RISE FOR FINANCIAL SERVICES

Project Brief – M2 Assessment

2024

Overview of Assessment

- In the Capstone, you will be assessed on the following Technical Skills & Competencies (TSCs), as stipulated under the SSG Skills Framework for the role of an AI/ML Engineer:

  1. Data Analytics

  2. Data Ethics

Data Analytics Ability Statements

- You will be assessed on the following Ability statements for the **Data Analytics** TSC:

Data Analytics TSC: Ability Statements	
A1	Apply predictive data modelling techniques to identify underlying trends and patterns in data using statistical computing tools, methods and procedures
A2	Identify patterns across multiple data sets to derive insights
A3	Develop prototype algorithms and proof of concept demonstrations
A4	Make decisions about which patterns are meaningful, and which to further analyse
A5	Assemble data aggregations to build data models to help test problem hypotheses
A6	Use machine learning techniques to gain new insights from data
A7	Mine data to find relevant insights to develop ongoing improvements
A8	Assess the business insights presented to determine impact of insights on organisation
A9	Manage the creation of interactive visualisations of data and data study outcomes
A10	Use industry standard tools and techniques for data visualisation in line with organisational procedures

Data Ethics Ability Statements

- You will be assessed on the following Ability statements for the **Data Ethics** TSC:

Data Ethics TSC: Ability Statements	
A1	Adhere to the organisation's code of conduct and the PDPA in the collection, use, retrieval and disposal of personal data
A2	Accept responsibility for own behaviour
A3	Recognise and report potential breaches in code of ethics
A4	Apply decision-making process to resolve ethical dilemmas

Corporate banking serves a wide spectrum of clients..

Corporate Clients
- Small
- Medium
- Large
- Multinationals

Institutional Clients
- Financial Institutions
- Public Sector
- Charities
- Principal Investors & Private Equity

With different needs depending on:

Size (annual turnover)

Degree of internationalization

Degree of maturity

Comprehensive range of products to address client needs




Manage cash / working capital | **Finance business** | **Manage risk** | **Advise on M&A**

Transactions	Deposits & investments		Lending	Trade Finance	Global markets			
Payments	Deposits	Investments	Short term	Receivables	Equity	Vanilla	M&A advisory	
Online banking	Demand deposit ac	Mandates	Working cap loan	Factoring	IPO	Derivatives	Buy side	
Liquidity and information	Multi-currency ac	Mutual funds	Revolv credit fac.	Invoice discounting	Follow on	Options	Sell side	
Value-added services	Money market ac		Long term	Payables	Debt	Swaps		
	Term deposits		Syndicated loans	Supply chain fin	Bonds	Futures		
			Bilateral loans	Dynamic discount	ABS/CMBS			
			Bridge loans	Internat. trade	EMTN			
			Project specific	Letters of credit				
			Project finance	Int guarantees				
			Acquisition finance	Open account				
			Asset lending					

Focus of project

Churn Prediction in Banking

- **A churn prediction model** predicts customers that are likely to **cancel** a product/service subscription (Hard Churn) or **decline** in engagement (Soft Churn)
- The full cost of churn includes both **lost revenue and marketing costs** involved with replacing those customers with new ones

Target Customer Profile is created out of two broad sources:

Interaction with the firm like average balances, # of transactions, # of product holdings,communication pattern, etc.

Focus of project

External information such as Share of Wallet, customer risk flags, firmographics, competitors data, macroeconomic, bureau data, etc.

Context

You are a data analyst working in one of the largest banks by assets in Southeast Asia, where the bank is also the largest payment bank in terms of transaction value.

The bank intends to analyse and prevent the CASA portfolio attrition of corporate customers by identifying declining customer relationship (i.e., soft churn) 3 months in advance.

