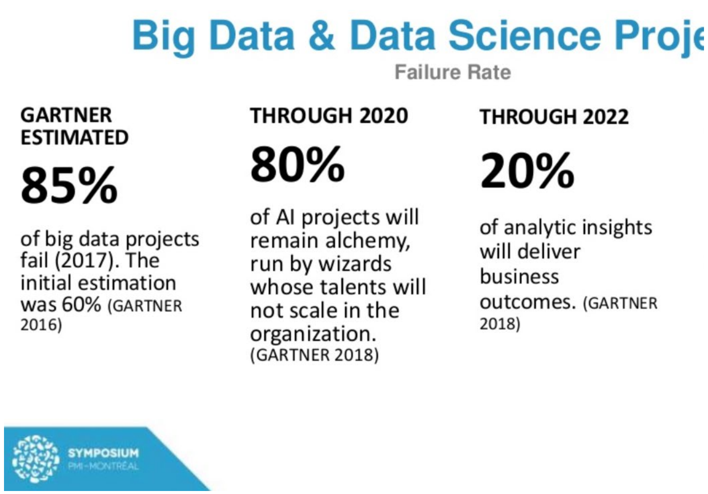
Organizations can gain more business value from advanced analytics by recognizing and overcoming five common obstacles.

Mayur P. Joshi, Ning Su, Robert D. Austin, and Anand K. Sundaram • March 02, 2021
READING TIME: 14 MIN

SUBSCRIBE BUY SHARE

Topics
Data, AI, & Machine Learning
Talent Management
AI & Machine Learning
Analytics & Business Intelligence
Performance Management

Image courtesy of Jean Francois Podevin/theispot.com



Business Analytics Projects Suffer from High Failure (cont.)

- Forbes – Solving The Last Mile Problem For Data Science Project Success:

  - In 2017, Gartner estimated that 85% of data science projects were failing.
  - Data science projects fail because of the last mile problem and not because of technology, data, talent or technical skills.
  - Data scientists struggle to deploy models into business processes and applications that are used by the business.
  - Successful data science projects marry technical excellence with soft skills such as collaboration and transparency.

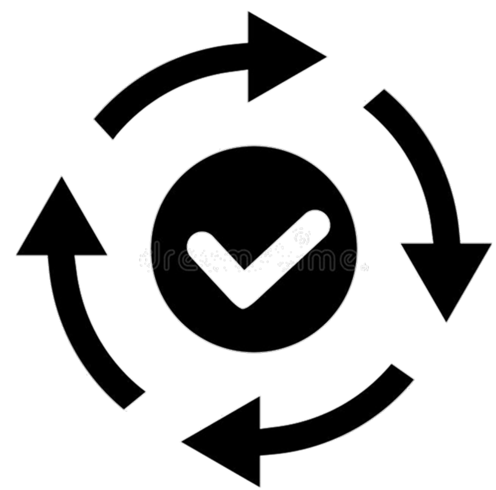
Business Analytics Projects Suffer from High Failure (cont.)

- MIT Sloan Management Review – Why So Many Data Science Projects Fail to Deliver:

  - The Hammer in Search of a Nail – Infatuation with data science solutions.
  - Unrecognized Sources of Bias – Using biased data.
  - Right Solution, Wrong Time – Lack of synchronization between data science and the priorities and processes of the business.
  - Right Tool, Wrong User – Lack of attention to how the outputs of analytical tools are communicated and used.
  - The Rocky Last Mile – Lack of coordination between data science team and business stakeholders in solution implementation.

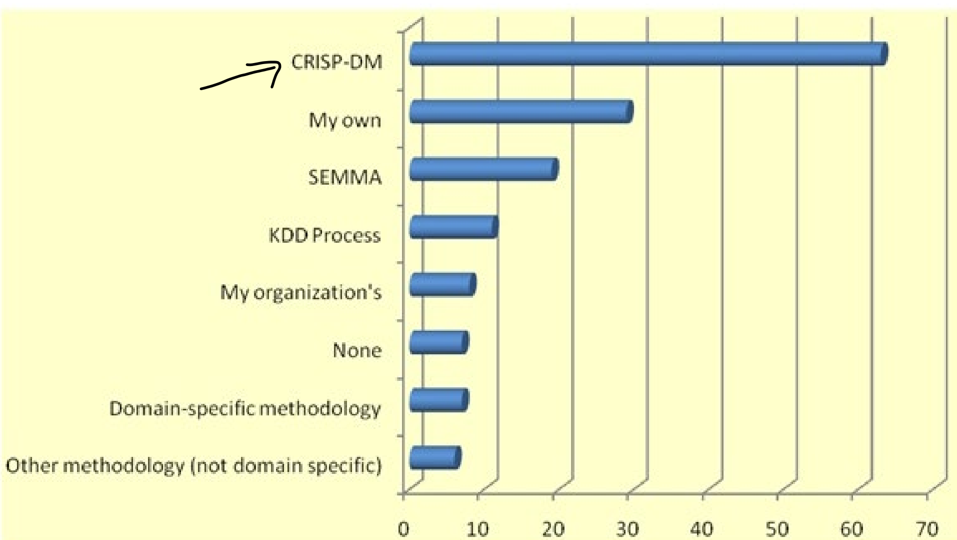
What is Business Analytics Project Lifecycle?

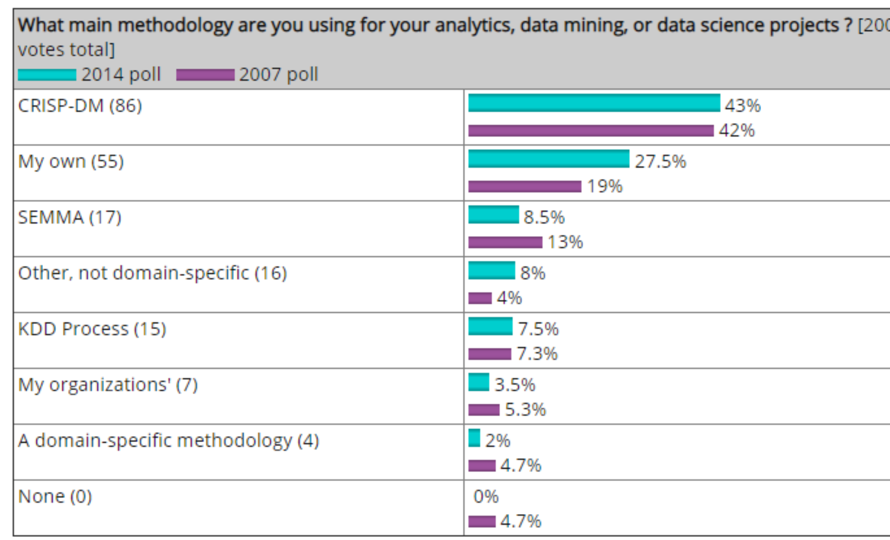
- A **business analytics project lifecycle** typically defines several key stages that are essential for ensuring the <u>success</u> of an <u>analytics</u> project.
- Generic stages include:
  - Problem definition and understanding.
  - Data collection and preparation.
  - Data analysis and modeling.
  - Interpretation and validation.
  - Deployment and implementation.
  - Monitoring and maintenance.
  - Communication and reporting.



