

# PRACTICE DAY

Naumaan Nayyar

# **OBJECTIVES**

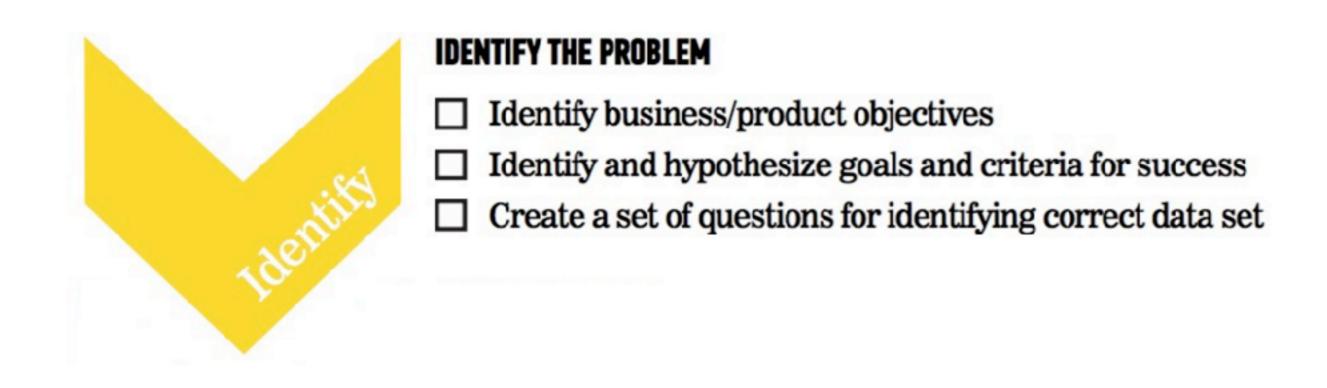
Practice the data science workflow on problems

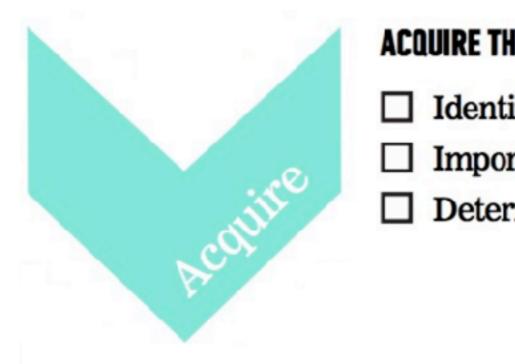
# THE DATA SCIENCE WORKFLOW

#### The steps:

- 1. Identify the problem
- 2. Acquire the data
- 3. Parse the data
- 4. Mine the data
- 5. Refine the data
- 6. Build a data model
- 7. Present the results

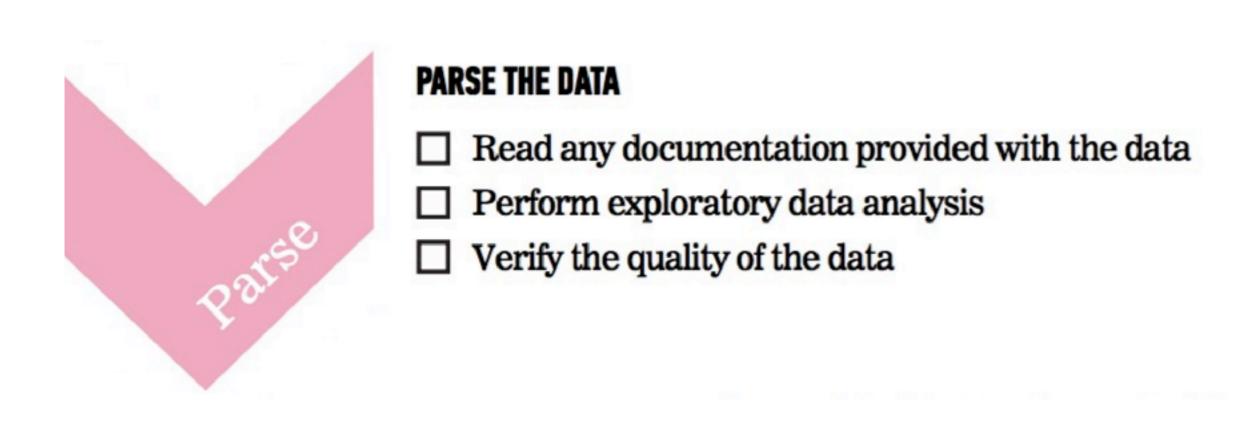
#### DATA SCIENCE WORKFLOW IDENTIFY THE PROBLEM ☐ Identify business/product objectives Identify and hypothesize goals and criteria for success. Create a set of questions for identifying correct data set ACQUIRE THE DATA ☐ Identify the "right" data set(s) Import data and set up local or remote data structure. ■ Determine most appropriate tools to work with data Read any documentation provided with the data ☐ Perform exploratory data analysis Verify the quality of the data. Determine sampling methodology and sample data Format, clean, silee, and combine data in Python Create necessary derived columns from the data (new data). RETINE THE BATA ☐ Identify trends and outliers Apply descriptive and inferential statistics □ Document and transform data JERON ATAL A CLIUR ☐ Select appropriate model □ Build model Expluste and refine model. PRESENT THE RESULTS ☐ Summarize findings with narrative, storytelling techniques ☐ Present limitations and assumptions of your analysis Identify follow up problems and questions for future analysis





#### **ACQUIRE THE DATA**

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

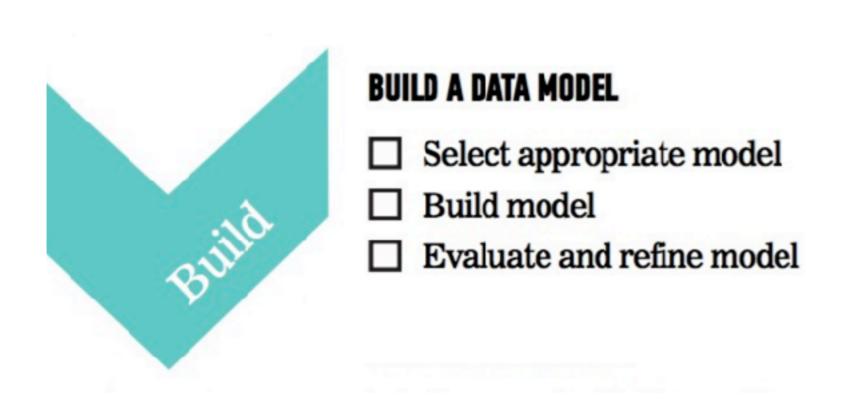




#### MINE THE DATA

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







#### PRESENT THE RESULTS

- ☐ Summarize findings with narrative, storytelling techniques
- ☐ Present limitations and assumptions of your analysis
- ☐ Identify follow up problems and questions for future analysis

# **GROUP ACTIVITY**

# PRACTICE

### **ACTIVITY: DATA SCIENCE PRACTICE**

#### DIRECTIONS (120 minutes)



- 1. Form groups.
- 2. Pick a problem among: loan funding prediction, sentiment analysis, housing prices prediction.
- 3. Go through the data science workflow.