Challenges faced

- Declining year-on-year CASA balance, resulting in low profits across segments of customers. This is coupled with the bank losing market share to competitors

Objective

Build a predictive model to identify corporate customers who are likely to soft churn

CASA = Current Account/Savings Account

Predictive model helps retain churners on 3 aspects

1. Detect churn risk 3 months in advance

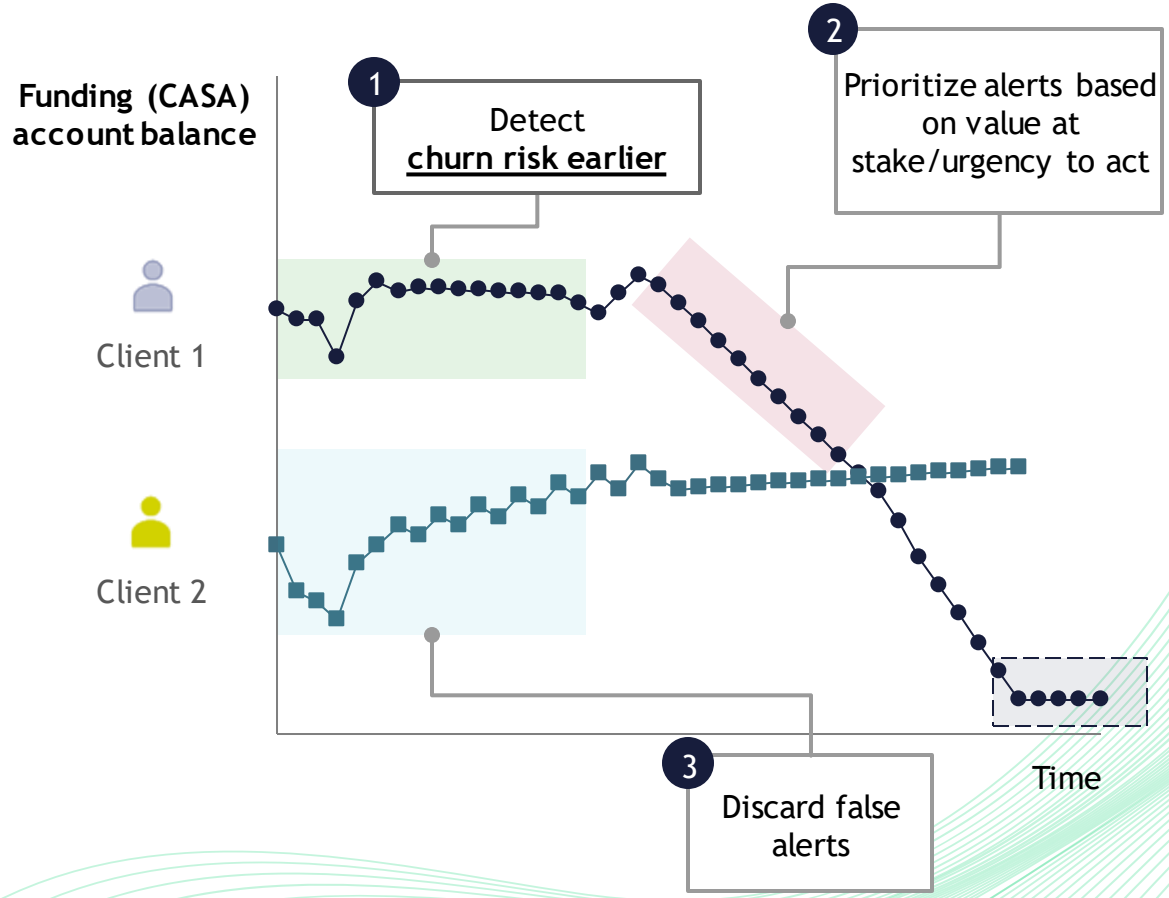
2. Recommend specific commercial actions, customized using churn drivers

   E.g., Customer predicted as churn:
   Churn drivers: decrease in revenue or decrease in # products
   Actions: provide incentives e.g., "maintain a min balance of $X across 3 accounts and get 10 BPS lower interest rate on lending account"

