Phase 1 Project Introduction

// FLATIRON SCHOOL

Agenda

- Building a Professional Data Science Portfolio
- Project Prompt
- Project Deliverables
- Schedule

Building a Data Science Portfolio

"When I'm evaluating a candidate, if they don't have [a PhD or experience as a data analyst] it's hard to say if they'll be able to do the job. But my favorite way to evaluate a candidate is to read an analysis they've done online. If I can look at some graphs someone created, how they explained the story, and how they dug into the data, I can start to understand whether they're a good fit for the role"

David Robinson, Principal Data Scientist
(personal website)
As quoted in *Build a Career in Data Science*

Project Prompt

Project Prompt

Your company is expanding in to new industries to diversify its portfolio. Specifically, they are interested in purchasing and operating airplanes for commercial and private enterprises, but do not know anything about the potential risks of aircraft. You are charged with determining which aircraft are the lowest risk for the company to start this new business endeavor. You must then translate your findings into actionable insights that the head of the new aviation division can use to help decide which aircraft to purchase.

Key Points

Three concrete business recommendations

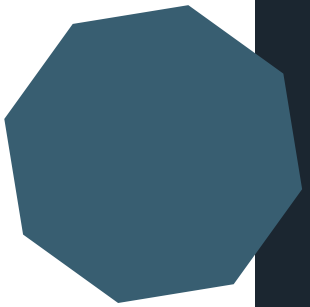
- Explicitly relate your findings to business needs by recommending actions that you think the business should take

Communicate effectively

- Create a storyline your audience can follow, highlighting only the most important points and skipping over the rest

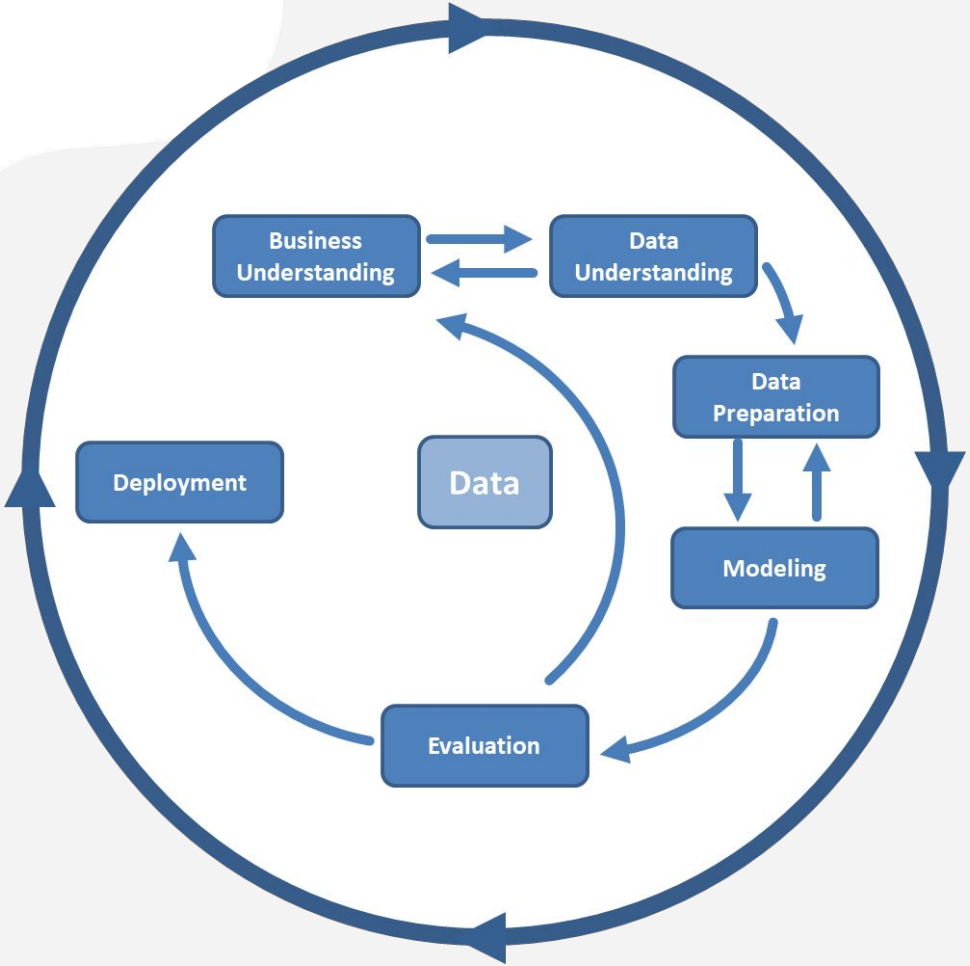
Use plenty of visualizations

- Spotlight visuals in your presentation, but only ones that relate directly to your recommendations
- Simple visuals are usually best (e.g. bar charts and line graphs), and don't forget to format them well (e.g. labels, titles)



