

# Phase 4 Project Introduction

// FLATIRON SCHOOL

# Agenda

- Overview Across Projects
- Project Deliverables
- Schedule

# Overview



# Key Points

## Advanced Data Types / Modeling

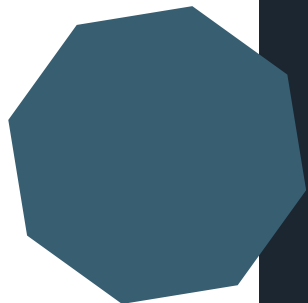
- Tailor your notebook to a technical data science audience, since this could be the most advanced project you complete from a technical perspective and is an excellent opportunity to demonstrate your understanding of an advanced modeling technique

## Validation Strategy

- You should demonstrate how you've ensured your model will perform well on unseen data, using validation strategies which make sense for your specific project type - this may or may not be just a `train_test_split`!

## Don't Get Bugged Down

- This is a great chance to demonstrate advanced skills - but remember that your much larger capstone project is just around the corner

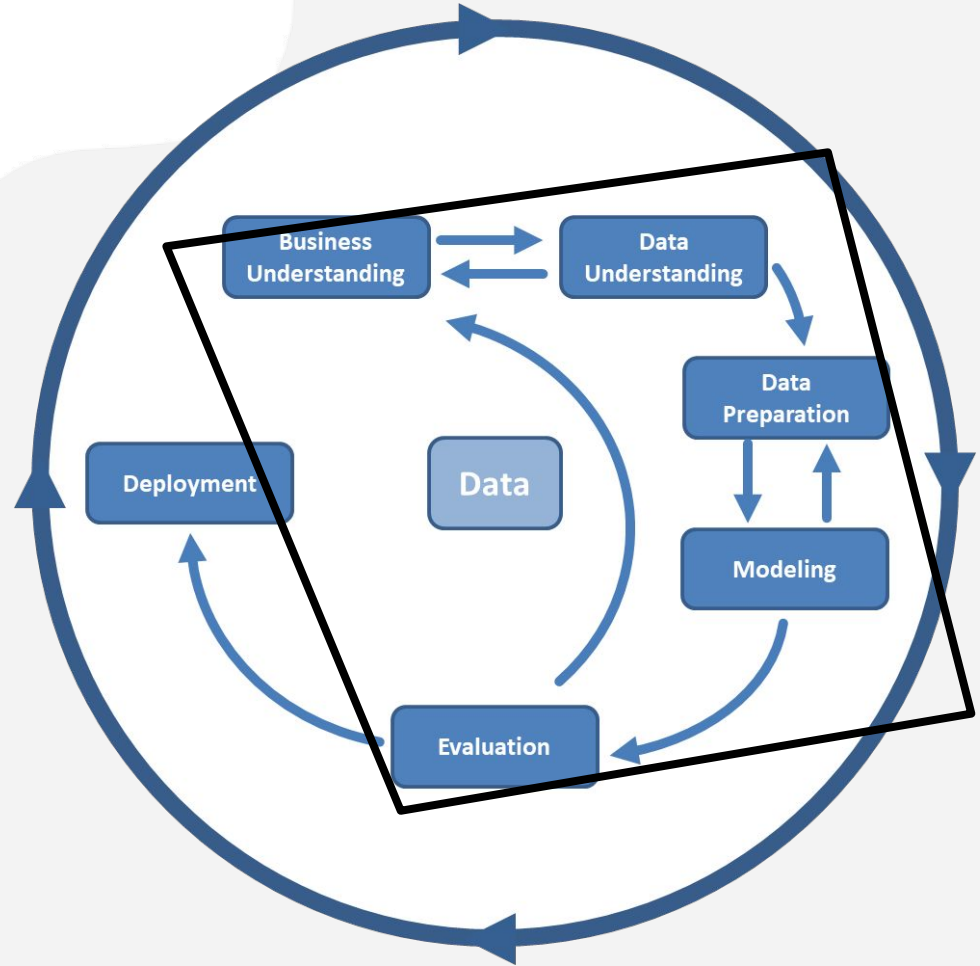


# Project Deliverables



# DS Process: CRISP-DM

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



# Project Deliverables



A diagram showing three overlapping circles representing project deliverables. The top-left circle is teal and contains the text 'Non-Technical Presentation'. The top-right circle is purple and contains the text 'GitHub Repository'. The bottom-center circle is red and contains the text 'Jupyter Notebook'.

