Flexible Class Session #2
Modeling

Ivan Corneillet

Data Scientist

GENERAL ASSEMBLY

Learning Objectives

After this lesson, you should be able to:

- Review Steps ❺ Refine the Data and ❻ Build a Model and more specifically
  - Linear Modeling (OLS)
  - Classification Modeling (KNN and Logistic Regressions)
- Have fun doing Data Science!

Announcements and Exit Tickets

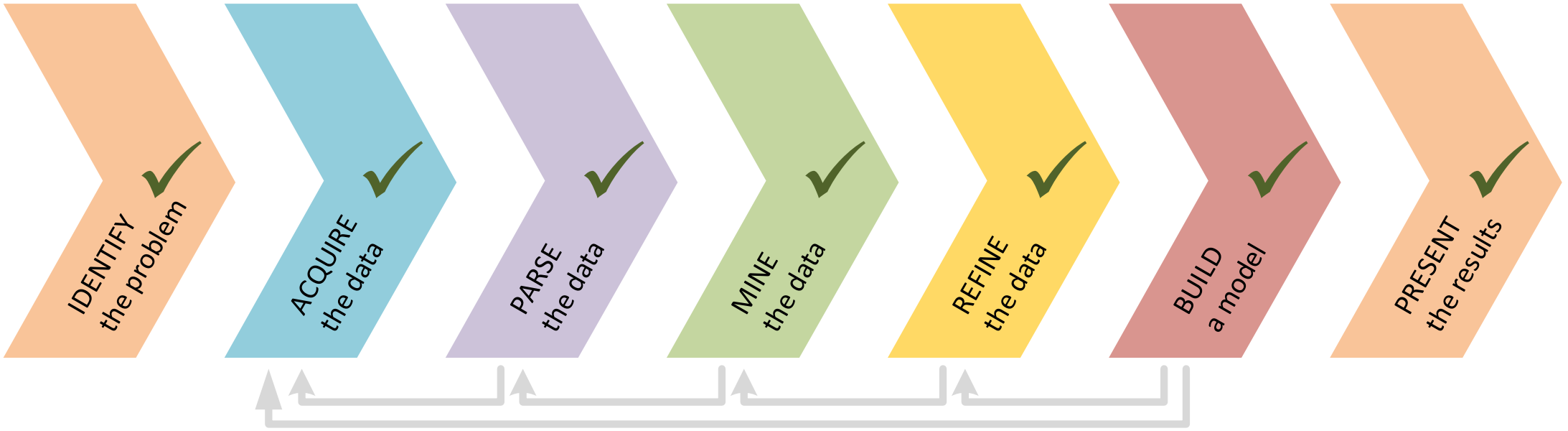
Q & A

Today

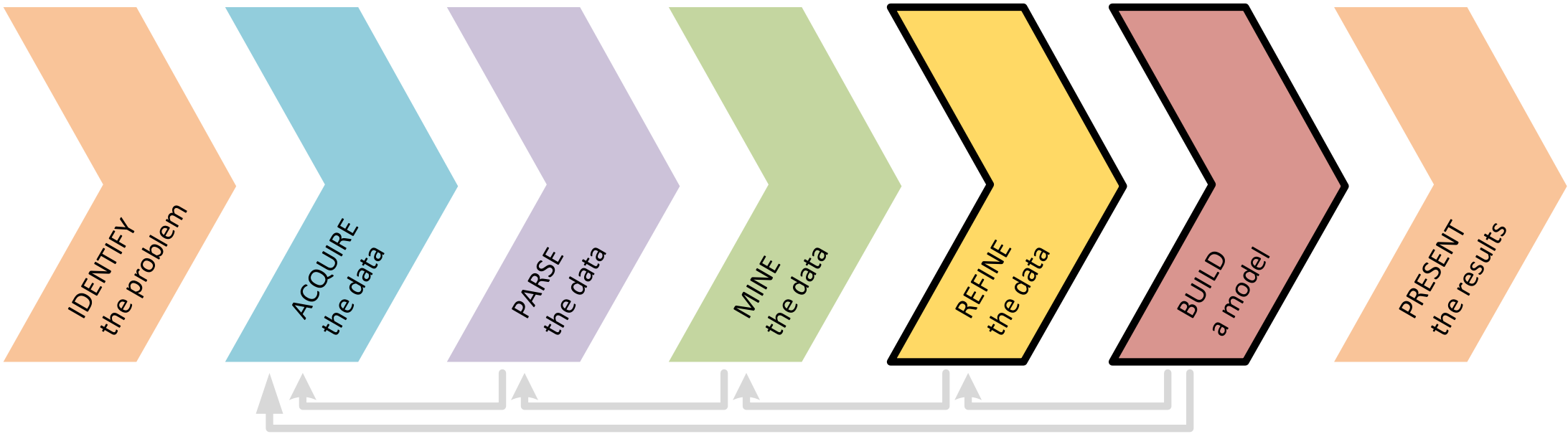
Today, we are wrapping Unit 2 – Foundation of Modeling

