DSI TEAM 2 PROJECT — MASTER OPERATING GUIDE

Links

Project Information

Project Name: Bank Marketing
Industry: Finance
Project Showcase Date: March 7, 2026

Shared Links

GitHub Repository TO CLONE: https://github.com/DzeDze/ds_team2_bank_marketing

Slides: 🟨 DSI Showcase - March 2026

Shared Drive: 📁 DSI Team Project

Meeting Notes: 📄 **Meeting Notes - DSI Project Team 2**

Team

- Kerensa Wong
- Vince Nguyen
- Kateryna Makieieva
- Jarsmeka Maneckarase
- Morolake Nwokoro

Requirement:

- Every member creates a PR
- Every member reviews & merges another PR

Dataset

Dataset Name: Bank Marketing: https://archive.ics.uci.edu/dataset/222/bank+marketing

Key Variables:

- Age, Job, Martial, Education, default, balance, housing, loan, contact, day_of_week, month, duration, campaign, pdays (last contacted), previous, outcome, y (subscribed?)

Business Case/Questions

1. **Possible Industry Problem:** What real-world problem or opportunity exists?
2. **Business Motivation:** Why would an employer/stakeholder care?
3. **Intended Stakeholders:** Who benefits?
   a. Executives
   b. Policy makers
   c. Customers
   d. Operations teams
   e. Public users
   f. Other:
4. **Expected Business Impacts:** What action could someone take after seeing this project? Example: Reduce cost, Improve efficiency, Predict outcomes, Optimize decisions

Questions must:

- Use dataset variables
- Produce measurable insight
- Support business decision-making
- Can leverage other datasets

BRAINSTORMING - Compiled: 📄 Meeting Notes - DSI Project Team 2
Jarsmeka
  a. **Business Problem:**
      b. How can the bank improve the effectiveness of marketing campaigns?
      c. How to run effective marketing campaigns using the banking system as a use case?
      d. **Business Motivation:** Save on resources, success rate, possible pivot on future marketing strategies
      e. **Questions:**

      i.   Demographic (helps us understand the audience + ideal targets to save bank resources)
1. Which age group is most likely to subscribe for a term deposit?
2. Does a customer's job influence purchasing trends?
3. Does a customer's education level impact subscription?
4. How does marital status impact subscription behaviour?

      ii.   Financials (helps us understand the audience + ideal targets to save bank resources)
1. Does how rich a person is impact the likelihood of subscribing?
2. What can be learned about customers with existing loans? Different rates? Age? Etc.
3. How many loans are for personal loan vs housing loans?

      iii.   Campaign Specific (helps us create better campaigns)
1. Does how we contact the customer matter? Cell vs tele/landline, etc.
2. What months/days have the lowest vs highest subscription success?
3. How many customers were contacted in the last year, month, week? How many were not contacted at all?
4. Can we look at the success rate of the previous campaign to predict future ones?

f. **Audience:** Business Decision Makers, Bank Marketing Team, Bank Executives
g. **Expected Impact:** All four - improve efficiency, predict outcomes, optimize decisions, and reduce cost (although the cost is not known)

Vince

h. **Business Problem:**
i. How can the bank optimize **who to contact, when to contact, and how often to contact** to increase subscription rates?
j. **Business Motivation:** Improve targeting precision, reduce unnecessary repeated contacts, increase subscription rate, improve operational efficiency of call campaigns.
k. **Questions:**

      i.   Demographic (helps us understand the audience + ideal targets to save bank resources)
1. Are retired customers more likely to subscribe than working professionals?
2. Does having credit in default reduce subscription probability?

      ii.   Financials (helps us understand the audience + ideal targets to save bank resources)
1. Does higher account balance increase likelihood of subscription?
2. Are customers with housing loans less likely to subscribe?
3. Does having a personal loan affect subscription behavior?

4. Is there an interaction between balance and loan status?
5. Are high-balance customers under-targeted or over-targeted?
iii. <u>Campaign Specific (helps us create better campaigns)</u>
1. Is there an optimal number of contacts in a campaign? (Is there a point where contacting customers becomes ineffective?)
2. Does subscription probability decrease after too many calls?
3. How does the outcome of a previous campaign affect current success?
4. Are customers with previous successful campaigns significantly more likely to subscribe again?
5. Are customers previously contacted but unsuccessful unlikely to convert again?
6. How does call duration relate to subscription?
7. Does time since last contact (pdays) influence success probability?
8. Do certain customer groups respond better in specific months?
9. Are certain days of the week more effective?
10. Is there evidence of seasonal patterns in subscription behavior?

- <u>Segmentation & Targeting</u>
1. Is there an optimal number of contacts in a campaign? (Is there a point where contacting customers becomes ineffective?)
2. Can we identify low-probability segments to avoid targeting?
3. What combination of age, job, balance, and loan status yields the highest conversion rate?
4. Can customers be ranked by likelihood of subscription?

