Phase 2 Project Introduction

// FLATIRON SCHOOL

Agenda

- Project Prompt
- Project Deliverables
- Schedule

Project Prompt

Project Prompt

Use multiple linear regression inferential modeling to **analyze house sales** in King County, Washington.

It is *up to you* to **define a stakeholder and business problem** appropriate to this dataset!

Key Points

Results include overall modeling performance (metric) as well as *at least* two important features for final model

- Your goal is to yield findings to support relevant recommendations - find important features that impact your target, and interpret their impact
- Make sure you are thinking about how a linear regression model adds value to your analysis. "The assignment was to use linear regression" is not an acceptable answer!

Iterative approach to modeling

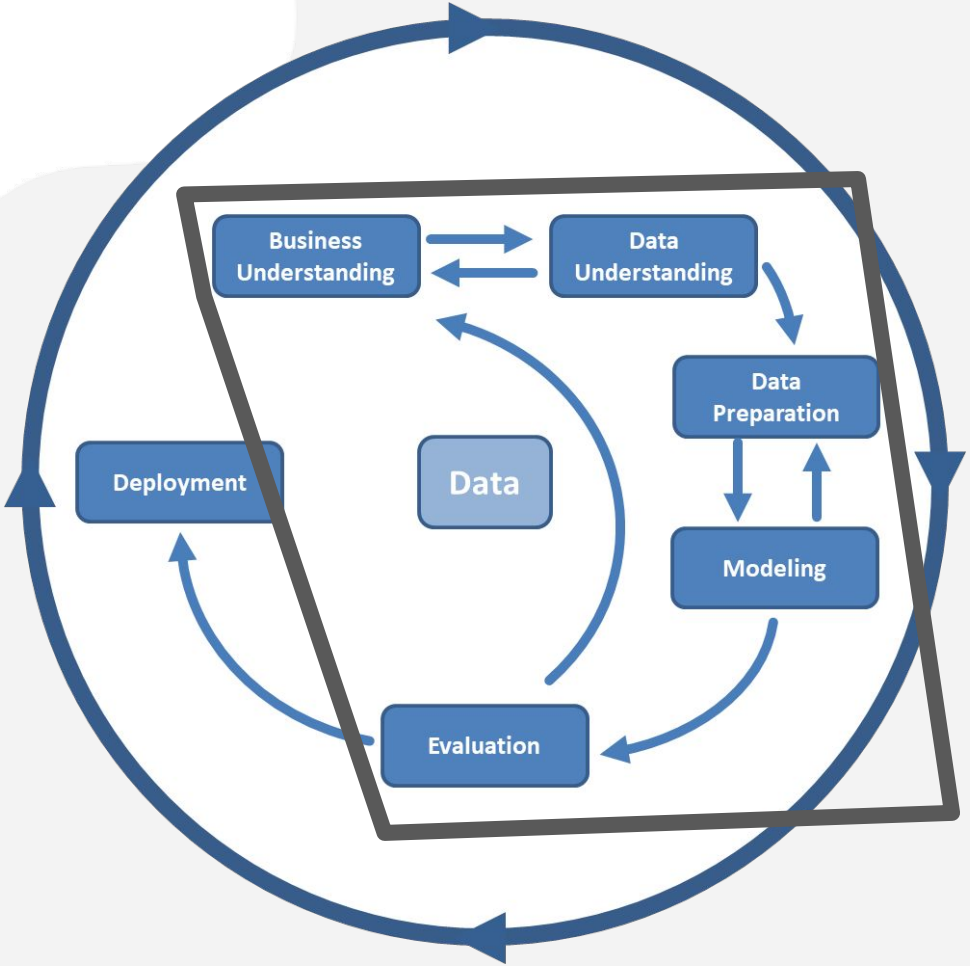
- You **must** build multiple models. Begin with a simple model, evaluate it, and then provide justification for and proceed to a new model until you wrap up with a ‘final’ model

Use plenty of visualizations

- Data visualization and analysis are no longer explicit project requirements, but they will help you build better models and tell a better story to your stakeholders