Traditional Lifecycle Frameworks

- Various standardized framework based on best practices:
  - CRISP-DM (Cross-Industry Standard Process for Data Mining)
  - SEMMA (Sample, Explore, Modify, Model, and Assess)
  - KDD (Knowledge Discovery in Databases)



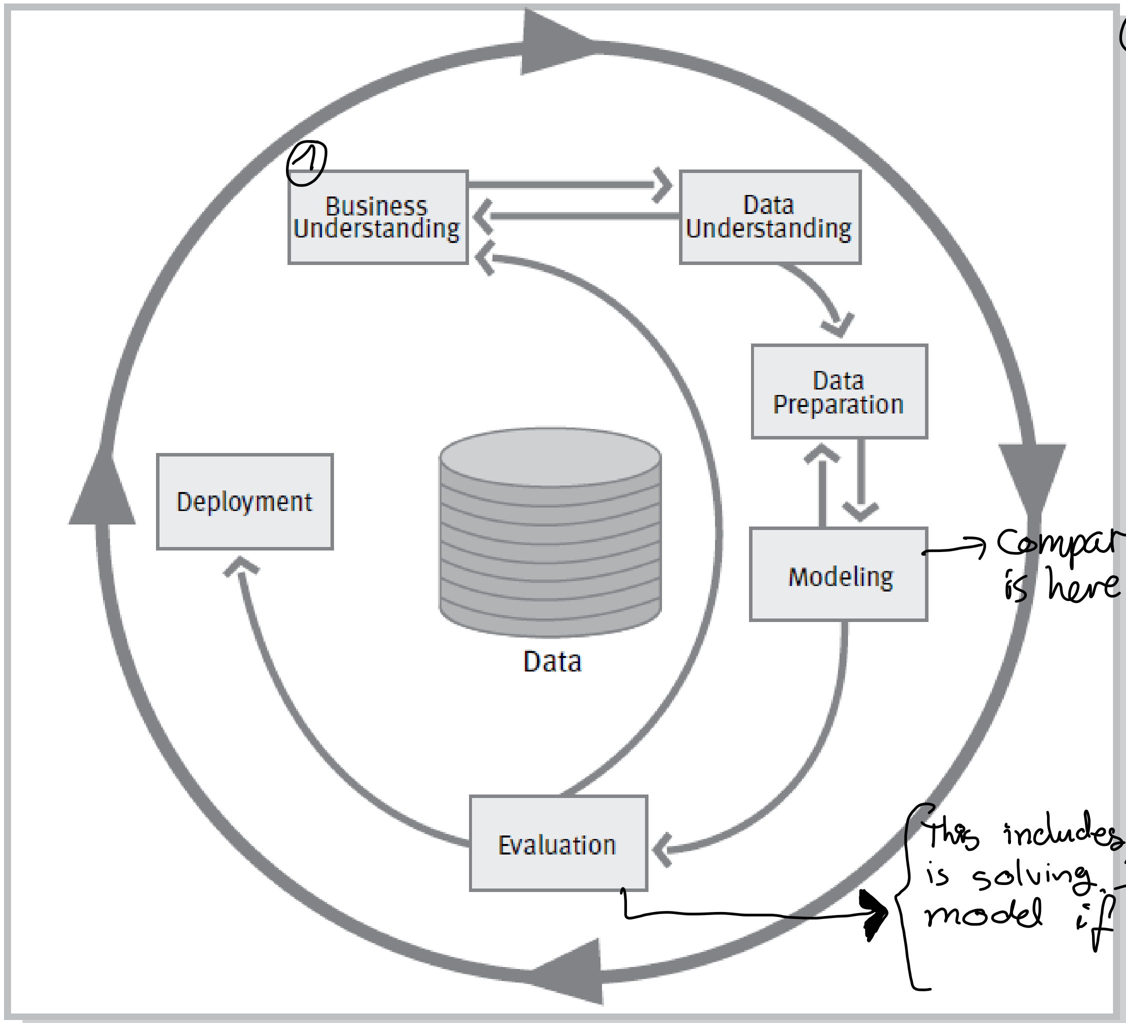


Source: KDNuggets.com, 2007 (L) and 2014 (R)

CRISP-DM

- **CRISP-DM (CRoss-Industry Standard Process for Data Mining)**:
  - Prescribes a set of activities to be done at each step together with the expected deliverables to guide the data analyst.
  - Emphasises on business problem instead of data.
- CRISP-DM consists of six steps:
  - The steps are sequential in nature but usually involve backtracking.
  - Whole process is iterative and could be time consuming.
  - The outcome from each step feeds into the next step and thus each step must be carefully conducted.

CRISP-DM (cont.)

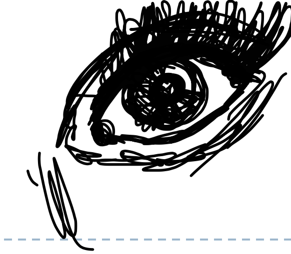


1 You first need to understand the b. problem.

Comparing models is here.

This includes how the model is solving the business model if we deploy it.

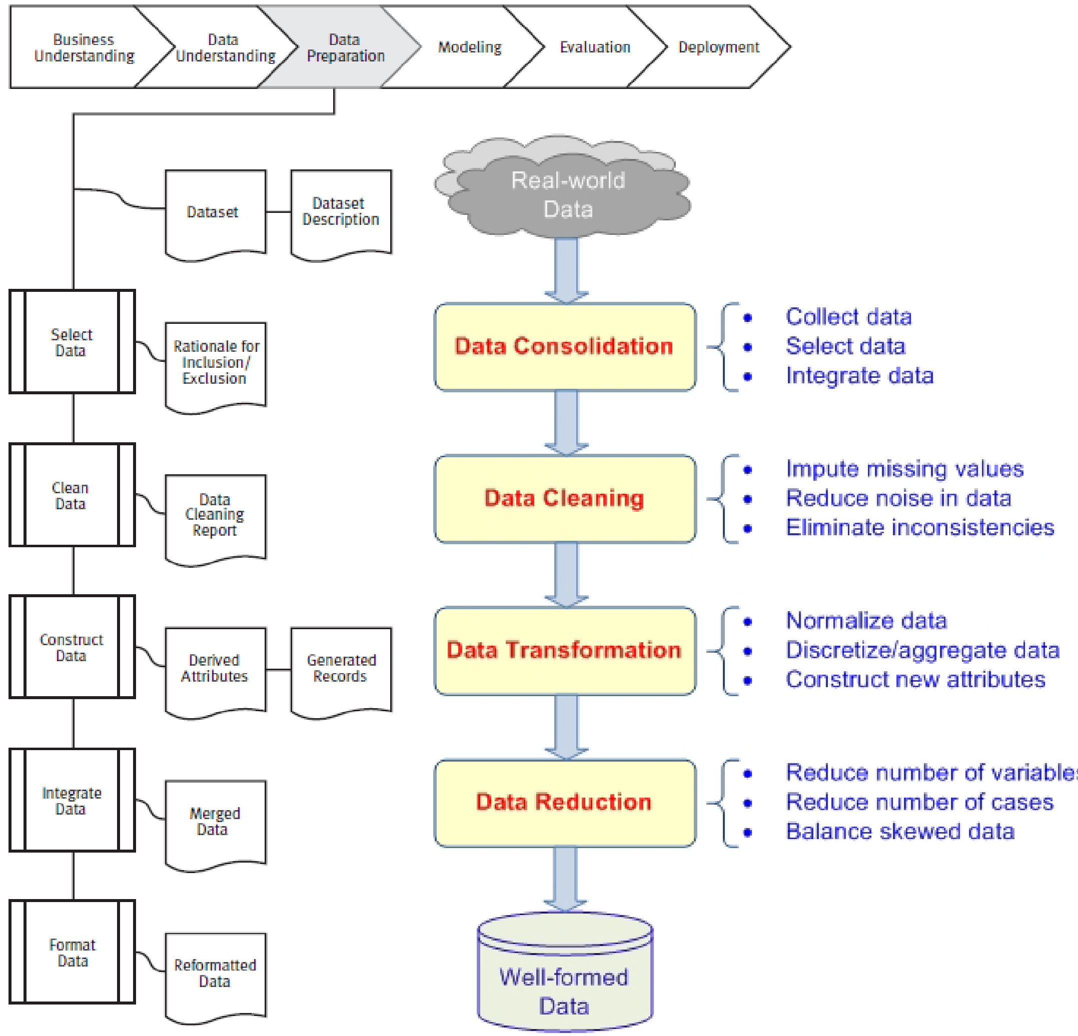
CRISP-DM (cont.)



Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits* **Determine Data Mining Goals** *Data Mining Goals* *Data Mining Success Criteria* **Produce Project Plan** *Project Plan* *Initial Assessment of Tools and Techniques*	**Collect Initial Data** *Initial Data Collection Report* **Describe Data** *Data Description Report* **Explore Data** *Data Exploration Report* **Verify Data Quality** *Data Quality Report*	**Select Data** *Rationale for Inclusion/ Exclusion* **Clean Data** *Data Cleaning Report* **Construct Data** *Derived Attributes* *Generated Records* **Integrate Data** *Merged Data* **Format Data** *Reformatted Data* *Dataset* *Dataset Description*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions* **Generate Test Design** *Test Design* **Build Model** *Parameter Settings* *Models* *Model Descriptions* **Assess Model** *Model Assessment* *Revised Parameter Settings*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models* **Review Process** *Review of Process* **Determine Next Steps** *List of Possible Actions* *Decision*	**Plan Deployment** *Deployment Plan* **Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan* **Produce Final Report** *Final Report* *Final Presentation* **Review Project** *Experience Documentation*

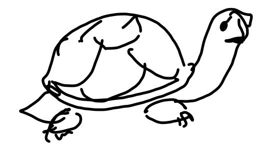
CRISP-DM (cont.)

- Although <u>business-oriented</u>, CRISP-DM focuses mainly on the <u>technicalities</u> of analytics:
  - E.g., data preparation entails many sub-steps.



Traditional Frameworks vs. More Modern Frameworks

- Traditional frameworks such as CRISP-DM:
  - Lower-level frameworks.
  - Guide the technicalities of the actual analytics process in a project.



Why we do what we want to do.

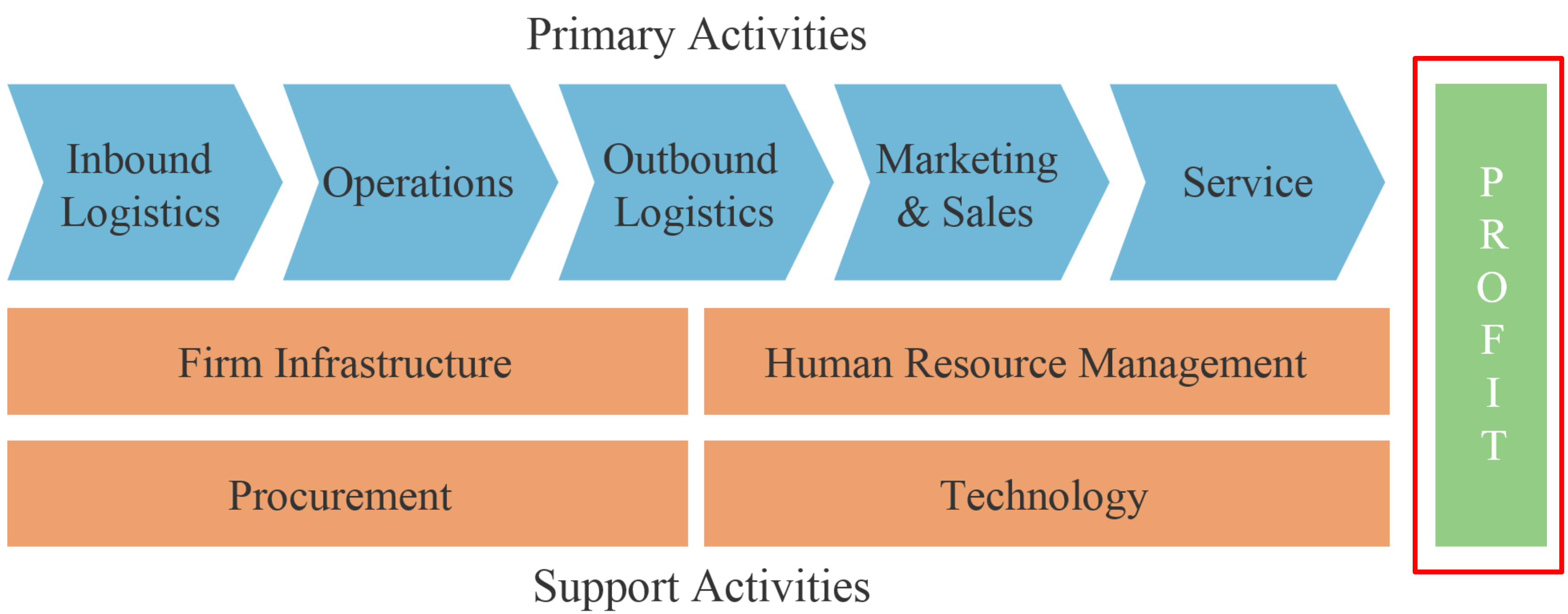
- More modern frameworks:
  - Higher-level frameworks.
  - Guide the overall governance process of a project including many important non-technical aspects.
- The two types of frameworks are **not** mutually exclusive and in fact we can use them **together**:
  - We can apply CRISP-DM to implement a specific project that is framed within a bigger governance process.

Analytics Value Chain

- The **Analytics Value Chain**:
  - Conceptualised by Judah Philips
  - A six-phase framework that explains the generalised process for "doing analytics".
  - Execution of each phase in the value chain completes necessary work for creating value with analytics.
  - It illustrates the various activities that the analytics team performs when it is "doing analytics".
- Why do we need the value chain?
  - In most analytics project, some level of preparation is required.
  - Data in an analytics project does not exist just because there is a business has a digital channel, e.g., website/social/mobile.
    ↳ what is the business indicator we want to tackle?

Analytics Value Chain (cont.)

- Porter's **Value Chain Model** is a more commonly used and established example of a value chain:



Analytics Value Chain (cont.)

- The <u>six phases</u> in the Analytics Value Chain are:

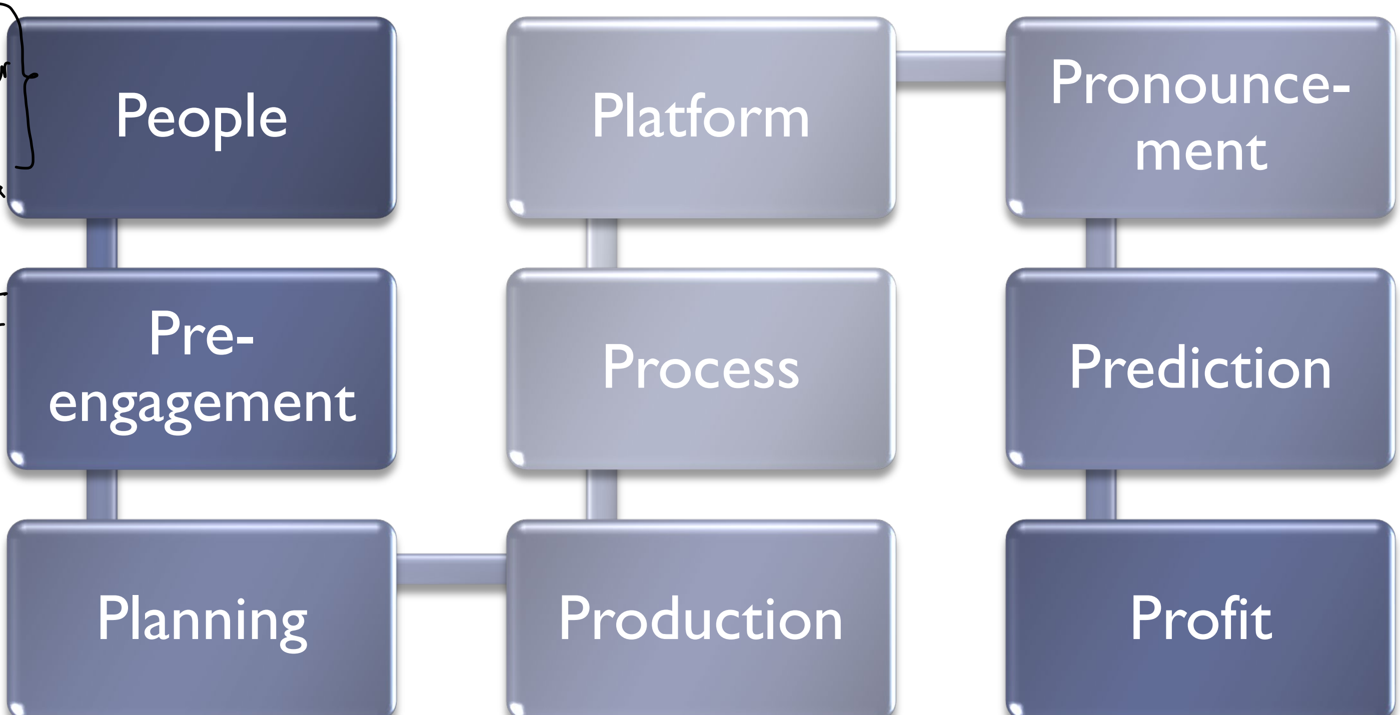
1. Understanding what to analyse.
2. Collecting, verifying, and governing data.
3. Reporting and dashboarding. → EDA (Expl. data analysis).
4. Analysing data, and communicating and socialising.
5. Optimising, predicting, and automating. → automate the delivery of insights.
6. Demonstrating economic value.

