Phase 1 Project Introduction

Agenda

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

Building a Data Science Portfolio

GitHub

LinkedIn

Medium

"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

Use exploratory data analysis to **generate three actionable insights** that the new head of "Microsoft Movie Studios" can use to decide what kinds of films to create.

Key Points

Three concrete business recommendations
- Explicitly relate your findings to business needs by recommending actions that you think the business (Microsoft) should take

Communicate effectively
- Create a storyline your audience ("Microsoft Movie Studios" new director) can follow, highlighting only the most important points and skipping over the rest
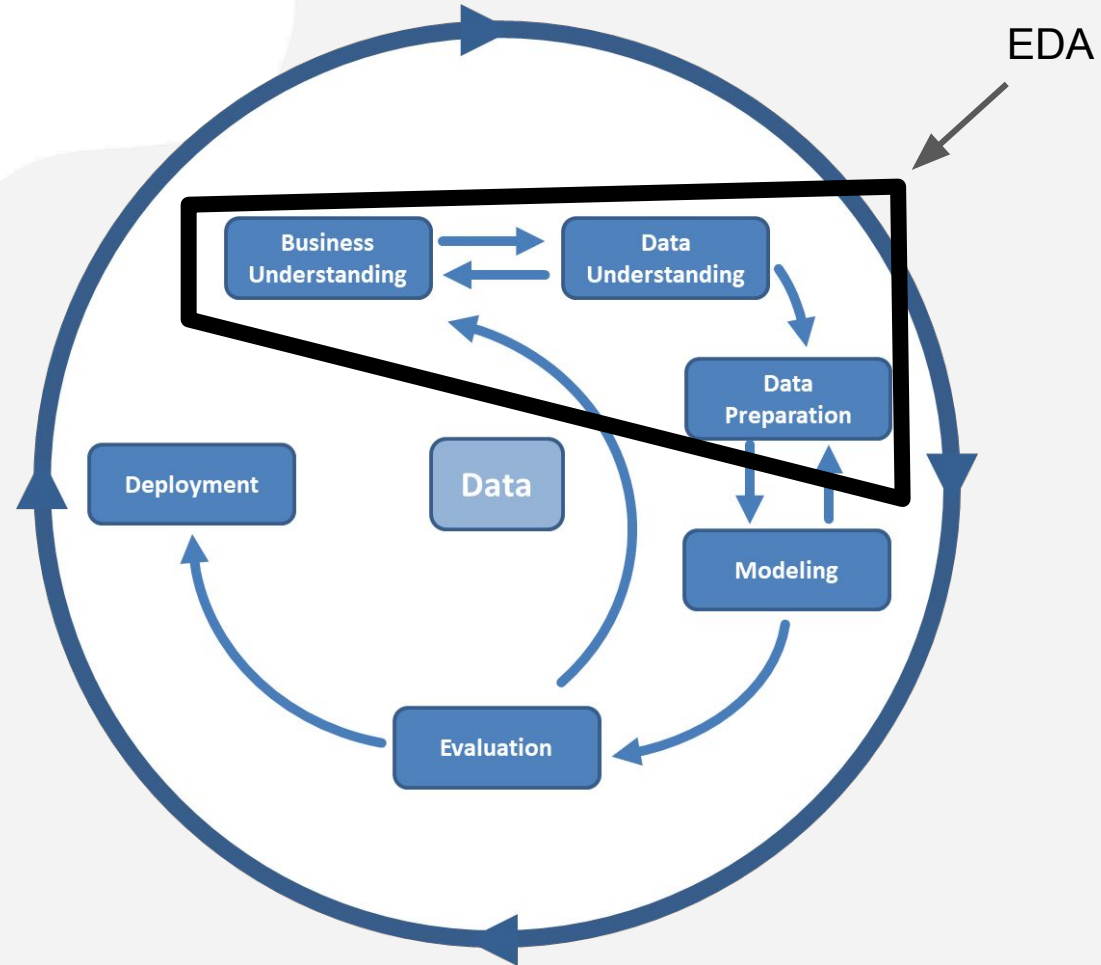
Use plenty of visualizations
- Spotlight visuals in your presentation, but only ones that relate directly to your recommendations
- Simple visuals are usually best (e.g. bar charts and line graphs), and don't forget to format them well (e.g. labels, titles)
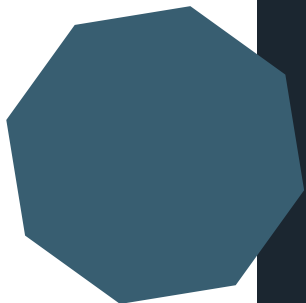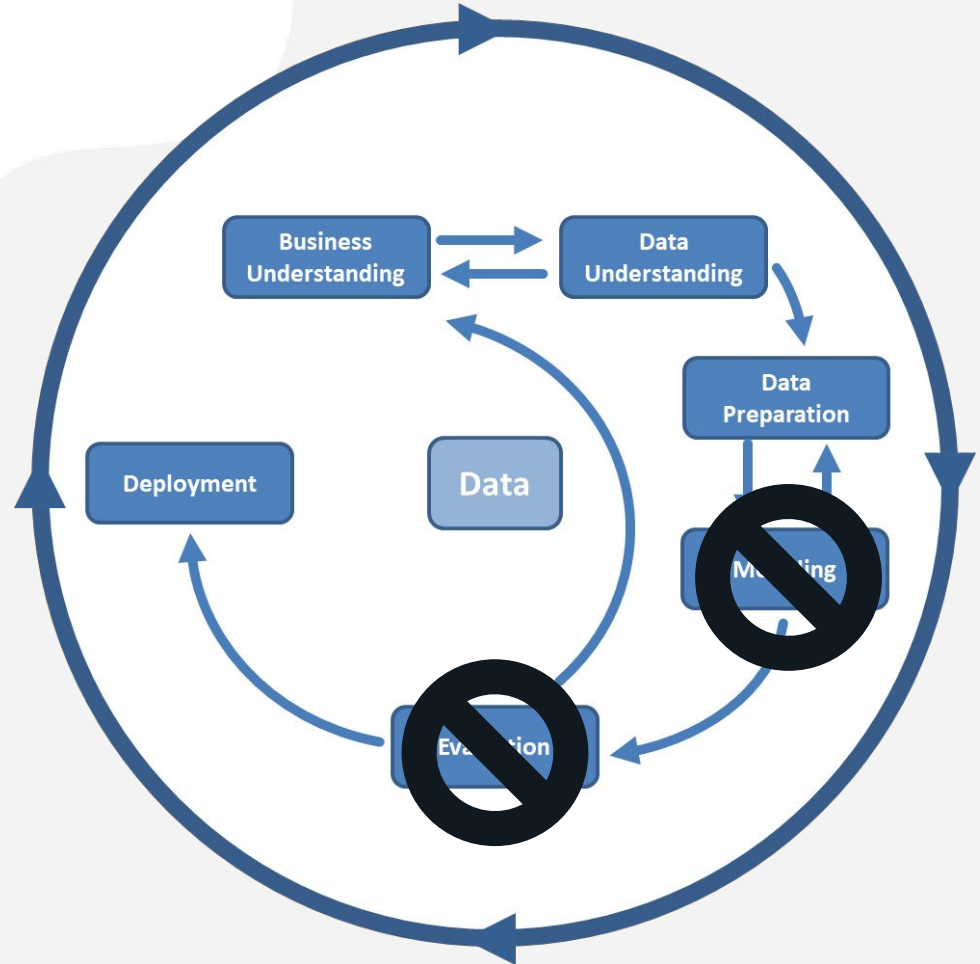
DS Process: CRISP-DM

