Pandas Consumer Sales Lab

Learning Objectives

After this lesson, you will be able to:

- Apply what you've learned in the datetime and joining lessons to a real dataset.
- Apply your charting experience to visualize insights based off of your EDA'd data.

To the notebook!

We actually will commence this lesson directly in the Jupyter Notebook, `pandas-consumersales.ipynb`, to walkthrough the what, why, and how all at once.

Here we have slides reviewing the key concepts.

Exercise Overview

- First, this exercise can easily take more than 60 minutes.
- Think of this as an opportunity to dive into the topic and apply datetime and joining operations to a real dataset.
- Budget time *outside of class* to continue work on this if you can:
    - Remember, the more comfortable you become with this, the more likely you'll use it in your day-to-day life!

Data Background

- This exercise uses a dataset originally used for qlik.
- If you want, try clicking (pun intended) around in their web-based solution to familiarize yourself with the data.
- This is a global food distribution company (canned goods, produce, meats, etc.).
- The data you have about their sales and inventory is distributed across multiple sheets, and even in different languages!
- This is an exercise *very* similar to what you'd be doing with relational databases with a larger enterprise company.

What Are We Looking For?

- Your boss, Joanna, has requested a report on the following:
  - Product Sales
    - Margin analysis, by region, by product group.
  - Sales by product group
    - Sales, and budget, year over year
  - Sales Reps
    - Sales and sales quantity, by rep, by customer
  - Supply Chain
    - Inventory vs. Lead Time for all products

Additional Resources

- Pandas documentation
- DataSchool 30-video series (by a former GA instructor!)
- Qlik Consumer Sales Dataset