Analytics Value Chain (cont.)

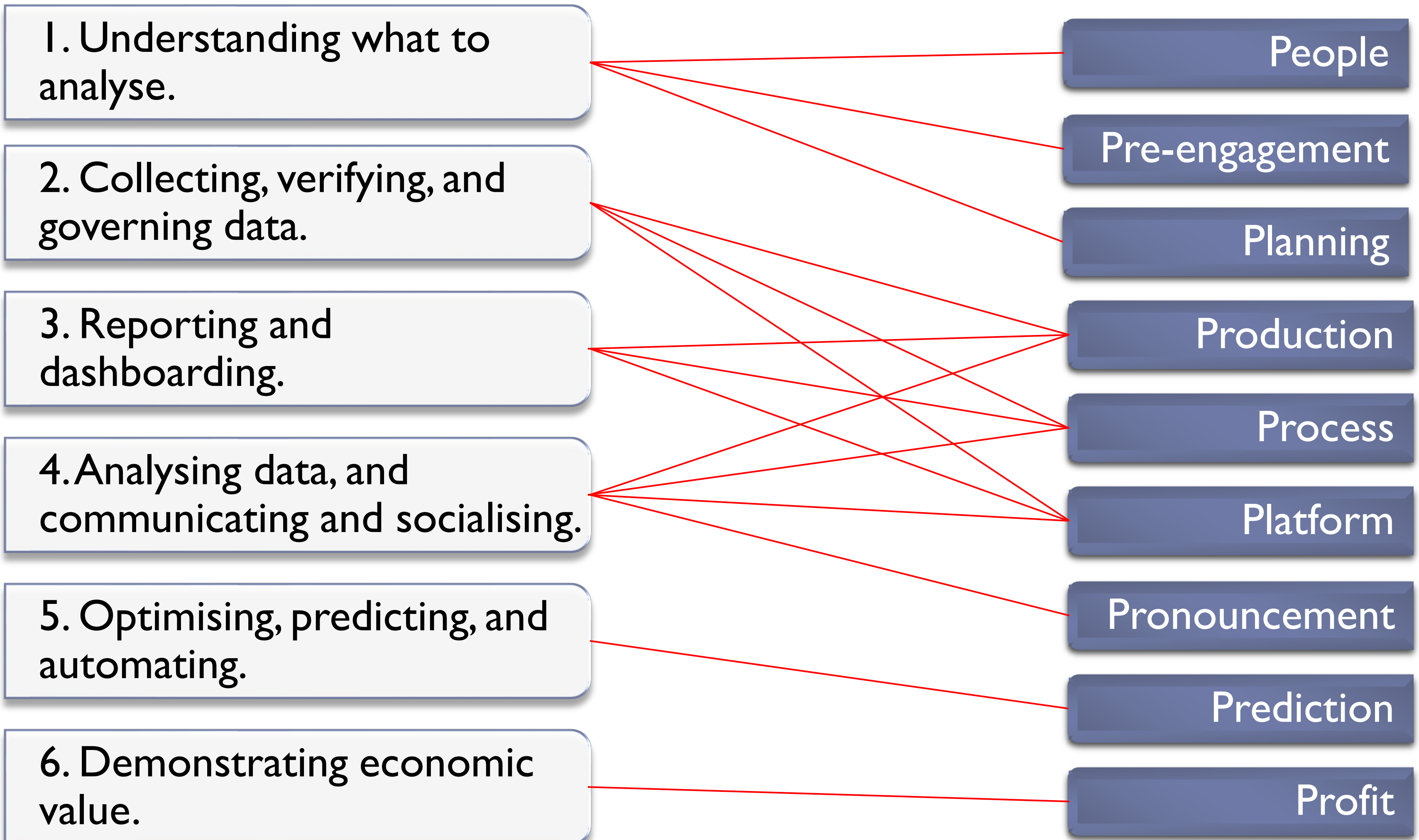
- The Nine P's of Analytics:
  - Identify the critical elements of a typical analytics project.
  - Constituent elements of the Analytics Value Chain.

Data Science
Business stakeholder
Technology team
↳ engineers help you get the data and develop the deployed solution.

Com el PID del Marto.
Si el projecte no té alta prob. the success, no ho farem.



Analytics Value Chain (cont.)



Mapping the six phases to the nine Ps

Understanding What to Analyse

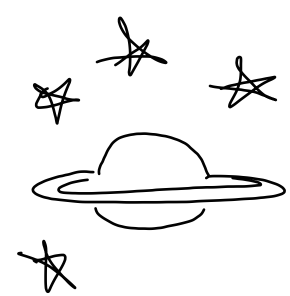
- Need to understand what the analytics team is being asked to analyse.
- Phrased in the form of a <u>business question</u> with a <u>relevant business challenge</u>:
  - Examples of typical questions asked that are **NOT** business questions:
    - How many unique visitors do we have?
    - How did X campaign do this week compared to last week?
  - Analyst needs to gradually evolve such <u>questions</u> into <u>business questions</u>:
    - What was the <u>impact on revenue and audience reach</u> from the <u>modifications made</u> to our inbound marketing campaign?

Understanding What to Analyse (cont.)

- Need to meet and discuss with business stakeholders and get buy-in.
- After formulating the business questions, the analytics team need to <u>define the data required for the analysis</u>:
  - Data definition need to be business-focused, operationally capable, technically accurate, and consensus-agreed.
  - The best data definition manifests in <u>three forms</u> – <u>business definition</u>, <u>operational definition</u> and <u>technical definition</u>.

Collecting, Verifying, and Governing Data

- After identifying business questions and defining data, the next step is to collect the data and verify that it is accurate and usable.
- Data collection is a complex tasks:
  - Collection of new data can be nuanced and complex.
  - Collation and integration of existing data from multiple sources can also be complex.
- Data collection can be automated by APIs (Application Programming Interfaces).
- However, support for data verification and data governance is limited.



Reporting and Dashboarding

- After data has been collected, integrated and verified, the analytics team will start exploratory analysis:
  - Segment data in multi-dimensional views, filter and drill-down.
- Reports are the main artifacts for the initial analysis:
  - Simple data tables, cross-tabulation and pivot tables.
  - Descriptive statistics.
  - Simple data visualisation.
  - Reports need to be reviewed and validated.
- Note that reports are **NOT** analysis themselves but the data and visualisation within the reports can be analysed.

