Phase 3 Project Introduction

// FLATIRON SCHOOL

Agenda

- Overview Across Projects
- Project Deliverables
- Schedule

Overview

Key Points

Classification and Metrics

- You are tackling a **classification problem** in this project
- **Choosing the right metric** is a key skill, and should be informed by data exploration and the business problem - you should **explicitly justify** why that metric is the most appropriate for **evaluating model performance** using both training and testing data for your project

Iterative approach to modeling

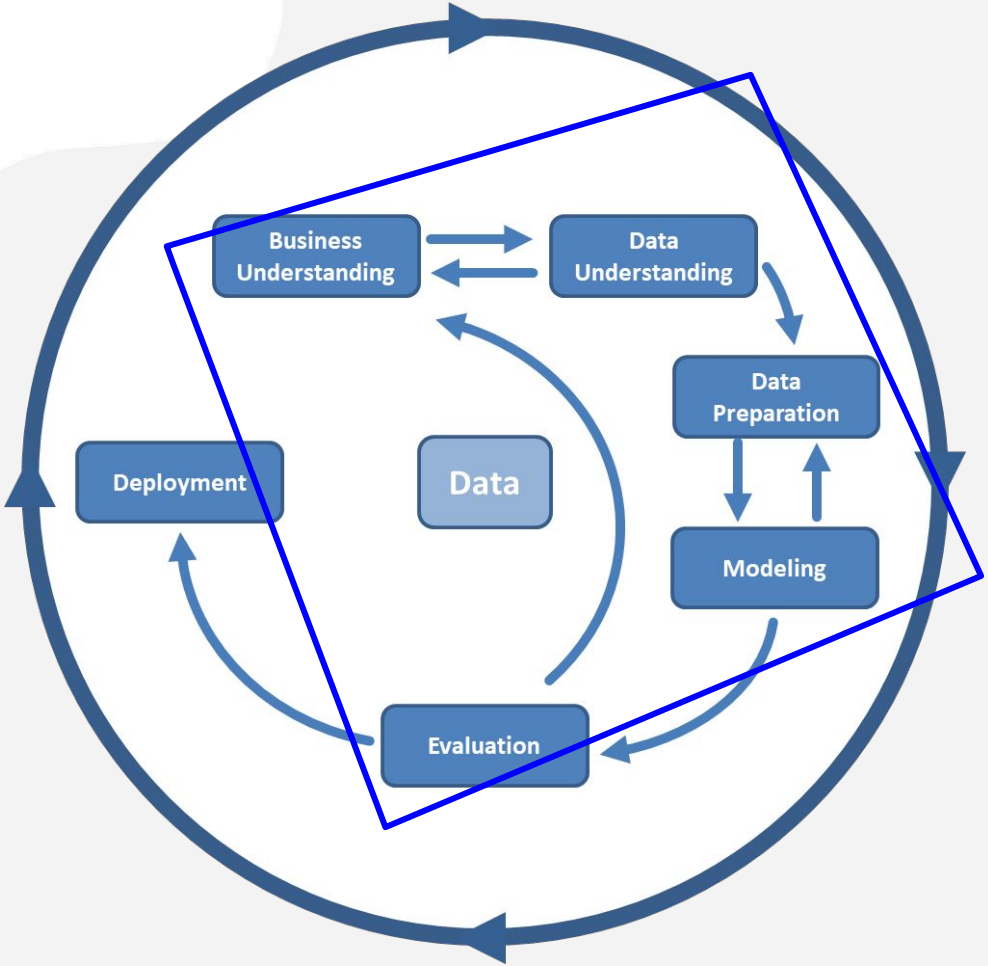
- Explore **different model types** (try simple models first - then, add complexity!)
- After choosing which model best fits your data, **iterate to find the best hyperparameters** for that model

Predictive Approach

- Frame your project's findings and recommendations through a predictive lens, focused on the **output** of your final model
- Can still include inferential elements if it lends support to the business problem

