

Institute of
Data



2020



Data Science and AI

Module X

Succeeding as a data scientist in the industry



Agenda: Data Science Industry Experience

- Introduction, definitions, purpose and objectives
- What do employers value and what do they complain about?
- Skills required and attitude to succeed in the industry
- Data Science process
- Case study
- Summary, conclusions and call for actions



Agenda: Succeeding in Industry

- **Introduction, definitions, purpose and objectives**
- What do employers value and what they complains about?
- Skills required and attitude to succeed in industry
- Data Science process in the industry
- Case study
- Summary, conclusions and call for actions



There are few careers as promising right now as a role in data science in Singapore

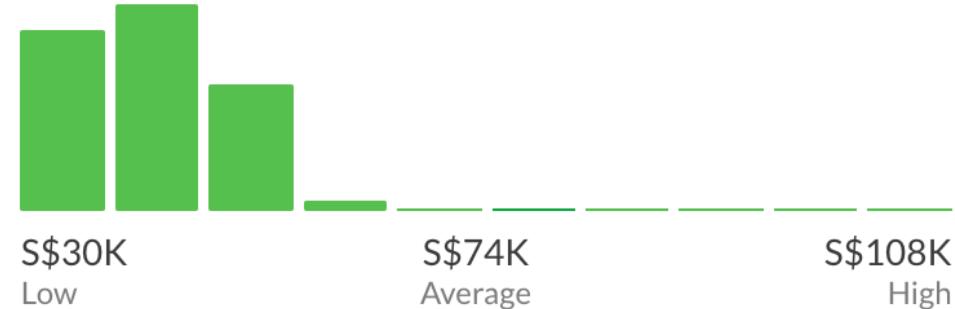
Data Scientist Salaries in Singapore Area

22 Salaries Updated 30 Jul 2020

 Very High Confidence

Average Base Pay

S\$74,323 /yr



How much does a Data Scientist make in Singapore, Singapore?

The average salary for a Data Scientist is \$74,322 in Singapore, Singapore. Salaries estimates are based on...

[More](#)

Salaries for Related Job Titles

Data Analyst	S\$55K
Quantitative Analyst	S\$120K
Senior Data Scientist	S\$110K

Source: [Glassdoor](#)



Working in industry versus research

- Working in the '**industry**' refers to working or consulting for **commercial entities** in **competitive sectors** such as **financial services, telecommunications** or **retail**.
- **Competitive pressures** in these sectors **heighten expectations** from Data Scientists, make it imperative to track **Return on Investment (ROI)** for all projects and **accelerate the pace of work**.
- **Typical** 'university' Data Science **education does not prepare graduates** to effectively work in the industry. This is due to **focus on theoretical** topics and the lack of emphasis on **softer skills** such as communication, collaboration and stakeholders management.



Purpose and objectives of this presentation

- The purpose of this presentation is to share my experience in working as a Data Scientist in the industry with the aim to help you **maximise the value of this course**.
- Objectives of the presentation:
 - Describe **what is valued** in the industry
 - **Prioritise the skills** you should focus on
 - Help **you to get hired**



What do employers value?

- Employers value Data Scientists or other data professionals who **use their technical skills and experiences** to:
 - **Asking the 'right' questions.**
 - **Taking initiative** to deliver **business value**.
 - **Manage Stakeholders'** involvement, communication and effective **team work**.
 - Understand **industry**.
 - Participate actively in delivering **solutions in production**.



What do employers complains about?

- Data Scientists care only about **theory**.
- They treat every project as a **6-month ‘PhD’**.
- They go down **rabbit holes**.
- They use **confusing language**.
- They **cannot** put solutions in **production**.



What should you do to meet these expectations?

- Focus on **business outcomes**.
- Be agile – effective **communication to stakeholders**.
- Understand the **business value** of projects.
- Use **simple models** and communicate in **business language**.
- Develop a small number of **effective and practical skills** and be prepared to learn on the job.



