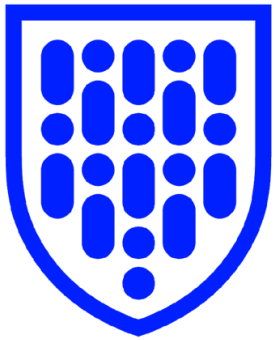


Institute of Data

2020

Data Science and AI

Module X

Capstone Project

Capstone Project

- You are required to **define, design and deliver** a **data science project** towards the end of the course.
- Project milestones:
  - <date tbd> : present **3 ideas for the project**
  - <date tbd> : decide on one option
  - <date tbd> : collect data
  - <date tbd> : present initial findings
  - <date tbd> : present an update
  - <date tbd> : Dry run of final presentation
  - <date tbd> : **Present final report**

What to present

- **Business perspective**
  - Business insights uncovered
  - Business scenarios for how the project can be deployed and used
  - Approach for estimating business value
- **Technical perspective**
  - Techniques used
  - Pipeline
  - Model validation results

Project evaluation criteria

The project is evaluated on the quality, clarity and completeness of the definition, design and delivery of the project.

- **Definition (20%)**
  - Business context, stakeholders and value
  - Data description, sources, quality
- **Design (30%)**
  - Data exploration, analysis and visualisation
  - Documentation: text document, presentation and Notebook
  - The project planning, effort allocation and next steps
- **Delivery (50%)**
  - Feature Engineering
  - Creation of an effective reproducible pipeline
  - Machine Learning model algorithms and accuracy
  - Overall end-to-end solution
  - Delivery of the presentation, poise and audience engagement

Questions?

Presentation Skelton

Project Title

Capstone Project

Presenter’s name and role (ideally Data Scientist)

Agenda

- Bio
- Project Context
- Define
- Design
- Deliver
- Summary, conclusions and next steps
- Appendix: list of supporting documents

The agenda should be ideally repeated as a transition slide between sections

Bio

- Education
- Professional experience
- Data science learnings and experience
- Relevance to the project

Project context

- Industry or domain
- Problem area
- Why is this area interesting?
- Previous work in this area