**Non-Technical  
Presentation**

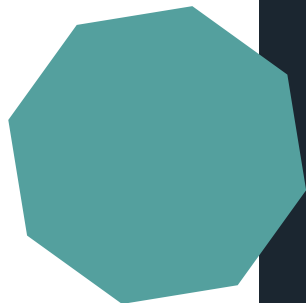
**GitHub  
Repository**

**Jupyter  
Notebook**

# 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!)**

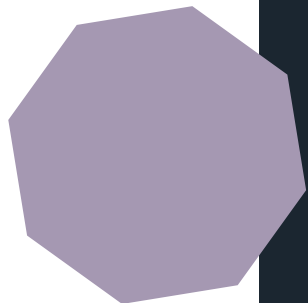
### End

- Recommendations
- Next Steps
- Thank You Slide

# GitHub Repository

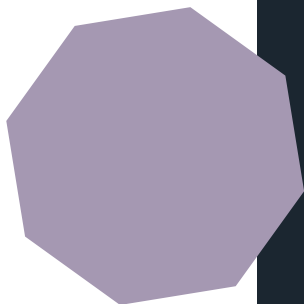
- Where your project lives and grows - want to see a consistent commit history throughout
- **This will be part of your portfolio at the end of this course!**
- Recommend **starting your repository from scratch** rather than forking the template repository

[Example repository and templates](#)



# GitHub Repository

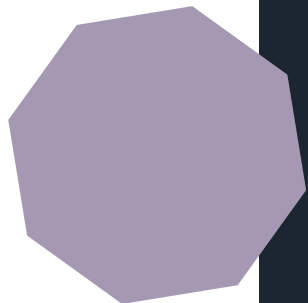
## Must-Haves



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**
  - 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

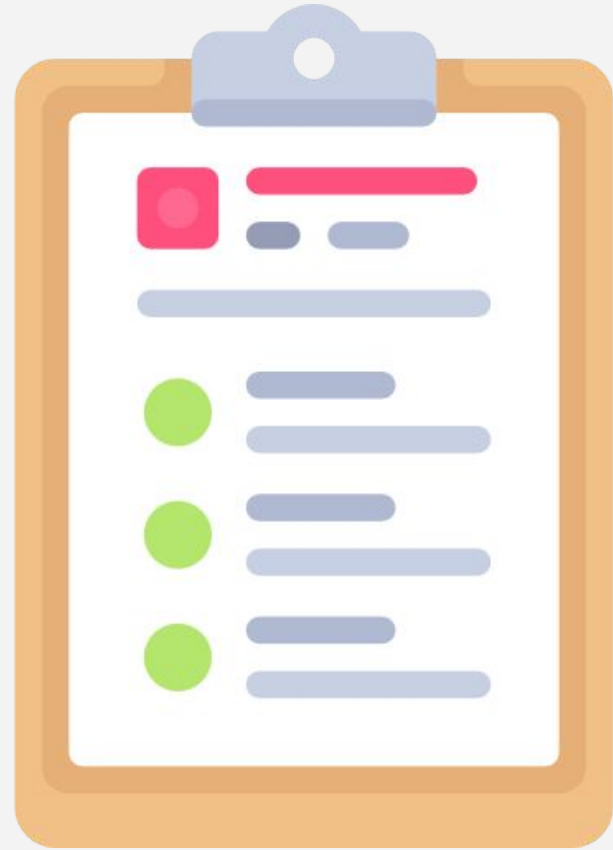
- **Project Description**
  - Explains the project goal, dataset, and deliverables
  - Contains rubric explanations
- **Rubric & Checklist Overview**
  - Use to check off requirements
- **Picking a Dataset**
  - 4 options of project
  - Can use your own or provided data

# 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





# Working Groups

- TBD: See slack survey

# Schedule

**Project Kickoff:** Right now!

**Group Check Ins:** Wednesday AM

**Office Hours:** Mon, Tues, Wed, Thurs

**Thursday PM:** Practice Presentations

**Friday PM:** Final Presentations

**Friday 7 pm ET:** Submit deliverables on Canvas!



Questions?