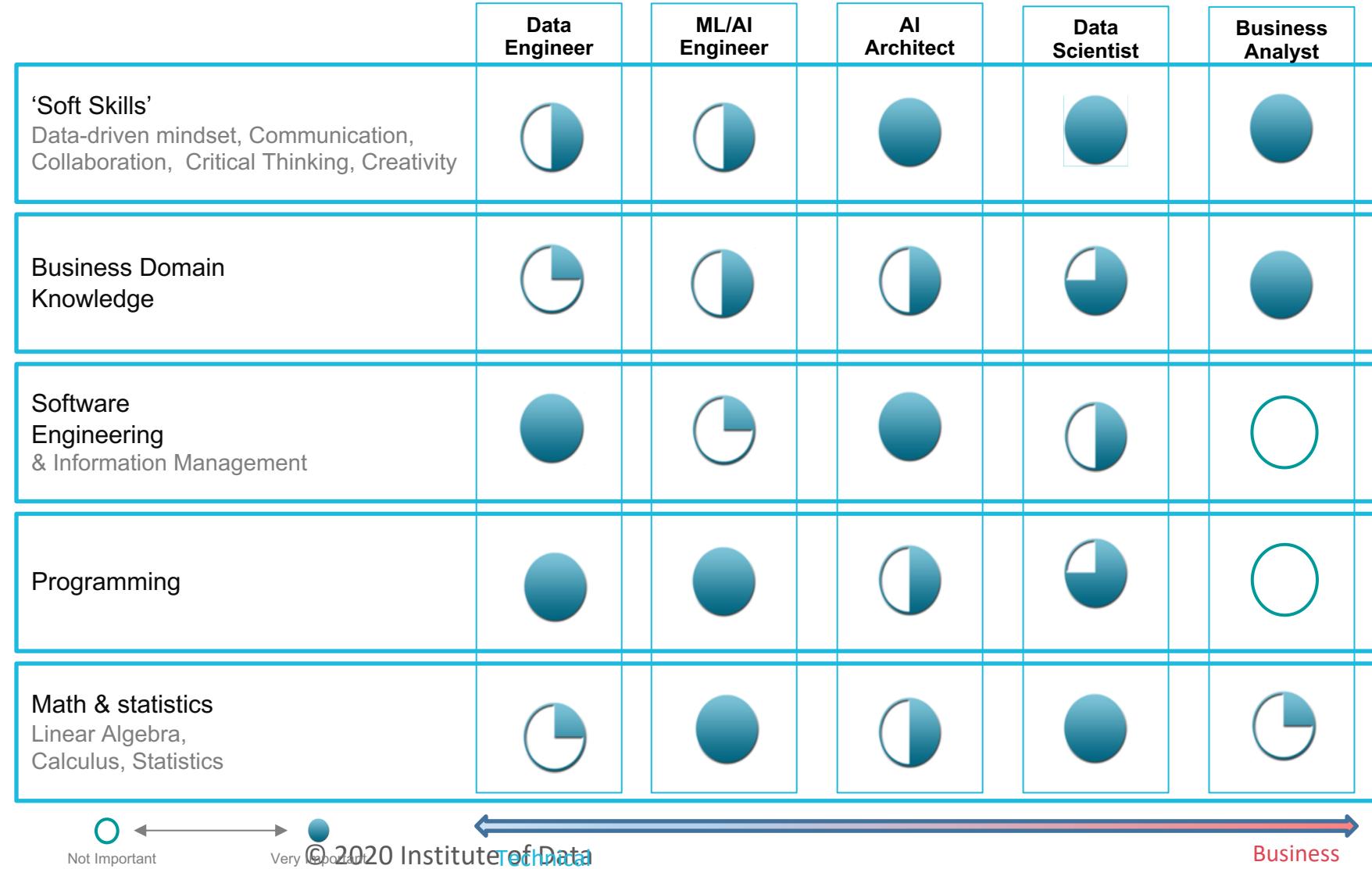
Agenda: Succeeding in Industry

- Introduction, definitions, purpose and objectives
- What do employers value and what they complains about?
- **Skills required and attitude to succeed in industry**
- Data Science process in the industry
- Case study
- Summary, conclusions and call for actions