Research Design and Data Analysis	Research Design	Data Visualization in *pandas*	Statistics	Exploratory Data Analysis in *pandas*
Foundations of Modeling	Linear Regression *(sessions 6, 7, and 11)*	Classification Models (KNN, Logistic Regression) *(sessions 8, 9, and 11)*	Evaluating Model Fit *(sessions 5, 6, and 7)*	Presenting Insights from Data Models *(session 10)*
Data Science in the Real World	Decision Trees and Random Forests	Time Series Data	Natural Language Processing	Databases

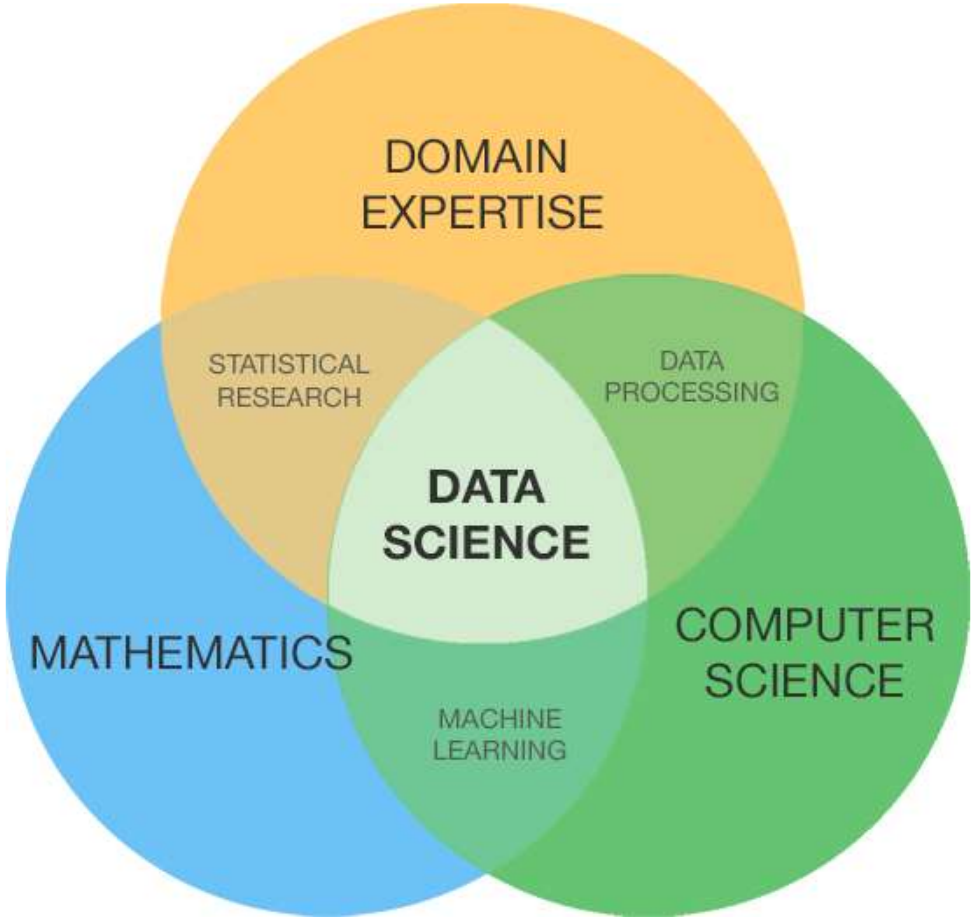
… as well as the first full pass of the Data Science Workflow