Kerensa

1. Are existing customers more likely to subscribe than new ones?
2. Can subscription likelihood be linked to other products (e.g. loans)?
3. What if the phone is not the only channel that customers are contacted (e.g. in-person in branch/email etc.)? ←may require other datasets
4. Does time of day (morning vs evening calls) influence success rate? ← no info
5. Does the combination of age+job+loan status yield stronger predictive power than single variables? (identify which variables most strongly influence subscription)
6.

Risks and Unknowns

(required Week 1 proposal)

- Missing values?

- Bias?
- Limited coverage?
- Privacy concerns?

BRAINSTORMING: Compiled 🗒 Meeting Notes - DSI Project Team 2

Jarsmeka
l. Risks
  i. Likelihood that data is heavily skewed with no subscriptions vs yes subscribers
  ii. May be unable to pull other datasets due dataset being based in Portugal
  iii. Data is older than 2013 - may not be very accurate for predictions 13 years later

m. Unknown
  i. Cost of campaign
  ii. Profit generated
  iii. Why don't customers subscribe (reasoning, etc.)?
  iv. How long do they subscribe?
  v. If loan is even utilized
  vi. Other reasons outside of demographic that lead to subscribing
  vii. Any changes to the campaign when calling

Vince
n. Risks
  i. Class imbalance

o. Unknown
  External economic factors

Kerensa
Risk
- Data Quality: "Unknown" variables in fields: "job", "education", "contact", "poutcome"
  - In "poutcome" there is also "other" variable which doesn't not have further explanation in it
- Outliers: extreme balances or extreme long call durations could skew results (need further handle)
- Dataset refer campaigns from a Portuguese Bank, the result might be applicable in Canadian context
- Predictive models may perform well on historical data but may fail to assist current campaign given it is 13 years ago (with financial crisis, COVID etc people psychology and behaviour might change)

Unknowns:
- Potential important variables aren't included (e.g. income, digital engagement, geographic location)
  - Whether calls made in morning or evening is not known

Analysis Approach

How will we answer our questions?

- Data Cleaning
- EDA
- Visualization
- Statistical Testing
- Regression
- Classification

Tools:
Python / SQL / Excel / Tableau / Other

BRAINSTORMING
Jarsmeka
p. All of the listed above, potentially classification as well - Python will be used primarily

Kerensa
Visualisation: Heatmap (e.g. show conversion rates by age group vs job category),

---

Repo Structure

```
data/
├── raw
├── processed
├── sql
experiments/
models/
reports/
src/
README.md
├── PROJECTPROPOSAL.md
.gitignore
```

Remove unused folders before submission.

GIT Workflow Plan

- No commits to main
- Feature branches only
- Small commits
- Pull requests required
- Clone repo, do not fork!

Branch naming:

Reproducibility Plan (Mandatory)

Anyone must recreate results from scratch.

Team must include:

- setup instructions
- environment file
- clean scripts
- raw + processed data
- execution order

Reproduction test owner:

Visualization Output

Output	Question Answered	Owner	Tool
Visualization 1			
Visualization 2			
Visualization 3			

Project must generate new insights.

Milestones Checklist

Week 1 — Proposal (Graded)

- ☐ Business motivation
- ☐ Dataset chosen
- ☐ Risks identified
- ☐ Analysis plan
- ☐ Roles assigned
- ☐ Questions finalized

Week 2 — Final Project

- ☐ Analysis complete
- ☐ Visualizations finalized
- ☐ README complete
- ☐ Repo reproducible
- ☐ Slides ready

Cleanup

Hidden work completed:

- ☐ Cleaning data
- ☐ Debugging
- ☐ Merge conflicts
- ☐ Documentation
- ☐ Environment setup

README File

Main README must contain:

- ☐ Project overview
- ☐ Business objective
- ☐ Dataset explanation
- ☐ Methods used
- ☐ Key findings
- ☐ Business recommendations
- ☐ Collaboration approach

☐ Setup instructions
☐ Member video links

Project Showcase (5 MIN)

Elevator Pitch Structure:

☐ Problem context
☐ Dataset overview
☐ Methods (brief)
☐ Key insight
☐ Business impact
☐ Future work

Individual Video

Each member answers:

☐ What did you learn?
☐ Challenges faced?
☐ How was it solved?
☐ What would you add?
☐ Team strengths?

Task Tracker

Task	Owner	Deadline	Status
Github Setup	**Vince**	Feb 25, 2026	Completed ▾
Data Cleaning	**May not be required**	Feb 26, 2026	Not Required ▾
Project Proposal	**Kerensa**	Feb 27, 2026	In Progress ▾
Question Analysis	**All**	Feb 27, 2026	In Progress ▾
Visualization	**All**	Feb 27, 2026	In Progress ▾
README	**Vince**	Mar 3, 2026	Not Started ▾

Slides	**Jarsmeka**	Mar 5, 2026	Not Started ▾
Deadline	**All**	Mar 7, 2026	Not Started ▾