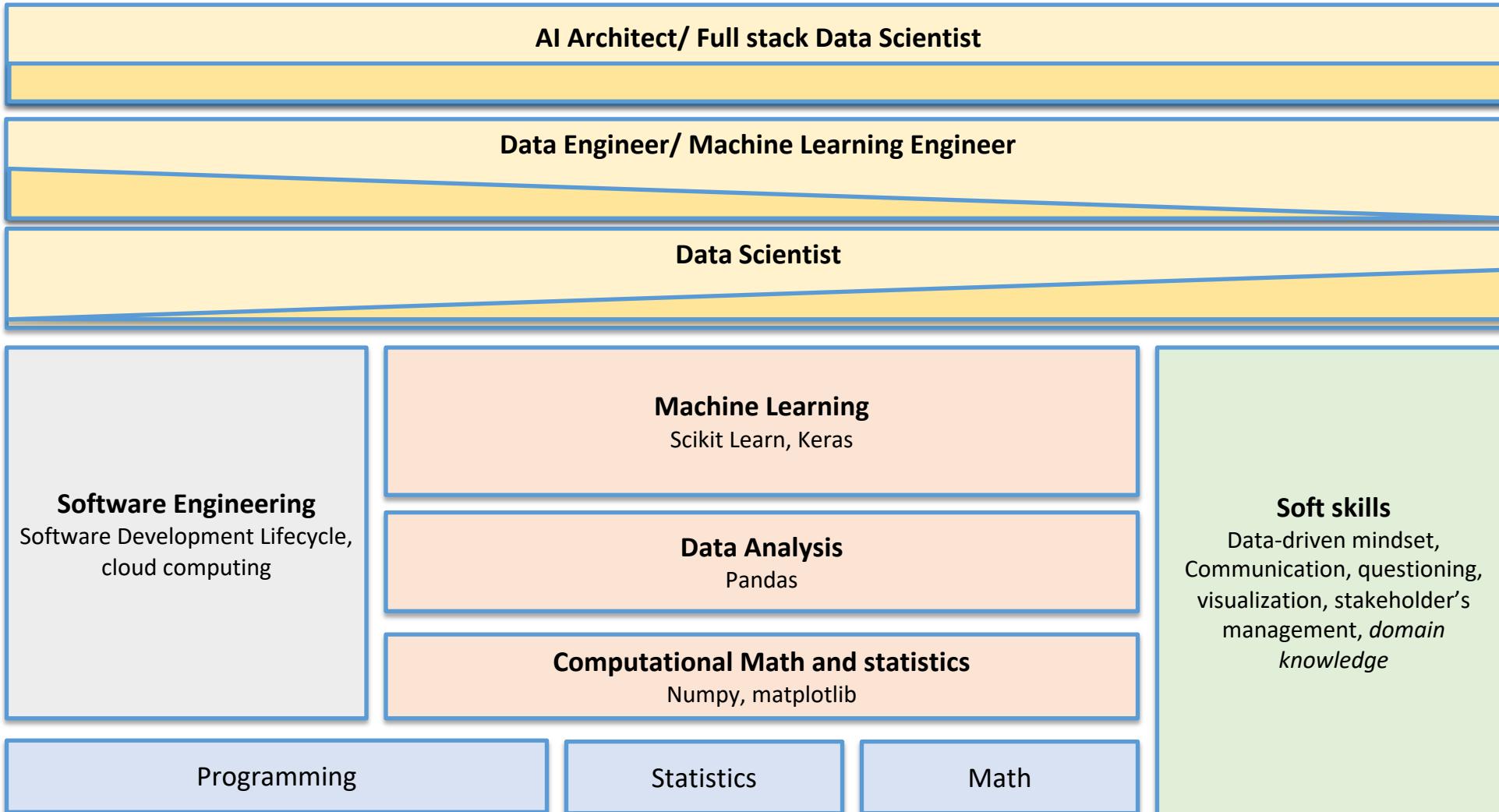
Skills of various roles in Data Science and AI

- There are a number of roles that are required to deliver Data Science/AI solutions.
- Some roles are closer to business while others are more technical.
- There is a growing demand for Data Scientists to be able to contribute directly to implementing systems in ‘production’.





Data Science skills for industry





Data scientist's responsibilities

- Identify business needs
- Analyse data
- Develop machine learning models and solutions
- Present insights
- Manage data science projects



Data scientist's responsibilities

- **Identify business needs**
 - Work with stakeholders to define business and information needs
 - Support the translation of business needs into data questions that can be addressed by available data
 - Defining what data is needed to answer the business question
- **Analyse data**
 - Collect, extract, query, clean, and aggregate data for advanced analytics purposes
 - Clean data to remove duplicate, outdated or irrelevant information
 - Perform statistical and visual analysis on data
 - Perform data validation and quality control checks
 - Mine data to identify trends, patterns and correlations



Data scientist's responsibilities

- **Develop machine learning models and solutions**
 - Build, implement, and evaluate advanced analytics problems solving using appropriate machine learning models and algorithms
 - Apply data mining techniques to investigate leads, identify patterns and regularities in data
 - Implement automated pipelines to create reproducible, scalable models
 - Identify areas of improvement of current analytics processes, products/services or models
- **Present insights**
 - Use data visualisation tools to communicate findings
 - Create clear and concise presentations reports for stakeholders
 - Design data reports and visualisation tools to facilitate data understanding
 - Assist with the development of actionable recommendations
 - Develop compelling, logically structured presentations, including story-telling of research and/or analytics findings
 - Guide stakeholders on how to act on findings
 - Use business consulting skills and frameworks in data science to assist managers and stakeholders understand the application of AI technology



Data scientist's responsibilities

- **Manage data science projects**
 - Assists in the conceptualisation of data science projects
 - Maintain project plans and status of the project
 - Provide feedback to stakeholders throughout the whole analytics lifecycle
 - Prepare documentation to outline data sources, models and algorithms used and developed



Mapping responsibilities to skills

Responsibility	Skills
Identify business needs	Applying data science in industry: <ul style="list-style-type: none">• Define projects Soft skills: <ul style="list-style-type: none">• Consulting, questioning and documenting projects
Analyse data	Core data science skills: <ul style="list-style-type: none">• Exploratory Data Analysis (EDA) and data wrangling• Visualisation• Unsupervised machine learning
Develop machine learning models and solutions	Core data science skills: <ul style="list-style-type: none">• Visualisation• Supervised machine learning (regression and classification)• Unsupervised machine learning Applying data science in industry <ul style="list-style-type: none">• Design projects• Deliver project