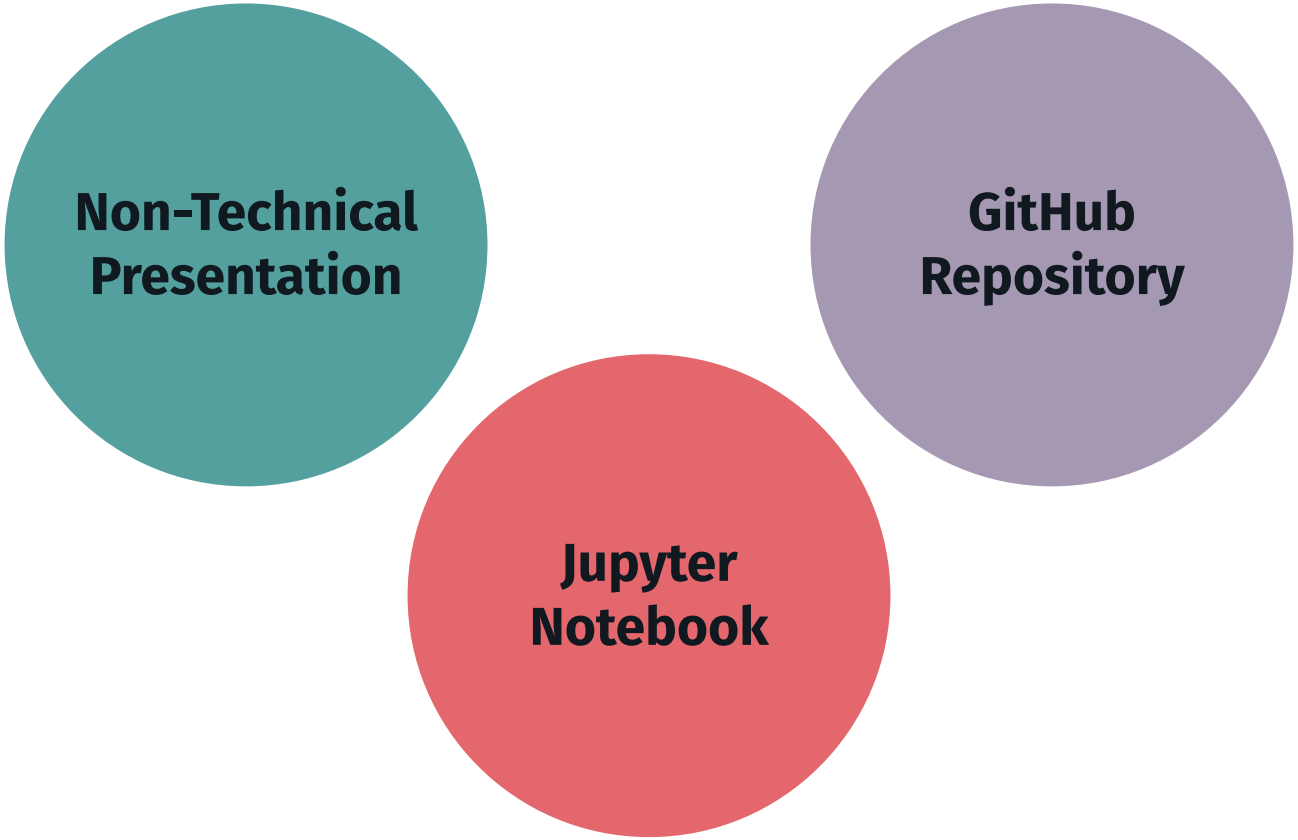
DS Process: CRISP-DM

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 (non technically!)**
- **Exploration of Important Features**

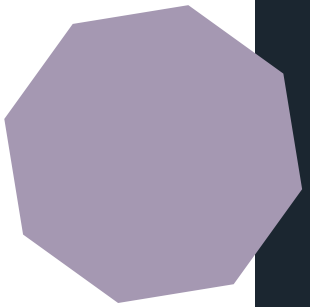
End

- 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 (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**

Things to Consider:

Outliers
Leave or drop?

Ordinal or
OHE?

Null values?

Categorical or
Continuous?

Multicollinearity

Location or
Not?

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

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

- Group 1 :
- Group 2 :
- Group 3 :
- Group 4 :
- Group 5 :

Schedule

Project Kickoff: Right now!

Office Hours: Mon, Tues, Wed, Thurs

Wednesday AM: Group Check-ins

Thursday PM: Practice Presentations

Friday PM: Final Presentations

Friday 5 pm ET: Submit deliverables on Canvas!

Questions?