Analysing Data and Communicating and Socialising Analysis

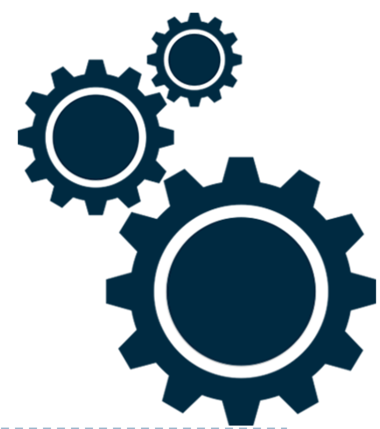
- **Analysis** is the ultimate deliverable of the analytics team.
- At a broad level, analysis is the process of breaking a complex business problem into smaller parts to gain a better understanding.
- The next step is to discuss and communicate the results of the analysis to the relevant stakeholders:
  - Prepare and verify the reports' accuracy.
  - Automating scheduled delivery of the reports can be done but should be minimised. → això és tipus enviar email, però es pot convertir en SPAM!
  - Instead, the analytics team should provide a self-service environment for users to request the reports with justification.

Optimising and Predicting

- A mature and competent analytics team should gradually move to optimisation, prediction and automation.
- **Optimisation**:
  - The baseline analysis answers the business question.
  - Optimisation involves answering further questions such as "What about this?" and "What about that?"
- **Prediction**:
  - Optimisation seeks to improve what we already know.
  - Prediction is the application of statistical methods for data mining and machine learning to identify what might happen next, or the next best course of action to take, i.e., find out about the unknown.

Optimising and Predicting (cont.)

- **Automation**:
  - The outcomes from analysis, testing, optimisation and prediction provide opportunities for using data to automate an online customer interaction.
  - Such data include previous behaviours, expressed preferences and propensities, and other digital expressions of interest.
  - Example of an automation:
    - A person self-identifies a preference for a specific topic during the initial visits.
    - A person's preference can also be predicted.
    - Upon the next and subsequent visits, the person will be offered dynamically customised content experience.

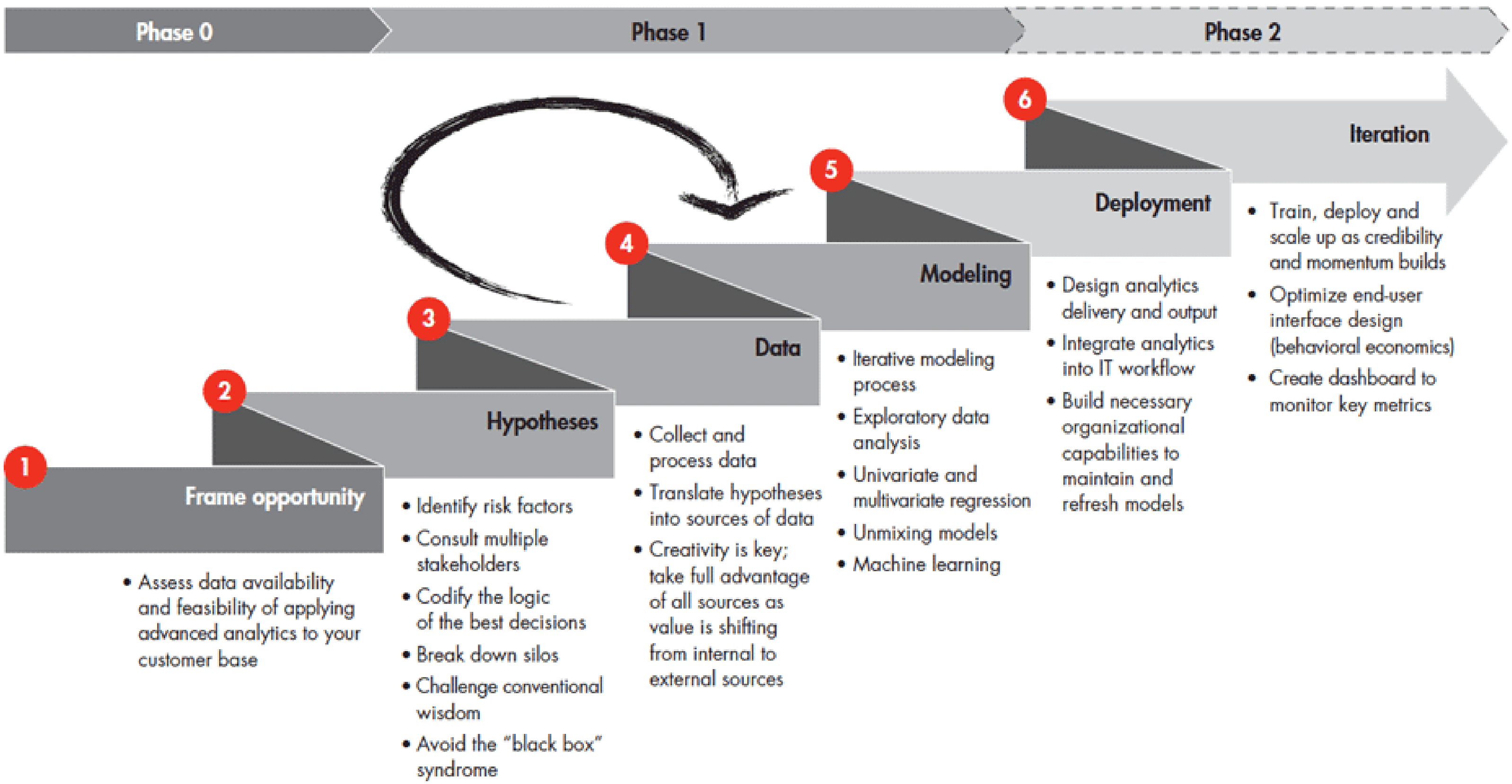


Demonstrating Economic Value

- The best use of analytics is to link back the customer interaction, events and transactions to financial metrics:
  - New and incremental revenue.
  - Cost reduction.
  - Increased efficiency.
  - Boosted or enhanced productivity.

More Examples of Business Analytics Project Lifecycle

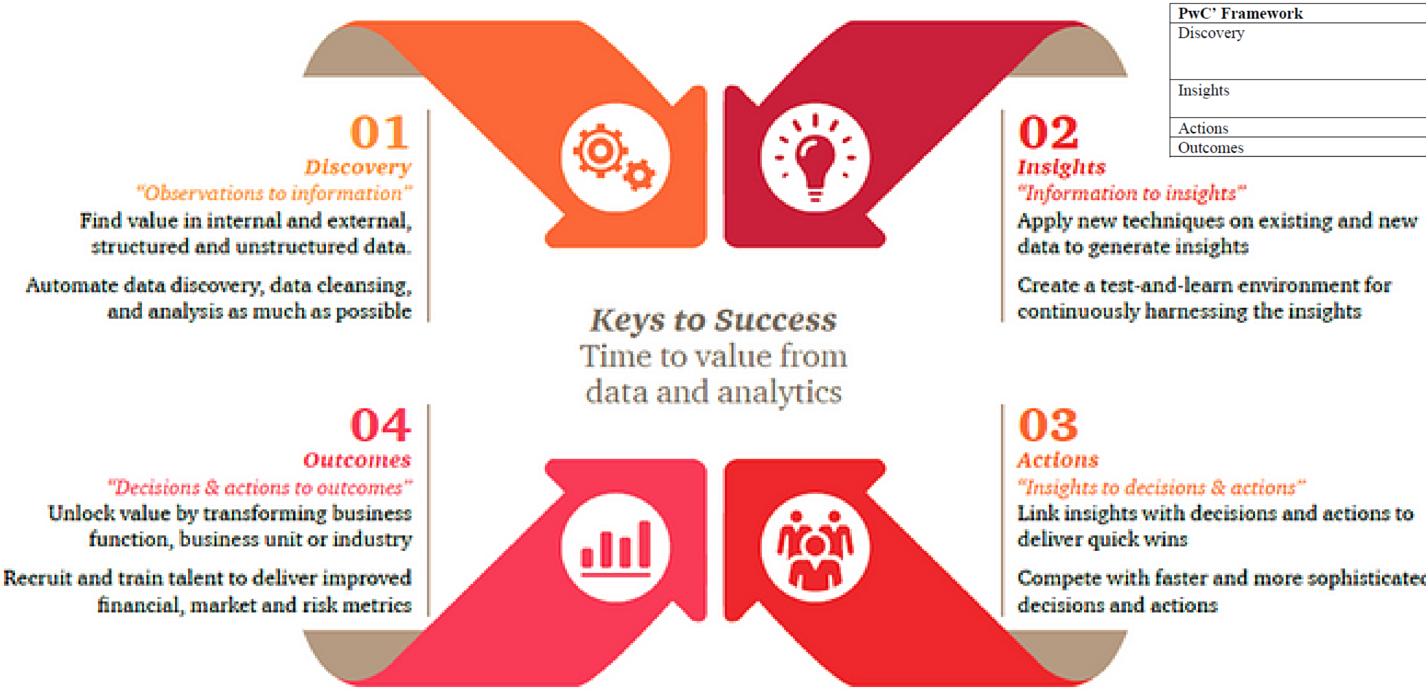
- **Bain & Company:**
  - A global management consulting firm.