Responsibility	Skills
Present insights	Soft skills: <ul style="list-style-type: none">• Presenting Core data science skills: <ul style="list-style-type: none">• Visualisation• Supervised machine learning (regression and classification)• Unsupervised machine learning
Manage data science projects	Applying data science in industry <ul style="list-style-type: none">• Define projects• Design projects• Deliver project

Skills for data science

Foundational skills <ul style="list-style-type: none">• Programming for Data Science (Python)• Maths and Statistics for Data Science	Core data science and AI skills <ul style="list-style-type: none">• Exploratory Data Analysis (EDA) and data wrangling• Data Visualisation• Database access• Application Programming Interfaces (APIs)• Supervised learning (Regression and Classification)• Unsupervised learning (Clustering and Dimensionality reduction)• Deep Learning• Natural Language Processing (NLP)• Artificial Intelligence• Data science industry practices	Applying data science in industry <ul style="list-style-type: none">• Applying data science on different data structures and domains• Defining a data science project• Designing a data science project• Delivering data science project• Optimising Machine Learning model algorithms• Overall end-to-end solution• Presenting to stakeholders and obtaining buy-in• Capstone project
Soft skills Critical thinking, Questioning, Documenting, Presenting		
Learning how to learn effectively framework Minimal Viable Learning (MVL), Multimodal learning, Learn-Create cycle		



Data Science skills for industry

- **Foundational skills** that are required to learn Data Science:
 - Programming
 - Math, Statistic
 - Basic software engineering
 - Soft skills



Data Science skills for industry

- **Core** Data Science skills
 - Computational math and statistics
 - Data Analysis
 - Machine Learning
- **Complementary** Data Science skills
 - Business domain knowledge
 - Software Engineering
 - Soft skills
 - Data-driven mindset
 - Critical Thinking
 - Communication
 - Curiosity



Programming Data Science in Python

Programming is:

the **process of creating a set of instructions** that tell a computer how to perform a task.
thinking **systematically and critically**
breaking a task into steps. Examples include: a recipe, directions to a destination and mathematical problem solving.

Python has a very **active community** with a vast selection of **libraries**, especially in scientific computing, data analysis and visualisation which makes it **very suitable for Data Science**.

There are a number of tools available to support the development of Python.

Jupyter notebook has emerged as an effective way to develop and share Data Science projects.

Visual Studio Code (VSC) is an alternative for developing reusable software modules.

Programming (**computational mathematics and statistics**) can be crucial for developing deep mathematical and statistical knowledge and skills.



Why statistics is important for a Data Scientist?

- **Statistical Thinking** is an essential component of a data-driven mindset which is crucial for a Data Scientist
 - Statistical analysis must start with the appropriate **data** (sample)
 - Statistical Inference (reasoning) should start with measurement, ideally, via **controlled experiments**
 - Statistics uses samples (a small subset of the population) and therefore always has a degree of **uncertainty**.
 - Sampling must be **random, and preferably, independent**.
- The best way to learn statistics is by **experimenting with data using Python code and visualisation**.



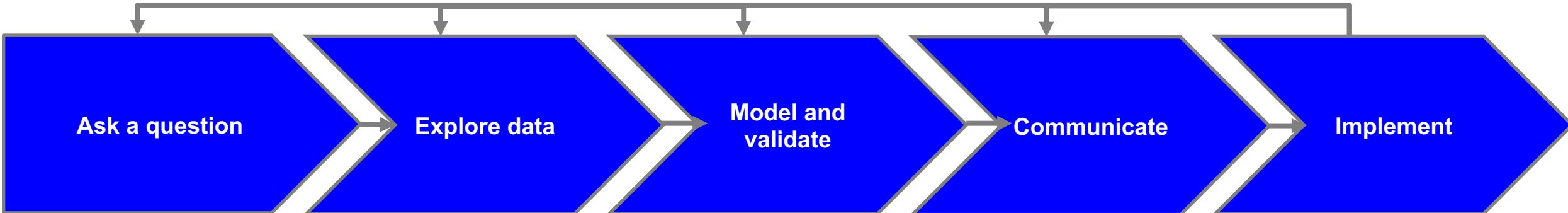
Agenda: Data Science Industry Experiences

- Introduction, definitions, purpose and objectives
- Working in industry versus research
- What do employers value and what they complains about?
- Skills required and attitude to succeed in industry
- **Data Science process in the industry**
- Case study
- Summary, conclusions and call for actions



Data Science Process

Iterate



- Ask a (business/data) question
- Define objective, strategy, value and effort budget.
- Translate business question to a data question.
- Identify and collect, clean and transform data
- Explore data
- Ascertain quality and ability to answer questions
- Feature engineering
- Select model, apply and validate
- Communicate to stakeholders and obtain buy-in
- Develop end-to-end solution
- Build, test, deploy and monitor

Most important step!

