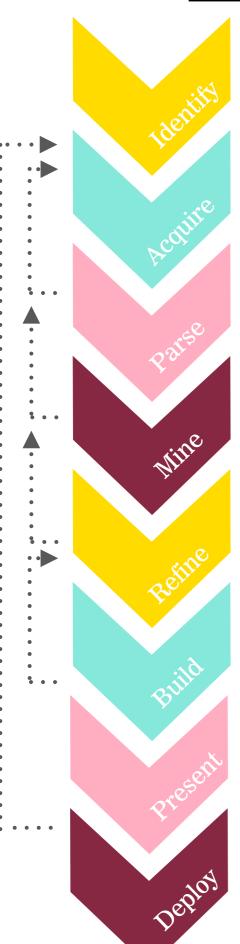
DATA SCIENCE WORKFLOW



IDENTIFY THE PROBLEM

- □ Identify business/product objectives
- $\hfill\square$ Identify and hypothesize goals and criteria for success
- Create a set of questions for identifying correct data set

ACQUIRE THE DATA

- ☐ Identify the "right" data set(s)
- $\hfill\square$ Import data and set up local or remote data structure
- $\hfill\square$ Determine most appropriate tools to work with data

PARSE THE DATA

- □ Read any documentation provided with the data
- Perform exploratory data analysis
- □ Verify the quality of the data

MINE THE DATA

- $\hfill\square$ Determine sampling methodology and sample data
- ☐ Format, clean, slice, and combine data in Python
- Create necessary derived columns from the data (new data)

REFINE THE DATA

- ☐ Identify trends and outliers
- Apply descriptive and inferential statistics
- Document and transform data

BUILD A DATA MODEL

- □ Select appropriate model
- Build model
- Evaluate and refine model

PRESENT RESULTS

- Summarize with narrative, storytelling techniques
- $\hfill\square$ Present limitations and assumptions of your analysis
- $\hfill\square$ Identify follow up issues for future analysis

DEPLOY AND VALIDATE

- □ Write unit tests and documentation
- Deploy stable production-ready code
- $\hfill\square$ Retrain and validate models over time