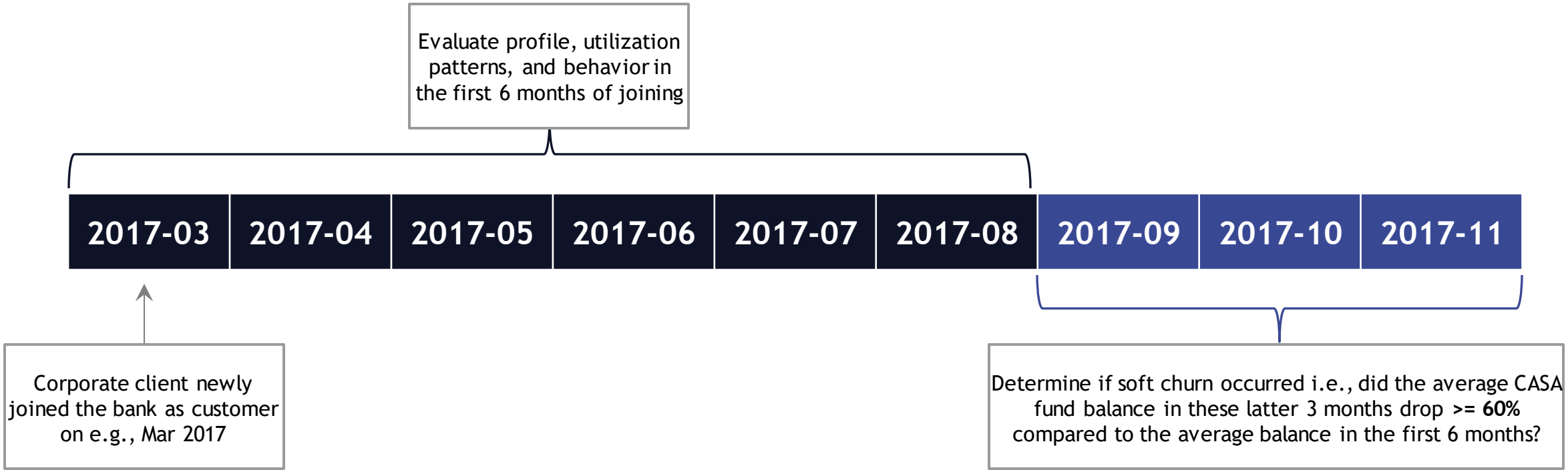
3. Limit false alerts as much as possible to focus on most valuable/urgent cases

BPS = Basis Points

Identifying **soft churn** by predicting **decline in relationship** in advance



Soft Churn Prediction for New Joiners



Detailed Steps for Model Development

1 Data Preprocessing

- Combine different datasets
- Define target Population (i.e., soft churners)
- Missing value treatment

2 EDA & Feature Engineering

- **Exploratory Analysis**
  - Bivariates
  - Statistics
- **Feature Engineering and Selection**
  - Correlation Matrix
  - VIF
  - Importance charts

3 Model Creation and testing

- Train-test split
- Cross validation folds
- Examples of models that can be built
  - Logistic
  - Random Forest
  - Ensemble of models

4 Model Testing and Tuning

- Out of Sample testing
- Model Tuning such as Regularization in Logistic Regression, hyperparameter tuning in Random Forest etc.

Linking Analytics to Business

- Use the combined customer data for target preparation, model development and tuning (including train-test split and cross-validation)
- Use your champion model to predict churn probability for submission
- Use final values to find and focus on the most valuable likely churners (e.g., based on average balance or transaction volume)

Datasets Available

File Name (CSV)	Description and Comments
Funding_2017_2019	Account and onboarding firmographics information about the corporate customer
Lending_2017_2019	Active loan account information and outstanding balances
Transactions_2017_2019	Customer transactions from their CASA accounts, providing information about the customer's interactions with the bank
Feedback_2017_2019	Customer feedback on bank's services and products