May consume large proportion of the total effort

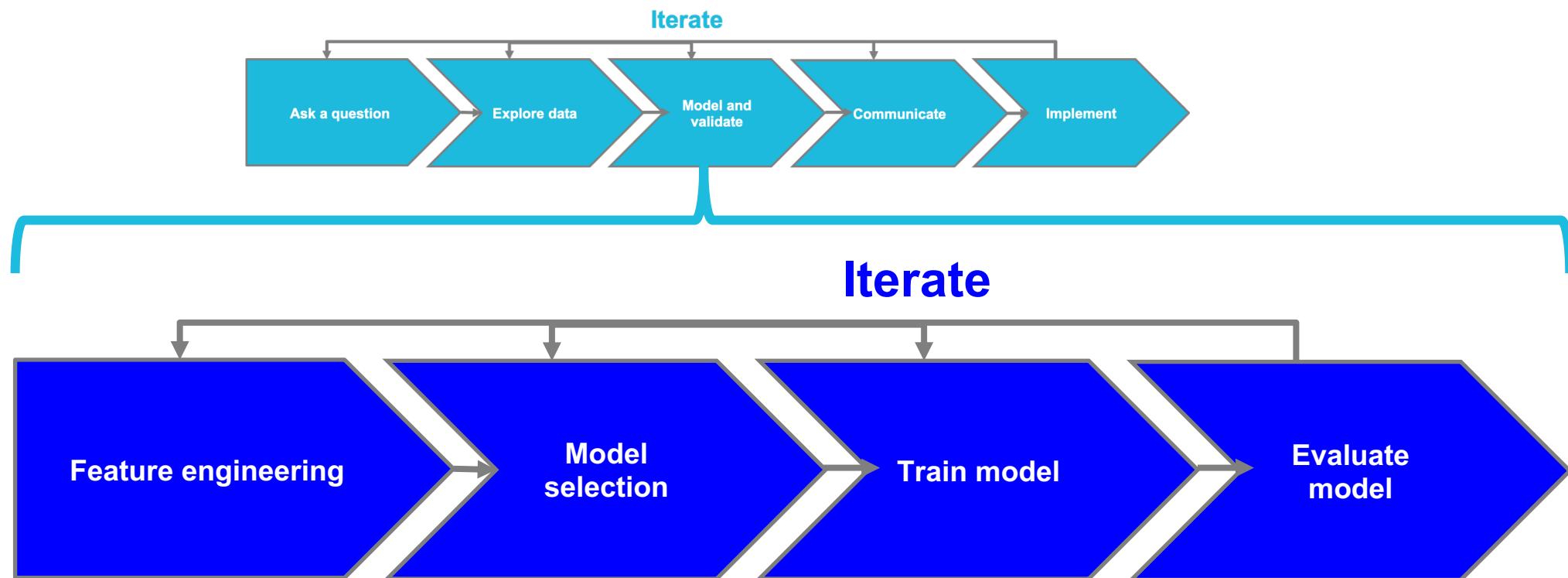
Feature engineering requires deep domain knowledge

Perform as early and as often as possible

Consider the total cost of implementation earlier in the process



Modelling Process



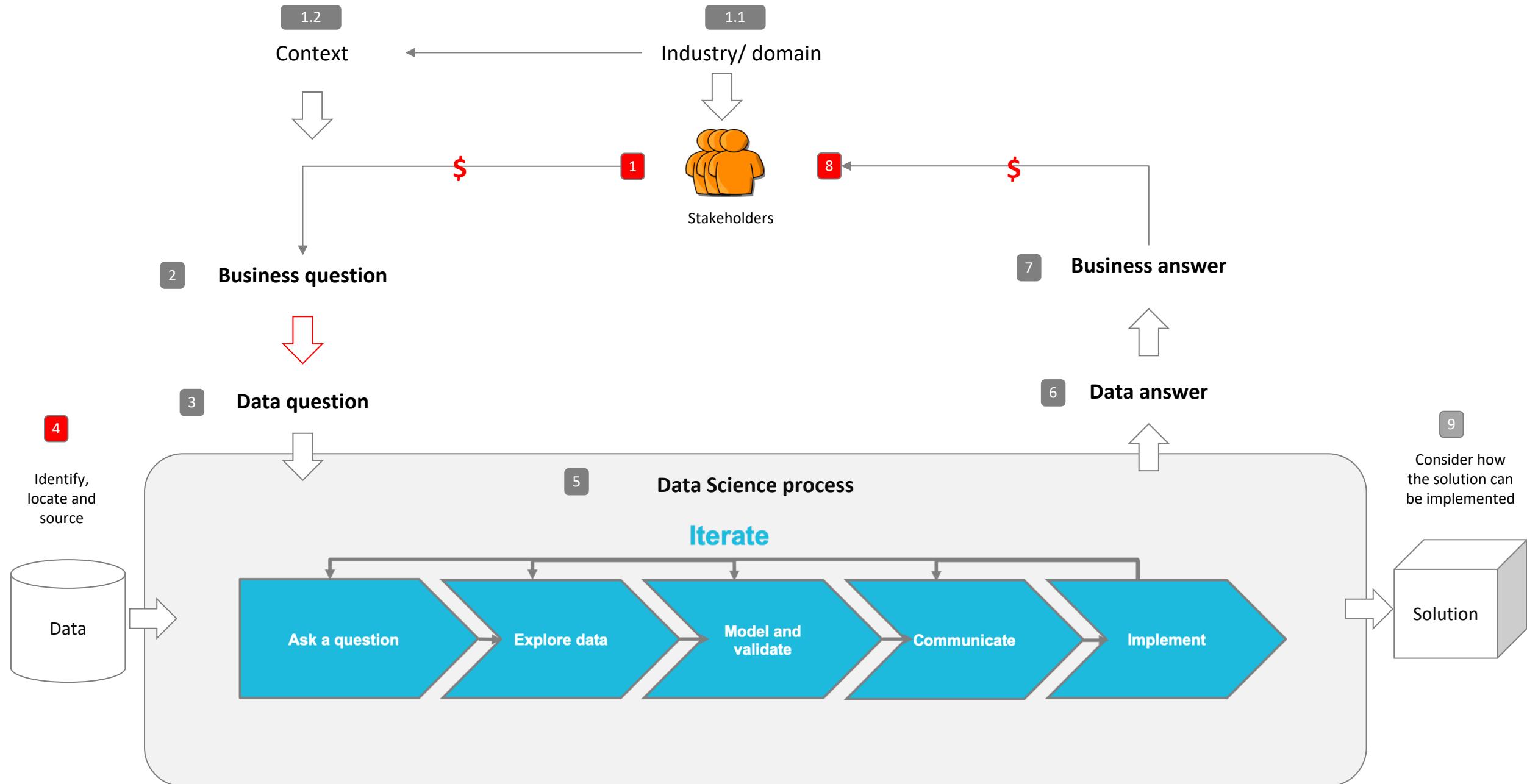
- Dimensionality reduction
- Remove noise
- Feature standardisation
- Categorical feature encoding