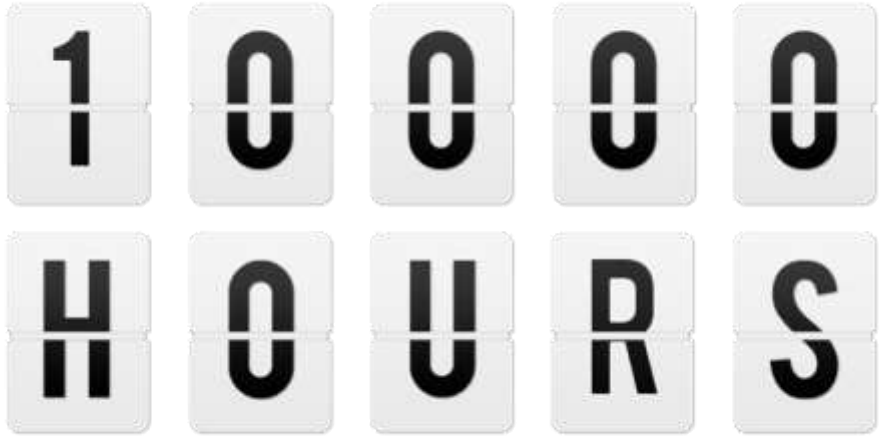


Today we will apply what we learned in Unit 2 and the REFINE the data and BUILD a model steps of the data science workflow on a new dataset



This session is about practice, practice, and practice…





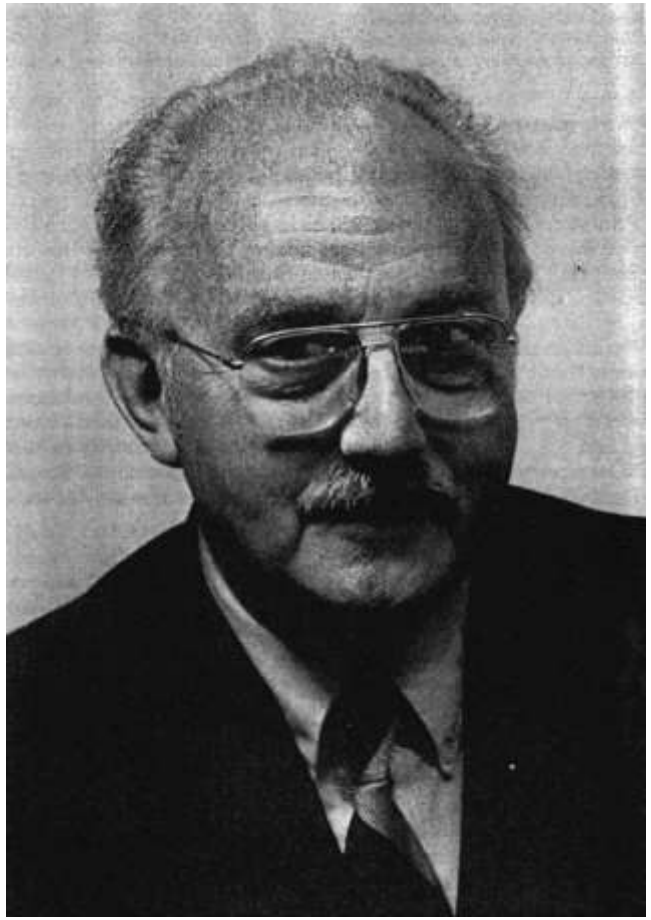
Here's what's happening today:

- Announcements and Exit Tickets
- Review
- ❺ Refine the Data and ❻ Build a Model
  - Linear Modeling (OLS)
  - Classification Modeling (KNN and Logistic Regressions)
- Final Project 3 (due next session on 6/14)

Auto-MPG

Modeling

And don’t forget that…



- “Essentially, all models are wrong, but some are useful” – George Box

Q & A

Before Next Class

Before Next Class

Before the next lesson, you should already be able to:

- Explain the concepts of cross-validation, logistic regression, and overfitting
- Know how to build and evaluate some classification model in *sklearn* using cross-validation and AUC

Next Class

Decision Trees and Random Forests

Learning Objectives

After the next lesson, you should be able to:

- Understand and build decision tree models for classification and regression
- Understand the differences between linear and non-linear models
- Understand and build random forest models for classification and regression
- Know how to extract the most important predictors in a random forest model

Exit Ticket

Don’t forget to fill out your exit ticket here