Data Dictionary

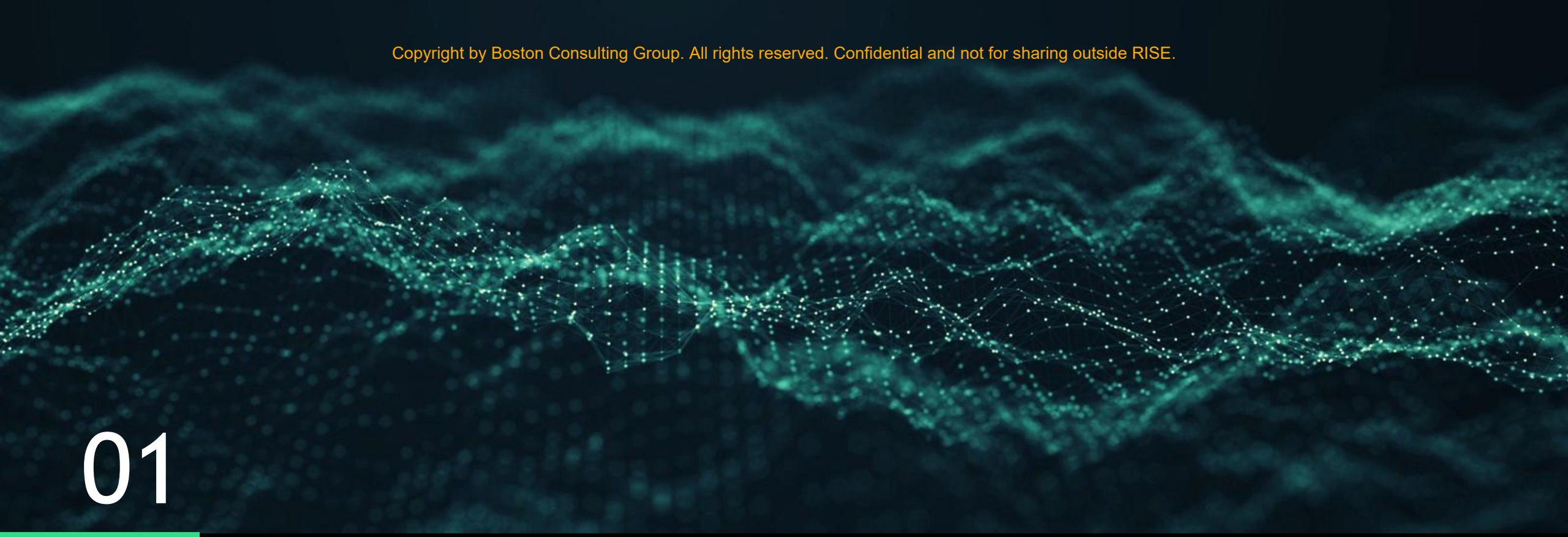
- The data dictionary for the datasets can be found in the attached Excel document

Microsoft Excel
Worksheet

- Alternatively, you can look for the `**Data Dictionary - M2 Assessment.xlsx**` file in the project folder

Assessment Objectives

- Apply the techniques involved in end-to-end predictive modelling (e.g., data preparation, feature engineering and selection, model building and tuning, model evaluation etc.)
- Interpret model findings and translate them into business insights
- Demonstrate the application of important Data Ethics policies and procedures based on relevant guidelines and frameworks