- Select appropriate model(s)
- Select hyper-parameters

- Train model

- Accuracy, recall

Applying data science in an industry project





Agenda: Succeeding in Industry

- Introduction, definitions, purpose and objectives
- Working in industry versus research
- What do employers value and what they complains about?
- Skills required and attitude to succeed in industry
- Data Science process in the industry
- **Case study**
- Summary, conclusions and call for actions



Industry Data Science use cases

- **Marketing, sales and customer services:** customer experience, acquisition, retention and life value.
- **Financial Services:** risk management, fraud detection and loan approval.
- **Telecommunication:** customer churn.

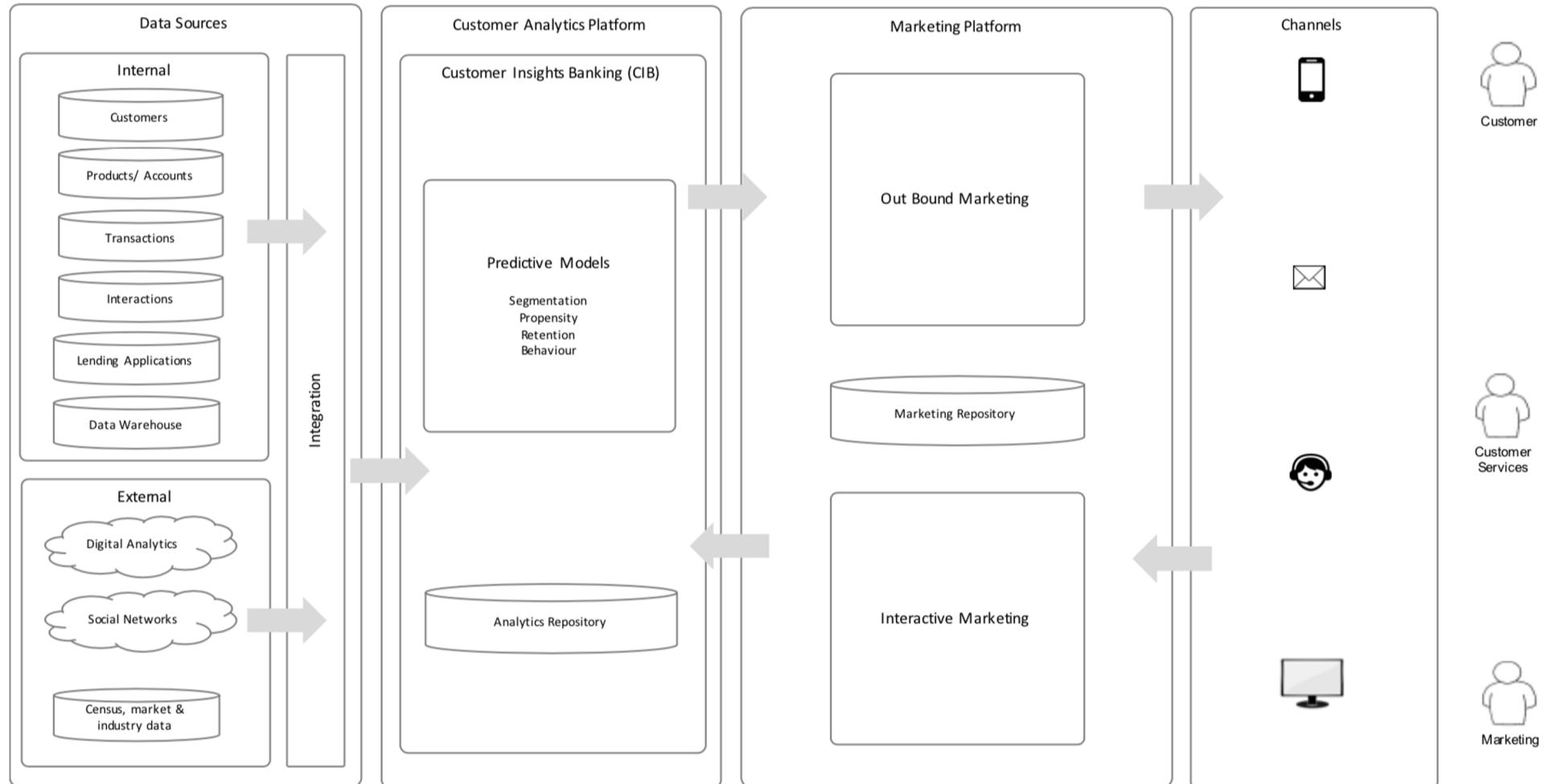


Case study: Home loans marketing

- **Use case:** The primary objective of the project is to develop models to **identify prospective customers** that are likely to take a new **home loan** or re-mortgage their existing loan with the bank within a set time horizon (up to 6 months).
- **Approach:** The models have been created and evaluated based on the **2-year historical data**.
- **Success criteria:** The model was tested on previously “unseen” customer data and successfully predicted customers who did purchase a mortgage.