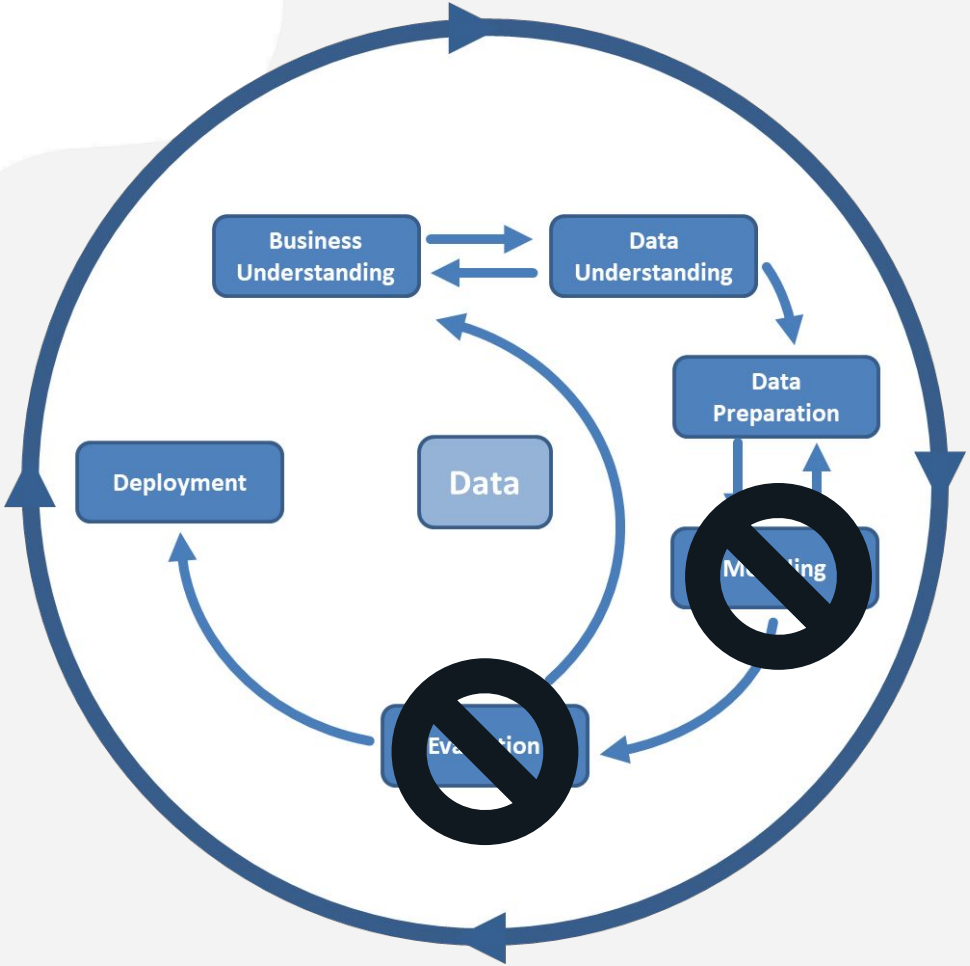
DS Process: CRISP-DM

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



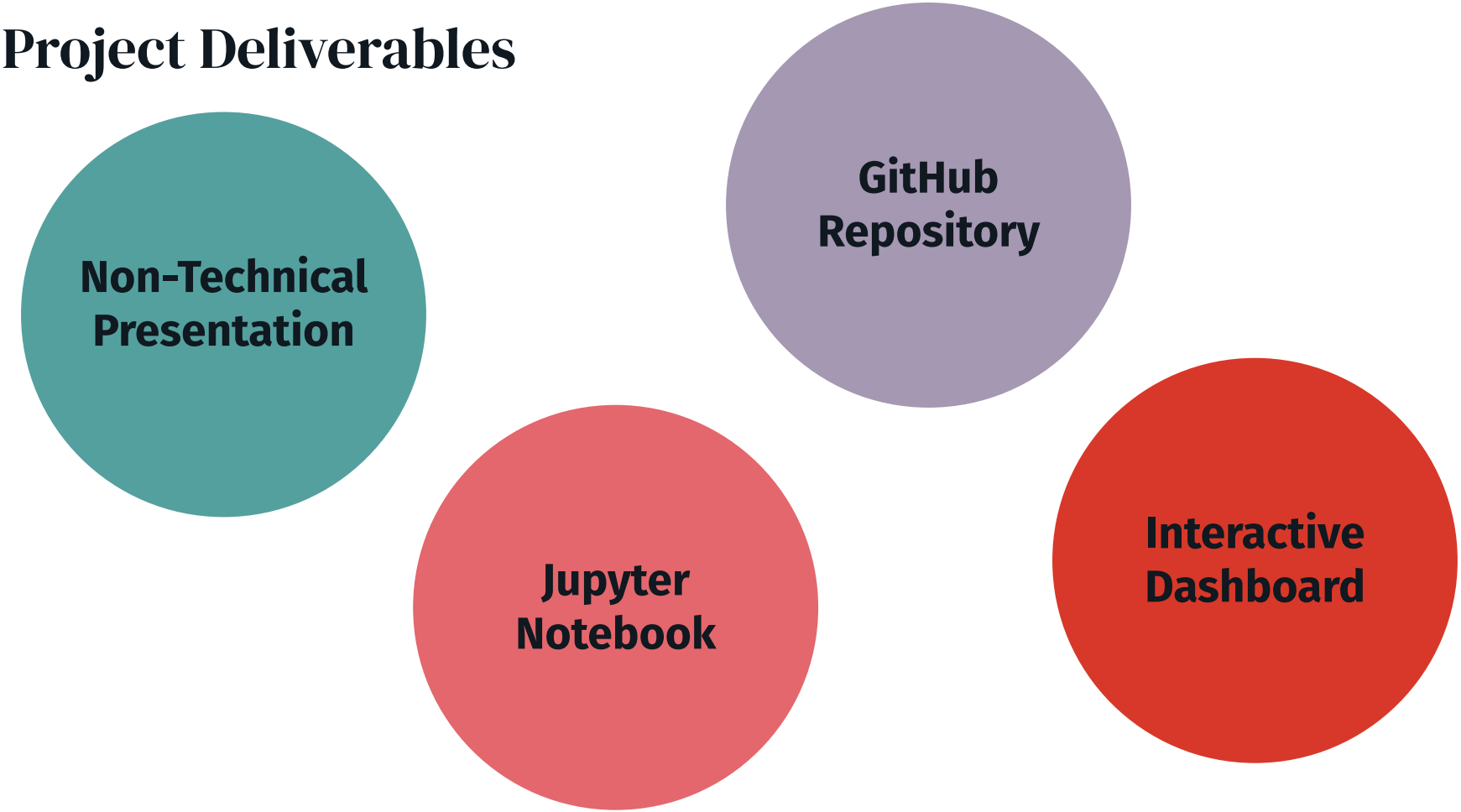
DS Process: CRISP-DM

Modeling and **Evaluation** are not steps within this Project's scope, and you can consider **Deployment** as the completed deliverables and your three recommendations.



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
- Key Statistics Supporting Findings
- Key Visualizations Supporting Findings

End

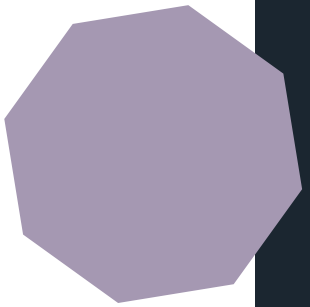
- Recommendations
- Next Steps
- Thank You Slide

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** or using the template

Template

Example



GitHub Repository

Must-Haves

01	**README.md**	• See next slide
02	Commit History	• Commit history with clear messages • Contributions from all team members
03	**Organization**	• Clear folder structure • Clear names of files and folders • Easily located technical notebook and presentation linked in the README.md
04	**Notebooks**	• Final technical notebook that runs without errors • Working notebooks from individual team members
05	**.gitignore**	• Ignores large files and pesky things like .ipynb_checkpoints • Github python gitignore

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)
- **Data Analysis/Recommendations**
  - Describe *interesting* techniques or methods
  - Written interpretation of results
  - Visuals that showcase your results
  - Interactive dashboard (Tableau)
- **Conclusion**
  - Summary of conclusions / recommendations
- **Repository File Structure**
  - See example

Jupyter Notebook

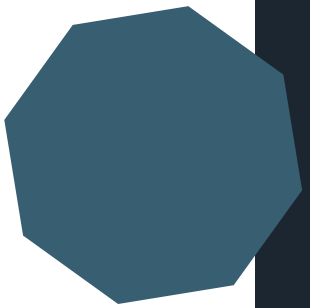