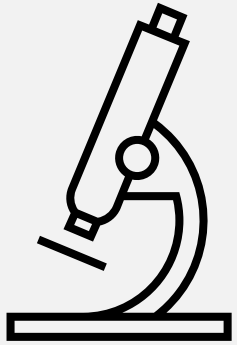


01

Data Analytics - Predictive Modelling

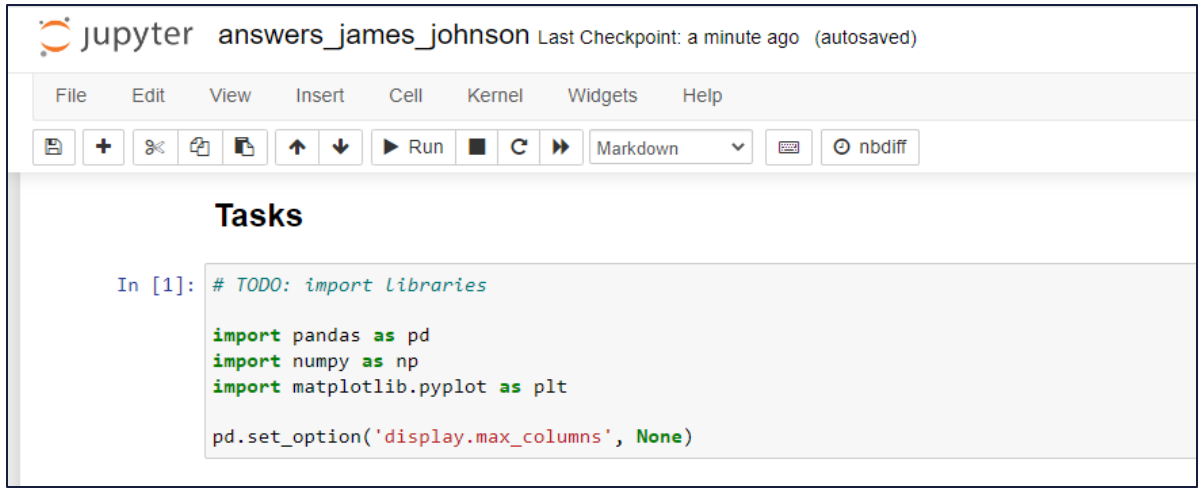
Task

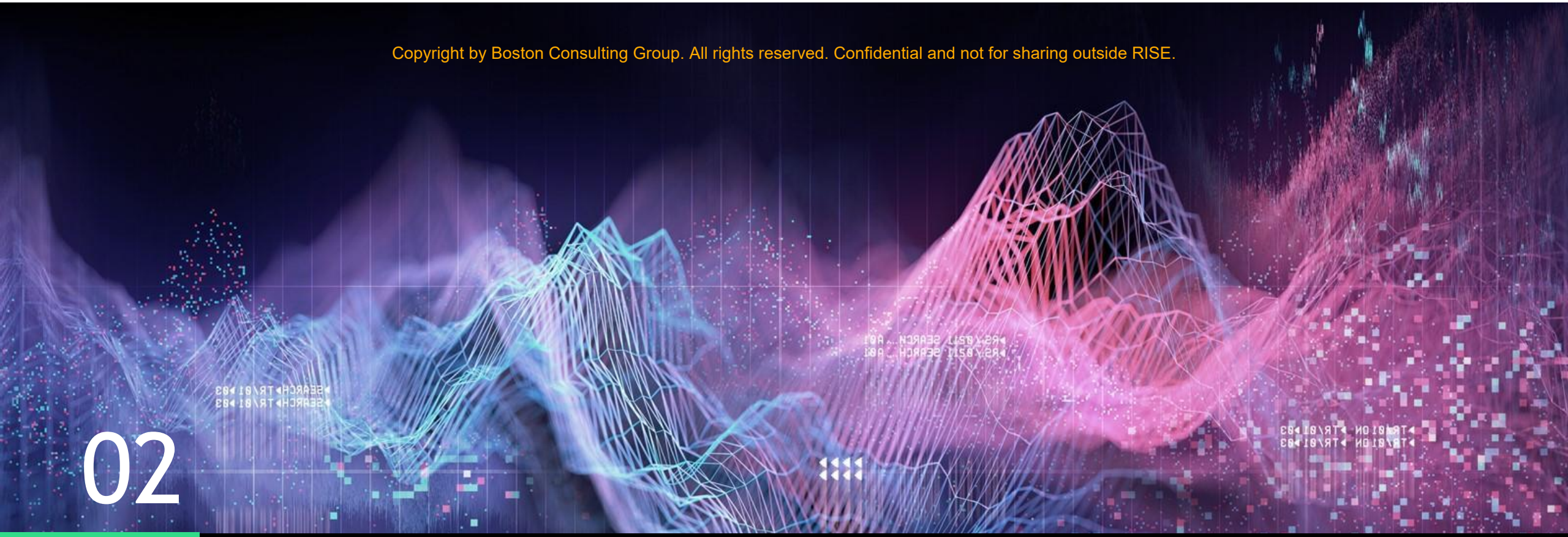
- Launch the Jupyter notebook titled '**Notebook - M2 Assessment (Learner).ipynb**'
- In this project, we will explore the practical implementation of predictive modelling techniques on a corporate banking dataset in an end-to-end manner



Expected Output – Jupyter Notebook

- You should code and save your answers in the Jupyter notebook provided
- Ensure that the Python codes in the notebook can be executed without any errors
- Save the .ipynb file with a filename that includes your name e.g., *m2_assessment_james_johnson.ipynb*





02

Data Ethics

Data Ethics Context

You are a Data Protection Officer who oversees all Data Ethics and Governance efforts. You work in the same bank as the data analyst who is building the predictive model to identify corporate customers who are likely to soft churn.

In the course of building the predictive model, you identified multiple scenarios where the data analyst may have failed to adhere to proper Data Ethics policies or procedures.

Objective

For each scenario, highlight any breach of Data Ethics policies or procedures and explain the breach with reference to specific data ethics frameworks and guidelines.

For each breach identified, provide recommendations to assist the data analyst to adhere to good Data Ethics practices.

Scenarios on Breach of Data Ethics

Scenario 1:
The data analyst has access to personal data of the corporate customers that are not relevant to the analysis. The data analyst is curious to know more about the personal lives of the customers and decides to look into their personal data.