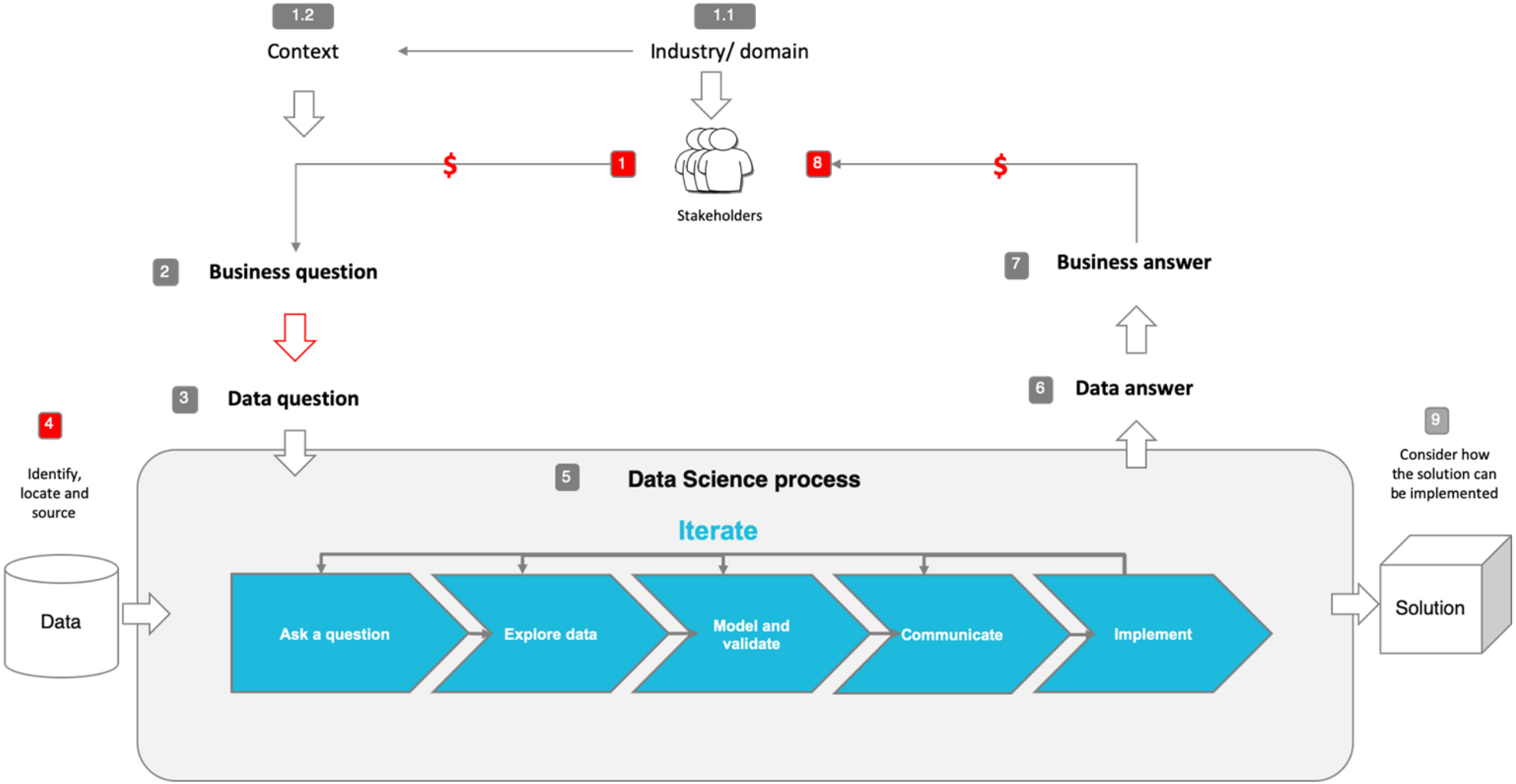
Define

- **Business aspects**
  - Stakeholders
  - **Business question**
  - **Business value**
- **Data science aspects**
  - **Data question**
  - Data required
  - Data sourced
    - Data source, description, volume and quality
    - How is the data generated?
    - How it can be sourced in the future?

Design

- **Data exploration, analysis and visualisation**
  - Include one slide for the **highlights** of your EDA
- **Overall process flow used**
  - See next slide

Applying data science in an industry project



You may use this diagram or your simplified version with project relevant information

Deliver

- **Feature engineering**
  - What are the most important features and what is the business significance
- **Machine models** used and their **evaluation metrics**
- How does the model fit in the **overall solution**
  - Who will use it?
  - When is it used?
  - How does it benefit the business?

Summary, conclusions and next steps

- **Summary**
  - A brief recap of the presentation
- **Conclusions**
  - What has been achieved?
- **Next steps**
  - How can this project be developed further and implemented in real life?

Questions

Appendices

Mini Project

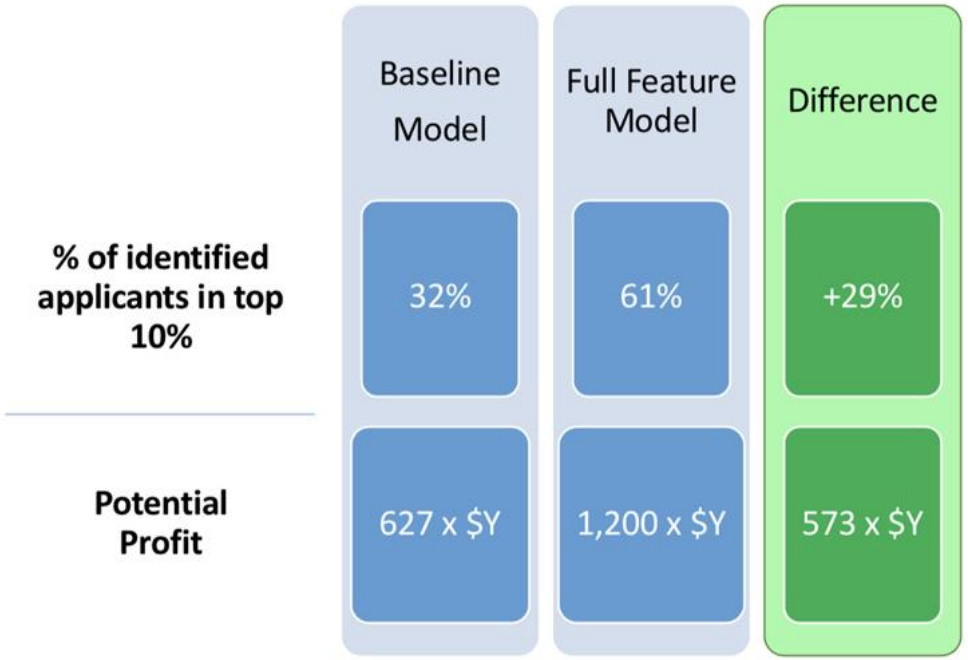
- The **objectives** of this mini-project is to:
  - Apply all what you have learned so far
  - Prepare you for the **Capstone project**
  - Understand your **learning gaps**
- Please present as if it is **your job interview take-away test** on:
  - **Business aspects** of the project
  - **Outcomes**
  - Potential **additional insights** that can be obtained from available data
  - **Data analysis** and **Machine Learning** techniques used
  - **Gaps** in your skills
- Please **listen actively** to your fellow classmate's presentation and **ask questions** as if you are the **hiring manager**

Case study: Home loans marketing

Results comparison and business case overview

Applying the model for Banking can lead to potential annual **revenue twice as big** as the current model.

Results overview



Business case overview based on the Final Model

Assumptions:

- Customer Value/year is $1000
- Customer base = 1.4 million
- Top 10% = 140,000 customers
- 2.8% applicants over 2 years ie. 1.4% annually
- 1.4% applicants in top 10% = 1,960
- 61% identified to target = 1,200
- 32% identified to target = 627

Potential profit = 573 x 1000 ≃ $500,000/ year
≃ $1.5m over 3 years

End of Presentation!