Source: Bain & Company

More Examples of Business Analytics Project Lifecycle (cont.)

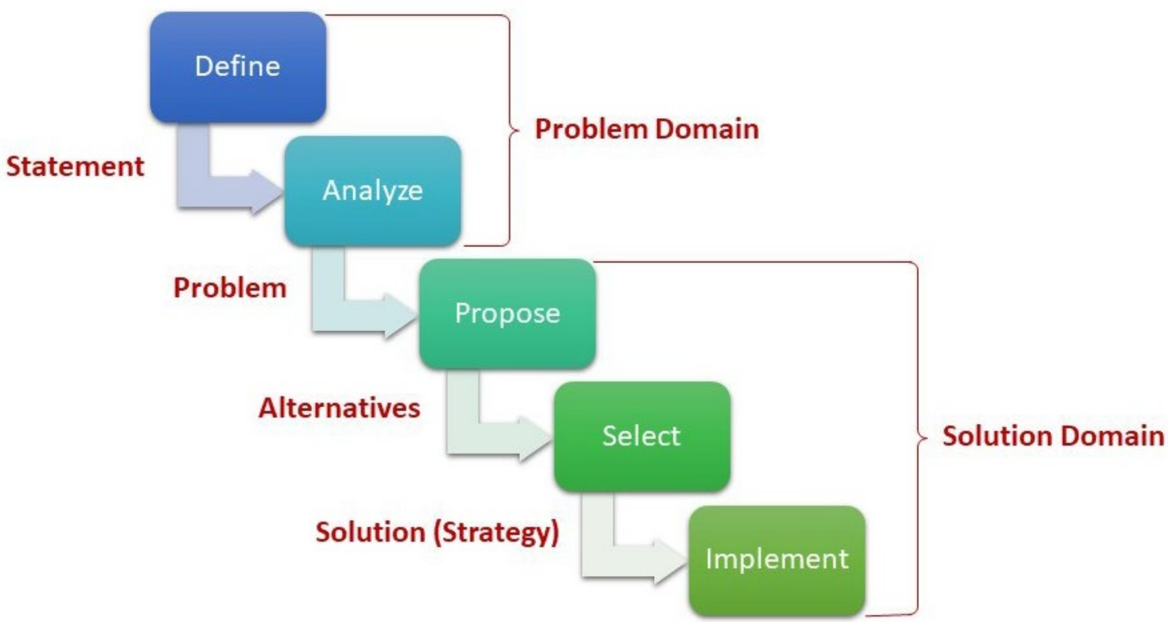
- **PricewaterhouseCoopers**:
  - A global professional services firm including accounting, tax and consulting.
  - PwC data and analytics framework:



PwC' Framework	Analytics Value Chain
Discovery	1. Understanding what to analyse. 2. Collecting, verifying, and governing data. 3. Reporting and dashboarding.
Insights	4. Analysing data. 5. Optimising, predicting, and automating.
Actions	4. Communicating and socialising.
Outcomes	6. Demonstrating economic value.

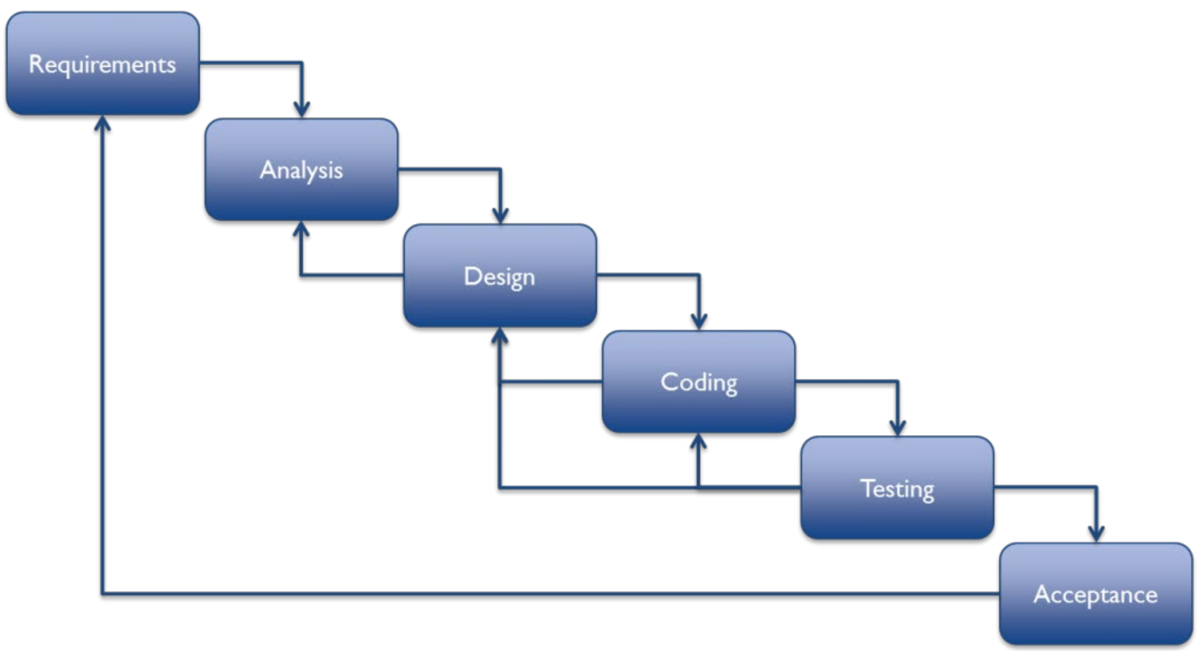
Are These Frameworks Good or Bad?

- In general problem solving:
  - We can adopt a methodology to improve the problem-solving process and increase the chance of successful outcome.
  - Most basic problem-solving methodology entails a waterfall model.