Scenario 2:
The data analyst comes across a pattern in the data that suggests the presence of potential fraud within the bank's system. The analyst is unsure if this should be reported to the management as it could potentially harm the bank's reputation.

Scenario 3:
The bank has a policy of deleting personal data of customers after a certain period of time, but the data analyst notices that some personal data is still being stored beyond the mandated time limit.

Scenarios on Breach of Data Ethics

Scenario 4:
The data analyst discovers that there is an error in the data used for the analysis, which could potentially lead to incorrect results. The analyst decides to ignore the error as it would take too much time to correct it.

Scenario 5:
The bank's system was recently breached and personal data of customers was compromised. The data analyst is tasked with investigating the breach and identifying the cause. The analyst discovers that the breach was caused by a weak password policy and inadequate security measures.

Scenario 6:
The data analyst is approached by a colleague from a competitor bank who offers a large sum of money in exchange for the bank's customer data. The data analyst is tempted to accept the offer.

Scenarios on Breach of Data Ethics

Scenario 7:
The data analyst discovers that the analysis conducted on the corporate customers' data has led to a biased outcome, favoring a particular group of customers. The analyst is unsure of how to proceed.

Scenario 8:
During the course of the analysis, the data analyst discovers that one of the factors that may be leading to customer churn is the bank's recent decision to increase fees for certain services. The analyst recommends to reduce or eliminate these fees in order to retain customers.

Scenarios on Breach of Data Ethics

Scenario 9:
As part of the analysis, the data analyst discovers that a large percentage of customers who are at risk of churning are in financial distress, and may not be able to continue doing business with the bank due to financial constraints. The analyst recommends to target these customers with promotions or offers in order to retain their business.

Scenario 10:
The data analyst discovers that certain customers who are at risk of churning are also in arrears on their loan payments. The analyst recommends to prioritize retention efforts for these customers over other customers who are not in arrears.

Expected Output – Microsoft Word Document

- For each scenario on Breach of Data Ethics, explain the breach with reference to specific data ethics frameworks and guidelines where possible.
- For each breach identified, provide recommendations to assist the data analyst to adhere to good Data Ethics practices.
- Document all the above in a Microsoft Word file.
- Save the MS word file with a filename that includes your name e.g., *m2_assessment_james_johnson.docx*

03

Project Logistics

Individual Q&A

1. Prepare for a 15-min individual Q&A session with the assessors.
2. The assessor will conduct oral questioning based on the assessed ability statements. They can be any ability statement, from any TSC.
3. Be ready to answer any questions that the assessor might have on your submitted work as well.
4. You are allowed to refer to any learning materials during the individual Q&A session.

Assessment Timeline

Date	Time	Session Description
13 May 2024, Monday	10:00am – 11:00am	M2 Assessment Briefing
16 May 2024, Thursday	1:00pm – 5:00pm	M2 Assessment Clinic
20 May 2024, Monday	1:00am – 5:45pm	M2 Assessment Clinic
22 May 2023, Wednesday	By 11:59pm	M2 Assessment Final Deliverables Submission
23 May 2024, Thursday	9:00am – 5:00pm *(exact time to be advised closer to date)*	M2 Assessment Individual Q&A (15-min per individual)

We suggest using these sessions for a mixture of trouble-shooting with trainers, and group work