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
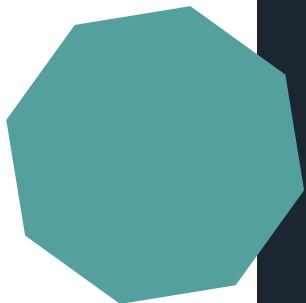
Project Deliverables

Non-Technical Presentation

Jupyter Notebook

GitHub Repository

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
- Summary of recs
- 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** rather than forking the template repository
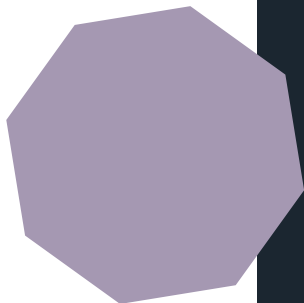
Example repository and templates

- **I would like every group to create a repository from scratch and use the template to copy stuff over**
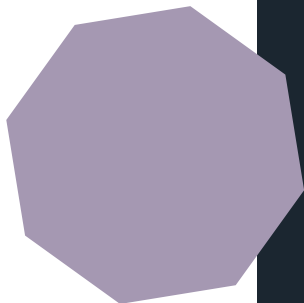
GitHub Repository

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
   - **ADD THE UNZIPPED SQL DATABASE**

GitHub Repository

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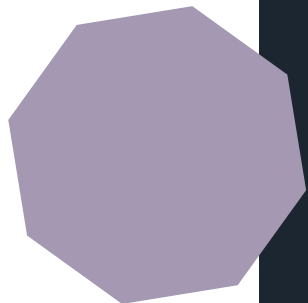
Your README should act as a **high-level technical summary**

- **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**
  - Describe *interesting* techniques or methods
  - Written interpretation of results
  - Visuals that showcase your results
- **Conclusion**
  - Summary of conclusions / recommendations
- **Repository File Structure**
  - Will be required for capstone project

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**
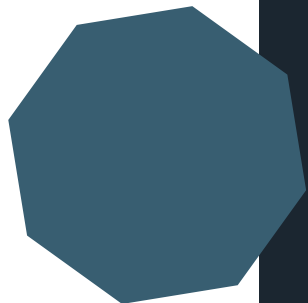
Example final notebook

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**.

Important Links

- **Project Description**
  - Explains the project goal, dataset, and deliverables
  - Contains rubric explanations

- **Rubric & Checklist Overview**
  - Use to check off requirements

- **Rubric & Checklist Details**
  - Use to read up on the requirements, including rationale and all the details

- **Submission and Review Instructions**
  - Note that you submit the GitHub repository link and PDF copies separately
  - Github link
  - PDF's

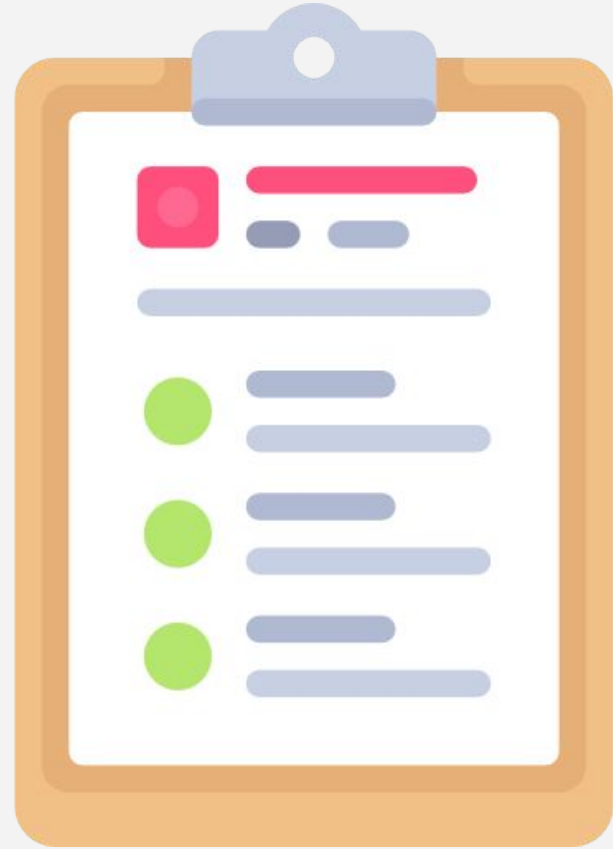# Working Groups and Schedule

Working Groups

- Group 1
  - Samuel, Andrew, Stefano
- Group 2
  - Juan, Jon, Holly
- Group 3
  - Raul, Eddie, Drew
- Group 4
  - Terry, Dermot
- Group 5
  - Anita, Eliot

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

Group Contract

1. **Make a copy** of [this contract](#)
2. Fill out with your group members
3. Send to Daniel via slack (one copy)
4. Submit by Monday the 3rd, EOD 6pm EST

Schedule

Project Kickoff: Right now!

Group Contracts: Monday EOD

Group Check Ins: Tuesday PM

Office Hours: Mon, Tues, Wed, Thurs

Thursday PM: Practice Presentations

Friday PM: Final Presentations

Friday 4 pm ET: Submit deliverables on Canvas!

Questions?