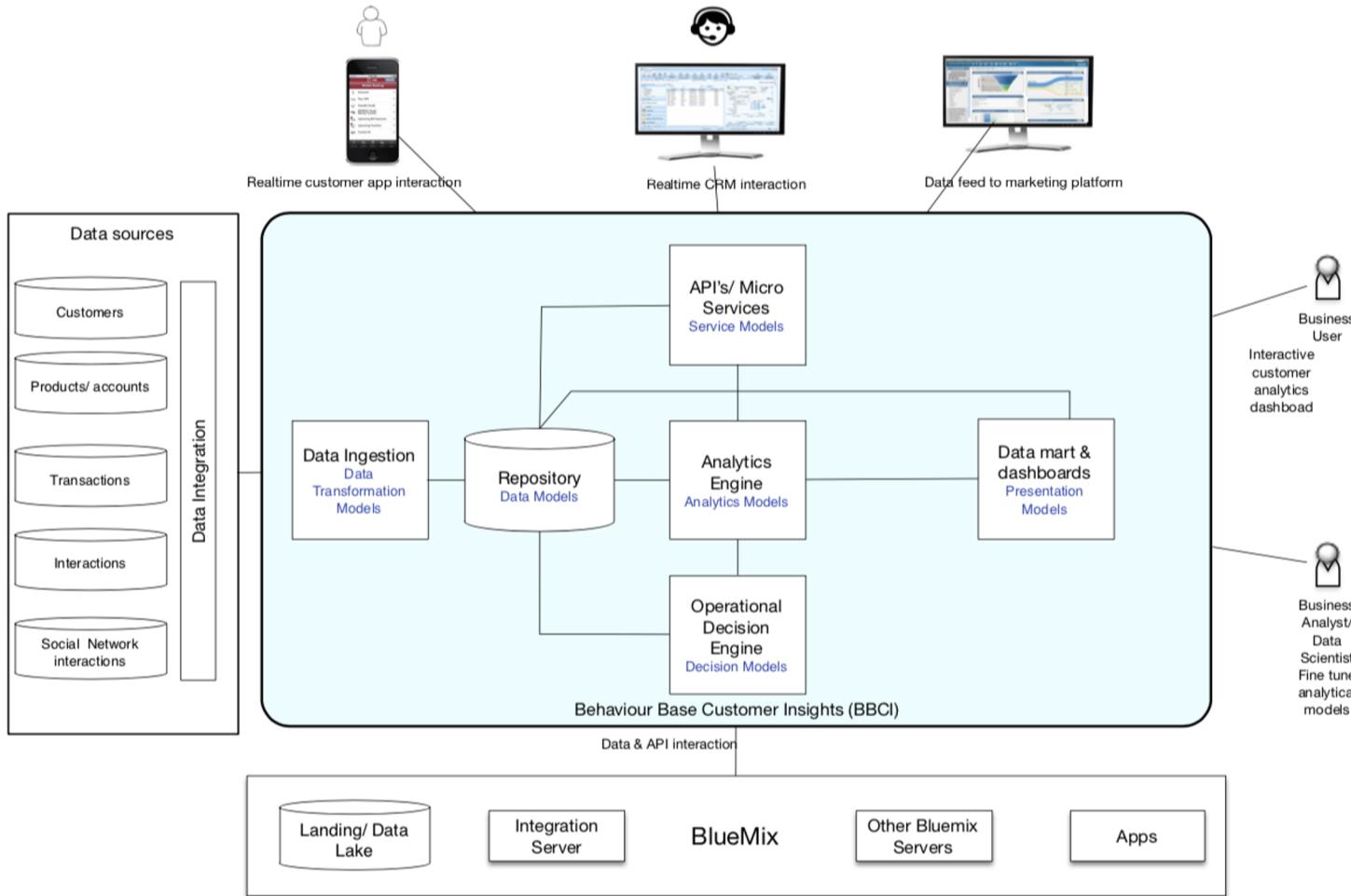


Case study: Home loans marketing





Case study: Home loans marketing

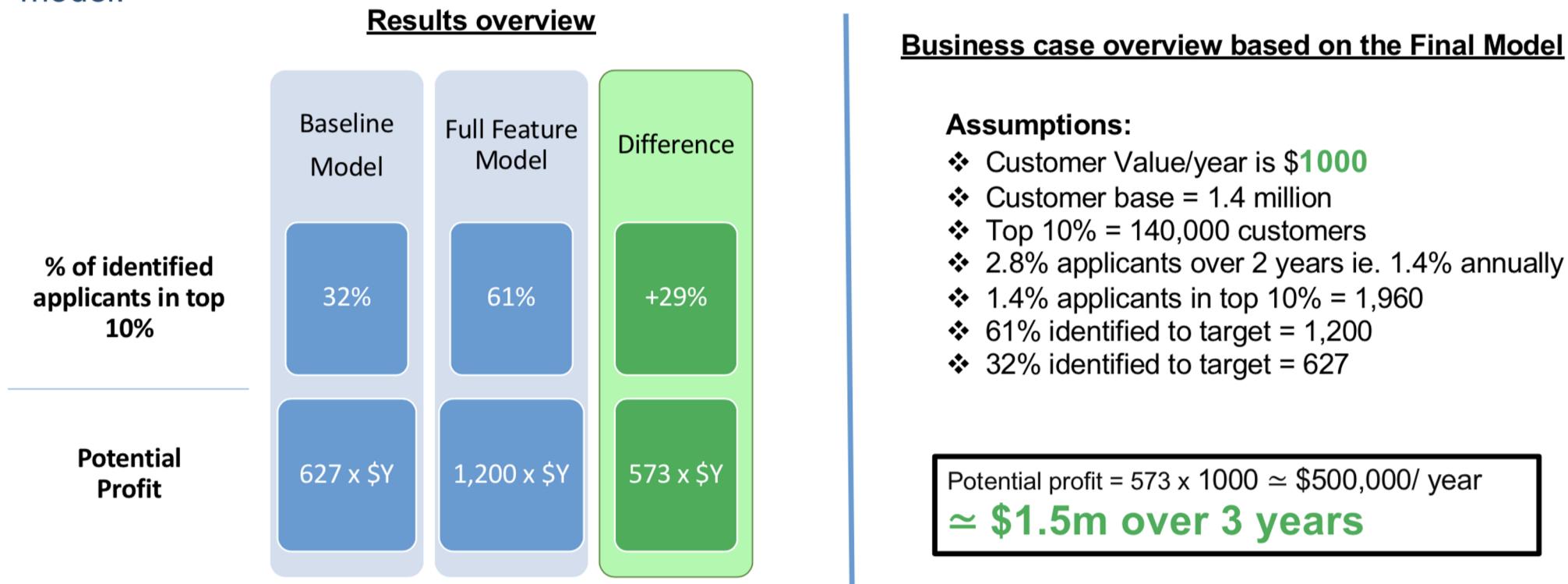




Case study: Home loans marketing

Results comparison and business case overview

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





Agenda: Succeeding in Industry

- Introduction, definitions, purpose and objectives
- What do employers value and what they complains about?
- Skills required and attitude to succeed in industry
- Data Science process in the industry
- Case study
- **Summary, conclusions and call for actions**



Summary, conclusions and call for action

- **Summary**
 - I have shared with you my industry experience and my views on how you could succeed in the industry by understanding what is required by employers.
- **Conclusions**
 - It is **not enough** to develop the Data Science '**technical**' skills, you need soft skills so you can apply these skills to deliver **value**.
 - To enable you to effectively work in industry, you need to:
 - Understand which are the most **important skills required in industry**
 - Discover your dream job and research what skills are needed for this job
 - Master a **small number of skills/tools** and
 - Decide on your **focus areas** including **domain (industry)**.



Summary, conclusions and call for action

- Call for action
 - *Start now!*
 - Identify/ refine your focus areas and skills,
 - learn,
 - create,
 - Identify gaps,
 - Iterate.



Questions?



End of Presentation!