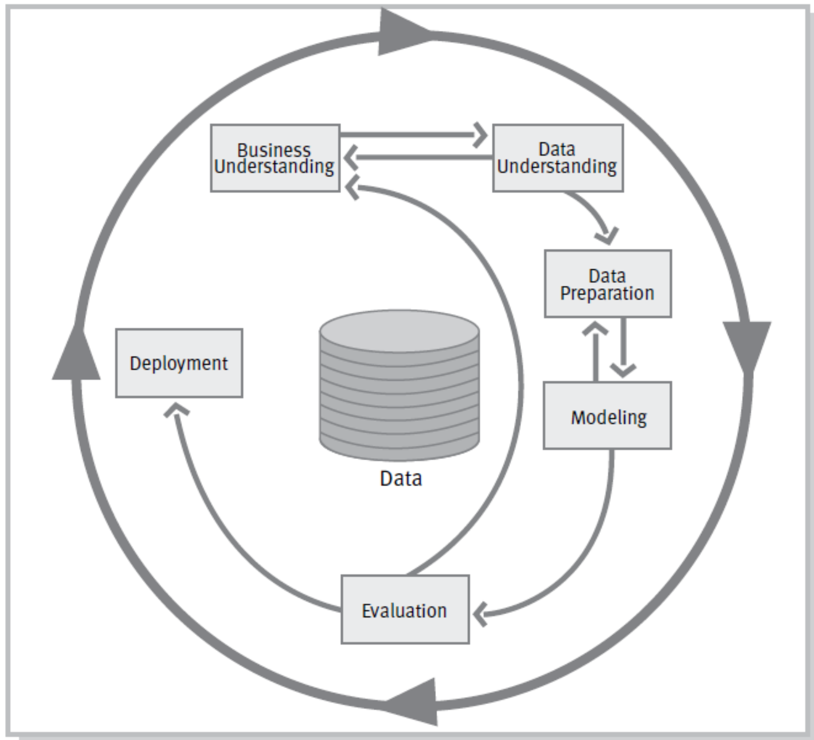
Are These Frameworks Good or Bad? (cont.)

- In software application development:
  - We can adopt a similar waterfall methodology.
  - Objective is to create a software application that meets all business requirement, on schedule and within budget.



Are These Frameworks Good or Bad? (cont.)

- In business analytics:
  - CRISP-DM resembles a waterfall model. Really?



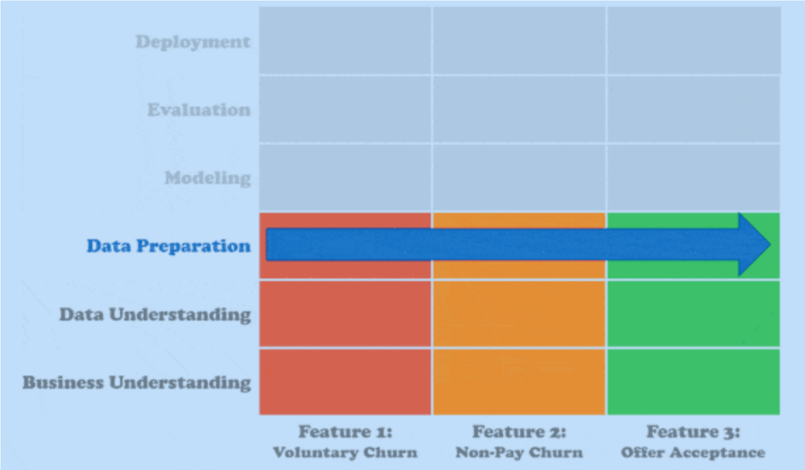
A CRISP-DM model is almost a waterfall model.

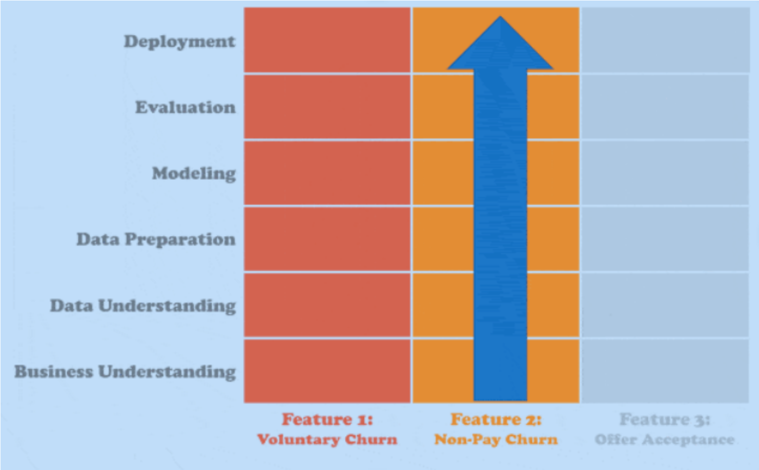
Waterfall approach is better for a mission-critical, and well-known requirements, since it gives more certainty

- So, is waterfall good or bad?

Are These Frameworks Good or Bad? (cont.)

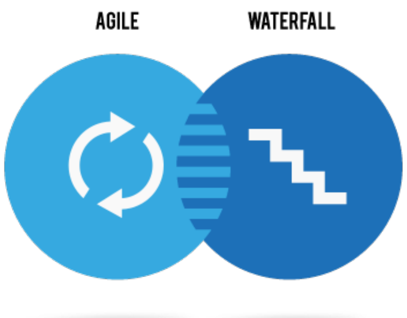
- Is CRISP-DM **waterfall** or **agile**?
  - One point of view:
    - CRISP-DM indirectly adopts agile principles and practices.
    - The sequence of phases is not rigid.
  - Another point of view – What a team does:
    - Horizontal slicing in waterfall (L) – A team does one specialised task.
    - Vertical slicing in agile (R) – A team does all tasks.





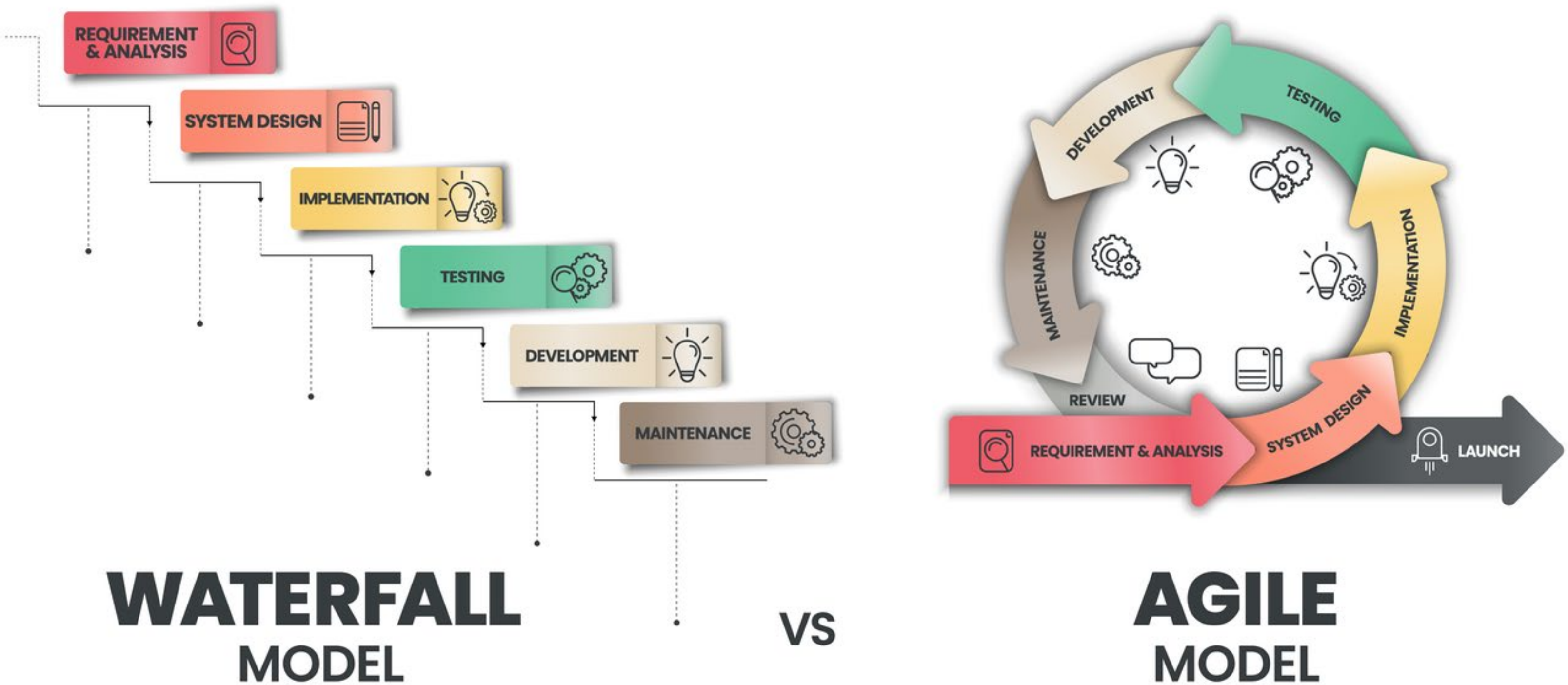
The Case for Agile

- Business analytics <u>project, project and project</u>…
  - Many organisations continue to approach analytics as a traditional discrete IT "project" of a finite duration.
  - The project mentality is understandable but prevents them from really embedding analytics into their cultures.
- Successful companies such as Netflix, Facebook, Amazon and Google view analytics as projects that never end:
  - Analytics need to be <u>continuously delivered</u>.
  - Analytics need to be consistently refined, audited, expanded, and eventually retired when it no longer makes sense.