Individual submission

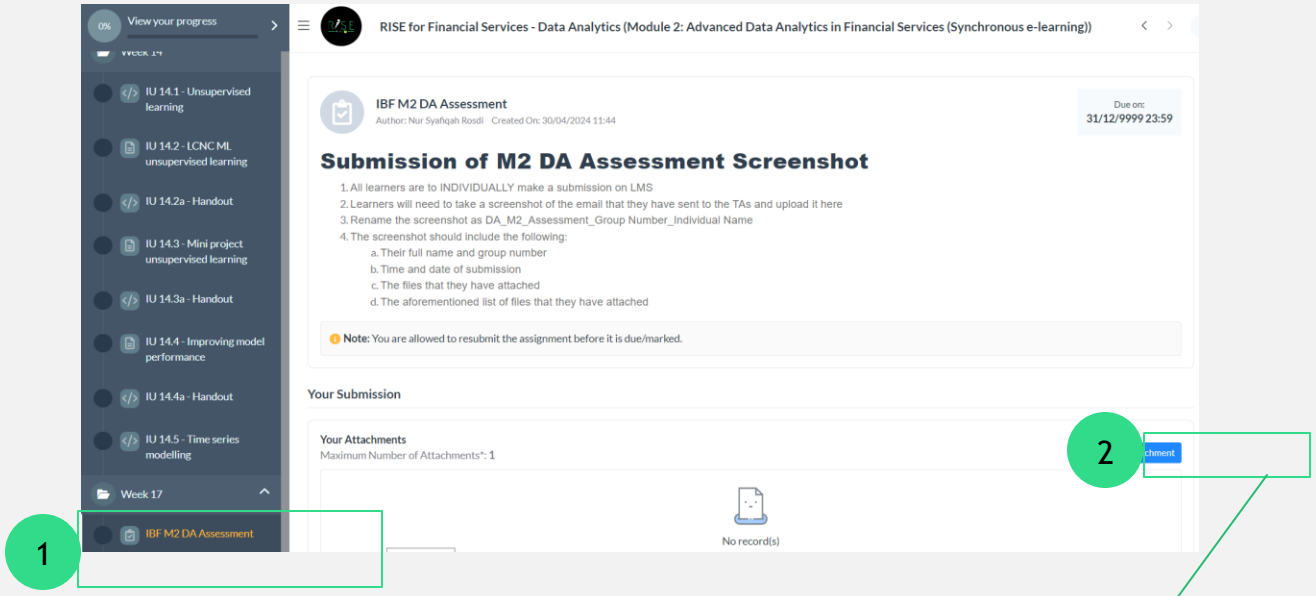
1. BCG is required to retain evidence of assessment submission, please refer to the following instructions:
2. If you are unable to submit all your files in one single email due to file size constraints, do it in multiple emails as needed.
3. Take a screenshot of the email to be used as submission on LMS (refer to next step)

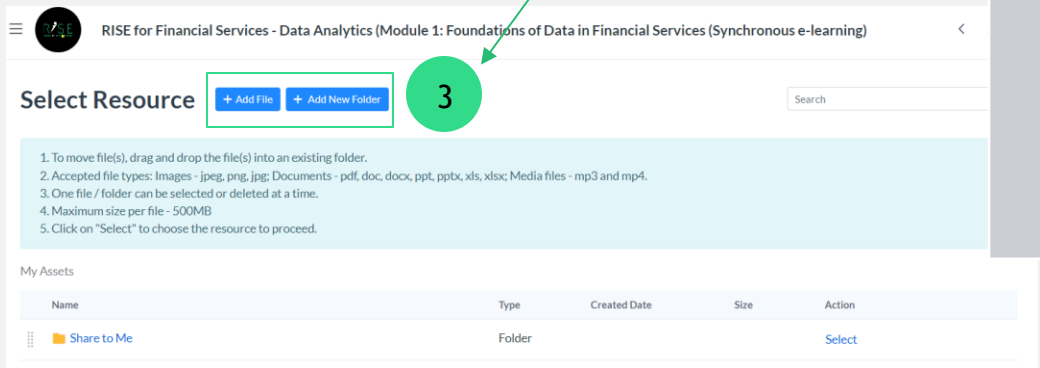
	Deliverables submission
Deadline	22 May, Wed, 11.59PM
Email subject	DA M2 Assessment_Group number_Individual Name • E.g. DA M2 Assessment_Group 4_Nur Rosdi
File naming conventions for (multiple) attachments	Group Number_Individual Name_File number and name • E.g. Group 4_Nur Rosdi_01 Workplan.doc • E.g. Group 4_Nur Rosdi_02 SQL script.sql • E.g. Group 4_Nur Rosdi_03 visualisation.pbix
Who submits	Each individual to submit all files (e.g., SQL, Jupyter Notebook) via email to your TA, followed by LMS screenshot submission.)