- Blends code, markdown, and visualizations to tell the **full story** of your project (content may overlap with your non-technical presentation and README)
- Includes **justifications and rationale** for every decision made throughout the project
- Notebook should be free of errors and run from top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

Example final notebook

Note on Notebooks and GitHub

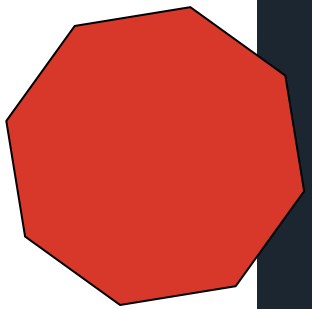
Your final notebook deliverable is **one .ipynb file** on the main level of your GitHub repository, which contains all important contributions from group members blended into a **seamless report notebook**.

However! You should create individual notebooks, kept in working subfolders, to **avoid merge conflicts**.



Interactive Dashboard

- Utilize Tableau to create multiple visuals
  - Highlight your key points
  - Align with recommendations/analysis
  - Use cleaned dataset
- Combine individual visuals into a dashboard
  - Desktop format
  - Can have other visuals or dashboards
- Should be ‘Interactive’ in some sense
  - Filtering etc...



Important Links

- **Project Description**
  - Explains the project goal, dataset, and deliverables
  - Contains rubric explanations
- **Checklist Overview**
  - Detailed grading checklist
- **Template**
  - Good starting place, contains dataset and template repo

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!

Check Ins: Tuesday AM/PM

Office Hours: Mon, Tues, Wed, Thurs

Thursday AM: Practice Presentations

Friday AM: Final Presentations

Final Due Date

Friday 11:30 AM ET:
Submit deliverables on Canvas!

Questions?