Flexible Session #1
Exploratory Data Analysis

Ivan Corneillet

Data Scientist

GENERAL ASSEMBLY

Learning Objectives

After this lesson, you should be able to:

- Review Step ❸ Parse the Data and more specifically
  - Descriptive Statistics and Exploratory Data Analysis
  - Apply *pandas* on a Kaggle dataset
- Have fun doing Data Science!

Announcements and Exit Tickets

Review

Python and *pandas*

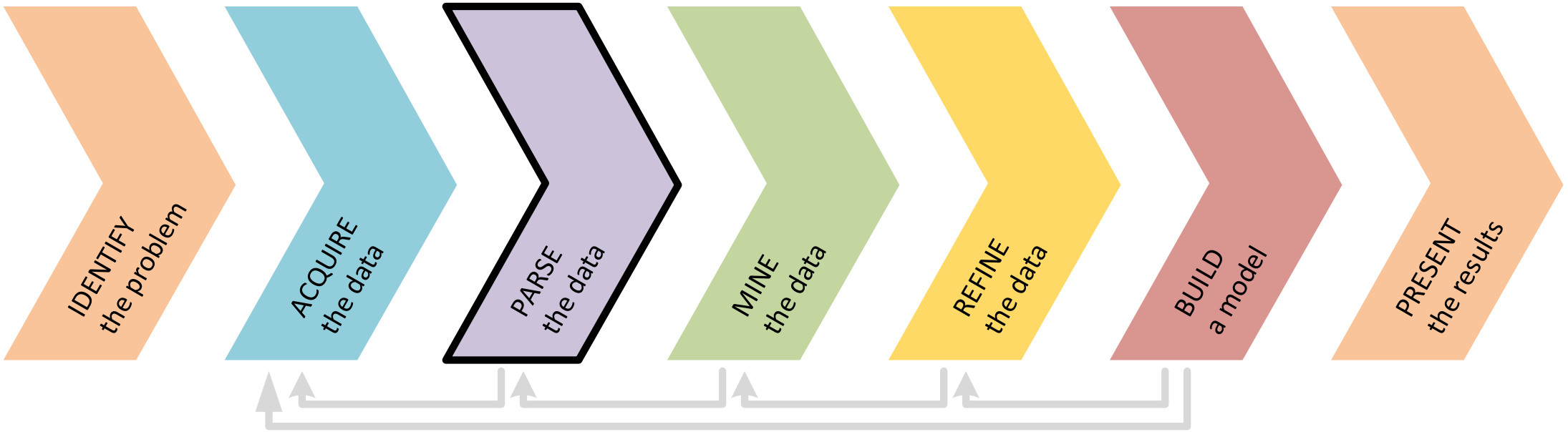
Measure of Centrality	`.mean()`	`.median()`	`.mode()`
Measure of Dispersion	`.var(), .std()`	`.min(), .max()` `.quantile()`	
Summary	`.describe()`		
Graphical Methods		`.plot(kind = 'box')`	`.plot(kind = 'hist')`
Correlation Matrix	`.corr()`		
Scatter plot	*DataFrame*`.plot(kind = 'scatter', x = '`*SeriesName*`', y = '`*SeriesName*`')`		
Scatter matrix	`pd.tools.plotting.scatter_matrix(`*DataFrame*`)`		

`.shape, .columns,`
`.set_index(), .drop()`
	`.count(), .sum(), .unique()`
	`.value_counts(), .isnull(),`
	`.notnul(), .dropna(),`
	`pd.crosstab(`*Series*`, `*Series*`)`
		`np.sort(), .apply()`	

Q & A

Today

Today we'll keep our focus on PARSE the data



And more precisely on the Exploratory Data Analysis using the *pandas* library

Research Design and Data Analysis	Research Design	Data Visualization in *pandas*	Statistics	Exploratory Data Analysis in *pandas*
Foundations of Modeling	Linear Regression	Classification Models	Evaluating Model Fit	Presenting Insights from Data Models
Data Science in the Real World	Decision Trees and Random Forests	Time Series Data	Natural Language Processing	Databases

Here’s what’s happening today:

- Announcements and Exit Tickets
- Review
- ❸ Parse the Data
  - Kaggle – Exploratory Data Analysis
- Review
- Exit Tickets

Kaggle

Exploratory Data Analysis

Q & A

Next Class

Inferential Statistics for Model Fit

Learning Objectives

After this next lesson, you should be able to:

- Explain the difference between causation and correlation
- Identify a normal distribution within a dataset using summary statistics and visualization
- Test a hypothesis within a sample case study
- Validate your findings using statistical analysis (t-tests, p-values, t-values, confidence intervals)

Exit Ticket

Don’t forget to fill out your exit ticket here