Submission of deliverables on LMS (1/2)

Note:

- All learners are to **INDIVIDUALLY** make a submission on LMS
- Take a screenshot of individual email submission in previous step
- Rename screenshot as DA M2 Assessment_Group number_Individual name and upload to the correct IU





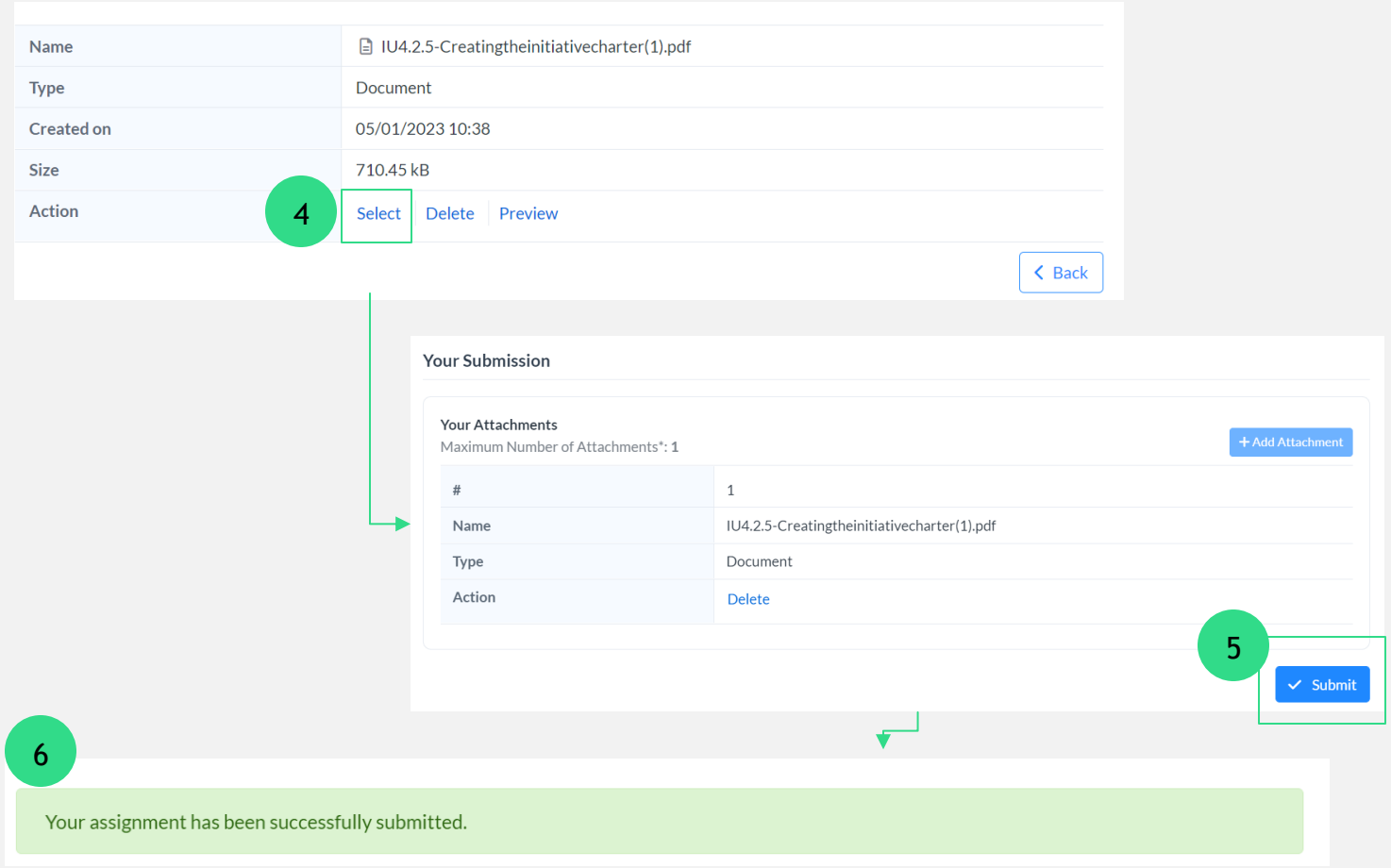
Steps

1. Select the IU on the course tab -
2. Click "Add Attachment"
3. When the new page pops up, click "Add File". Kindly upload your screenshot file

Submission of deliverables on LMS (2/2)

Note:

- All learners are to **INDIVIDUALLY** make a submission on LMS
- Take a screenshot of individual email submission in previous step
- Rename screenshot as DA M2 Assessment_Group number_Individual name and upload to the correct IU



Steps

4. Scroll down and click "Select"
5. You will be redirected back to the main submission page. Scroll down here and click "Submit"
6. You should see the pop up that your assignment has been successfully submitted

RISE FOR FINANCIAL SERVICES