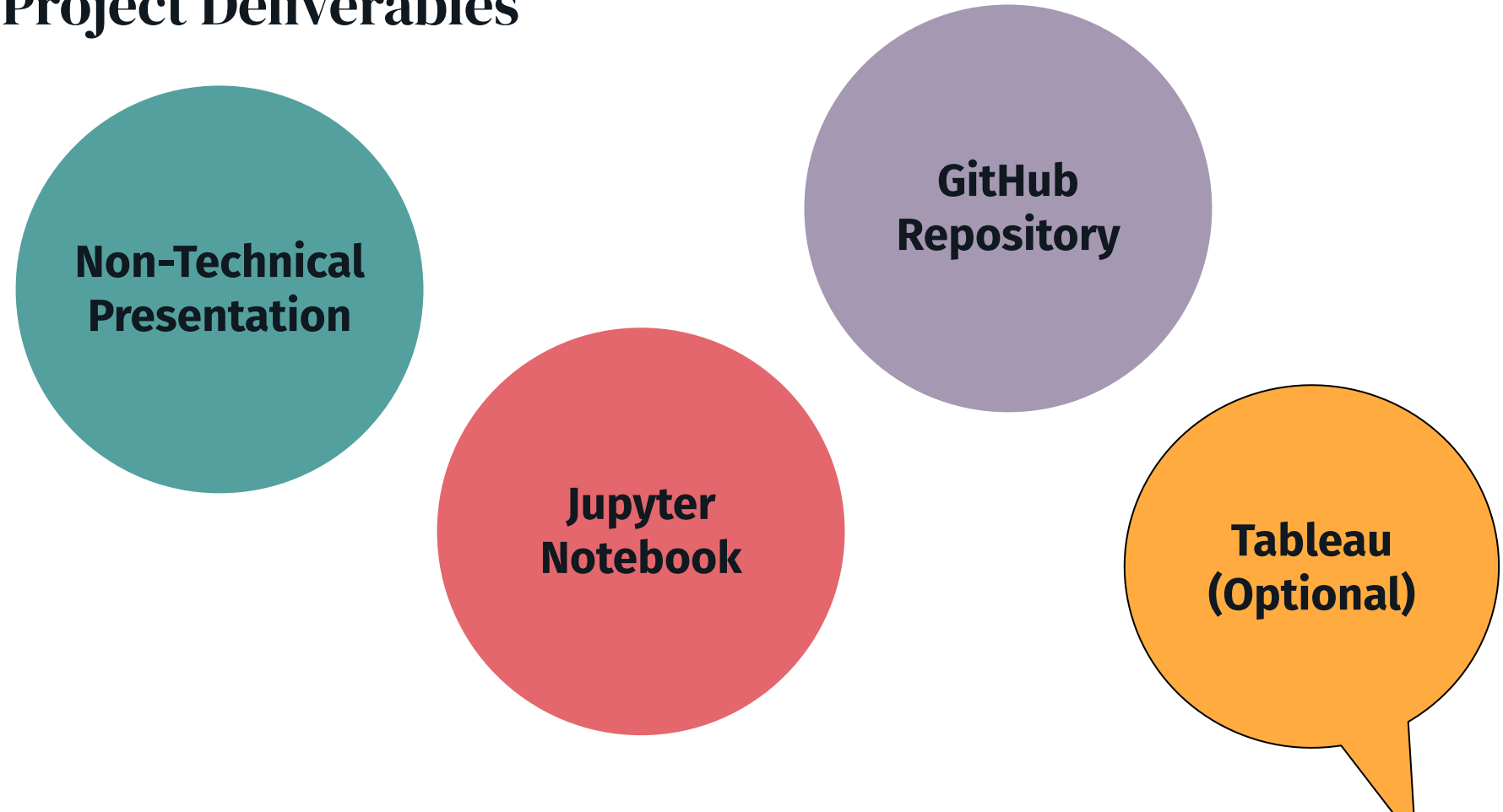
DS Process: CRISP-DM

Consider the **CRISP-DM** process and headers while creating each deliverable.



Project Deliverables

Project Deliverables



Non-Technical Presentation

- Slide deck for a **five minute** presentation
- **Non-technical audience**
- Professional style
  - Light on text
  - Effective template
  - Legible and labeled visualizations

Example slide deck

Non-Technical Presentation

Tell a Story:

Beginning

- Overview
- Business Understanding
- Stakeholder
- Key Business Questions

Middle

- Data Understanding
- **Final Model Results (nontechnically!)**
- Discuss considerations for **metric choice (nontechnically!)**

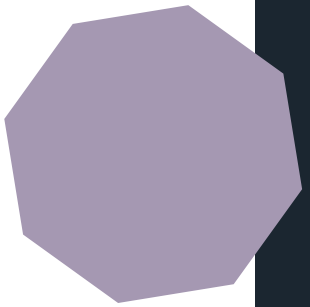
End

- Recommendations
- Next Steps
- Thank You Slide

GitHub Repository

- Where your project lives and grows - want to see a consistent commit history throughout
- **This will be part of your portfolio at the end of this course!**
- Recommend **starting your repository from scratch** rather than forking the template repository

Example repository and templates



GitHub Repository

Must-Haves

1. **README.md**
   - More detail on the next slide
2. **Commit History**
   - Commit history with clear messages
   - Contributions throughout the project period
3. **Organization**
   - Clear folder structure
   - Clear naming conventions for files and folders
   - Technical notebooks and presentation file are easily located
4. **Notebook**
   - Final technical notebook on main level of repo
   - Working notebooks (if applicable) in subfolders
5. **.gitignore**
   - Ignores large files as well as junk files (like .ipynb_checkpoints or .DS_Store)
   - GitHub's python .gitignore template

GitHub Repository

README Sections

Your README should act as a **high-level technical summary**

- **General Overview**
- **Business Understanding**
  - Include stakeholder and business questions
- **Data Understanding**
  - Source of data (either describe or link)
  - Description of data (high level, go into more detail in your technical notebook)
- **Modeling + Evaluation**
  - Describe techniques or methods
  - Written interpretation of results (final model)
- **Conclusion**
  - Summary of conclusions / recommendations
- **Repository File Structure**
  - (nice-to-have not need-to-have)

Jupyter Notebook

- Blends code, markdown, and visualizations to tell the **full story** of your project
- Includes **justifications and rationale** for every decision made throughout the project
- Notebook should be free of errors and run from top to bottom
- Use CRISP-DM steps as markdown headers to divide your final notebook into **sections**

Important Links

- **Project Description**
  - Explains the project goal, dataset, and deliverables
  - Contains rubric explanations
- **Checklist Details**
  - Use to read up on the requirements, including rationale and all the details
- **Choosing a Dataset**
  - 4 (5) Options
  - All classification
  - OR choose your own

Working Groups and Schedule

Group Project Best Practices

1. Get to Know Your Group Members
2. Define Individual Project Contributions
3. Meet Regularly
4. Communicate Actively, Clearly, and Transparently

Schedule

Project Kickoff: Right now!

Data Check: IF using own data, I need to see it by Monday 1 PM ET

Group Check Ins: Wednesday AM

Office Hours: Mon, Tues, Wed PM

Thursday PM: Practice Presentations

Friday PM: Final Presentations

Friday EOD: Submit deliverables on Canvas!

Questions?