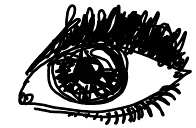


The Case for Agile (cont.)

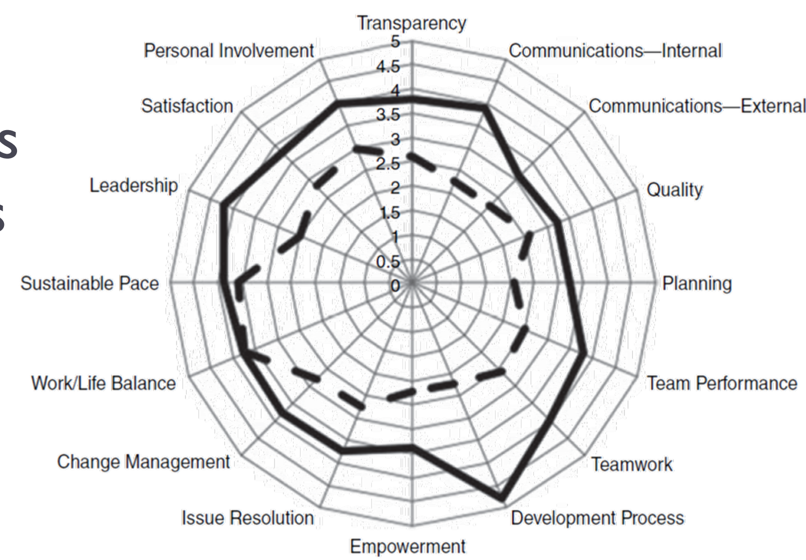
- **Waterfall** emphasizes a linear and deliberate process.
- **Agile** emphasizes a continuous cycles of incremental and iterative process.



The Case for Agile (cont.)



- With high failure rates among waterfall projects:
  - Academics and practitioners have been advocating the adoption of agile methods.
  - Agile methods has been gaining popularity since mid-1990s and early 2000s ➔ This is year 2024.
- According to Rico et. al. in their book "The Business Value of Agile Software Methods":
  - Agile software development teams reported more positive outcomes compared to non-agile teams.

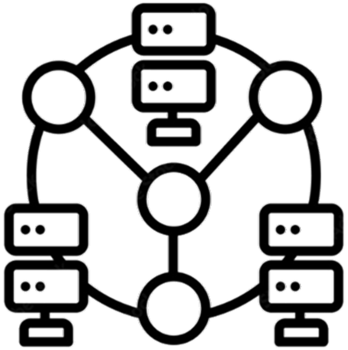


The Case for Agile (cont.)

- **Agile analytics** remains focused on finding business value in data:
  - Just a faster and more free-flow approach that attempts to respond to rapidly changing business environment.
  - We will discuss more about agile analytics in the next lecture.

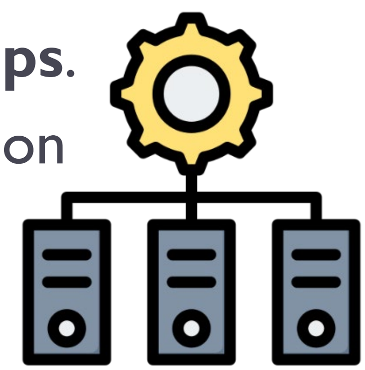
Beyond Agility

- There are <u>two</u> other major problems associated with business analytics project that we will attempt to address.
- **Connecting the data**:
  - Modern business analytics projects are characterised by **big data**.
  - Big data are characterised by at least **three Vs – Volume, Velocity and Variety**.
  - Moving data from source to model training is a complex task.
  - **DataOps** is required to automate data integration and data transformation.

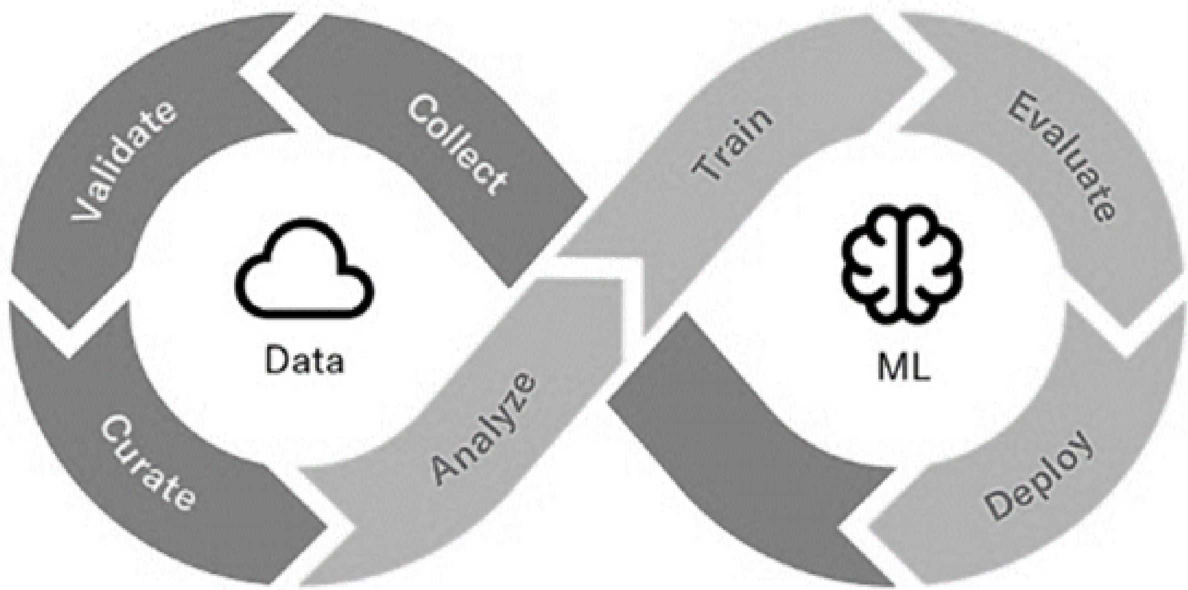


Beyond Agility (cont.)

- **Infrastructure**:
  - Developing machine learning models involve more than data science expertise.
  - Considerable IT and infrastructure skills are also required:
    - Huge data sets have to be aggregated, stored, moved, protected and managed.
    - Training and testing the models requires very high levels of compute capacity and performance.
  - Accelerating the lifecycle requires abstracting the infrastructure layer from the data science with **MLOps**.
  - Productionising models also require close collaboration between data scientists and engineers.



Beyond Agility (cont.)



DataOps

MLOps

Summary

- Business analytics project lifecycle prescribes essential steps to be taken to ensure success.
- Traditional lifecycle frameworks focus on the technicalities of business analytics projects.
- More modern frameworks address higher-level governance and other important non-technical issues.
- Agile analytics frameworks are better suited to ensure continuously delivery and improvement of solutions.
- DataOps and MLOps work together within agile analytics frameworks to automate data management and optimise infrastructure management.

Q&A

Next Lecture…

- **Learn about:**
  - Agile methods.
  - Integrating agile methods with analytics.