Python Data Analysis Lab

In this lab, you will use Pandas and other Python libraries to generate and communicate business insights from a data set.

You will have the entire class period to complete the lab. At the end of class, you'll share your final deliverable with your instructor for review. You can choose to work on the lab individually or with your classmates, but each person must turn in their own work.

Data Set Options

Choose one of the following for your lab:

Data Set	Problem Statement
Option No. 1: **Grocery Sales**	A grocery sales company is looking for patterns in its sales data regarding regions, sales reps, and product categories. Conduct an analysis to surface any notable patterns, particularly related to sales results.
Option No. 2: **Home Credit**	Home Credit is looking for patterns in its data regarding the default risk of consumer loans. Conduct an analysis to surface any notable patterns related to default risk.
Option No. 3: **College Board**	A new format for the SAT was released in March 2016. As a consultant for the College Board — the company that administers the SAT — you are tracking statewide participation to recommend where to spend money to improve participation rates. Conduct an analysis to recommend how the College Board can increase participation in a state of your choice.

Deliverables

No matter what data set you choose, you will submit a Jupyter Notebook containing:
- Imported data, including at least two DataFrames and at least one combination of DataFrames via methods such as joining.
- Functions for cleaning the data set, with explanations for how null values are being handled in each field.

● At least two visualizations accompanied by textual descriptions of the business insights they communicate.

Process

Follow this approach to start addressing your chosen problem statement. Feel free, as you begin exploring your data, to develop different questions than the ones provided. Let the data guide your analysis.

Data Set	Process
Option No. 1: **Grocery Sales**	1. Clean the data with rationale-backed handling of null or missing values. 2. Join the data sets together into a single DataFrame. 3. Analyze sales results and supply chain logistics in relation to other data points of your choosing. Based on your analysis: Propose an hypothesis around patterns and correlations revealed during the analysis.
Option No. 2: **Home Credit**	1. Clean the data with rationale-backed handling of null or missing values. 2. Join the data sets together into a single DataFrame. 3. Analyze and group the default and successful loans according to segments of your choosing. 4. Create more detailed breakdowns of a specific segment to explore more nuances. Based on your analysis: Propose a hypothesis around patterns and correlations revealed during the analysis.
Option No. 3: **College Board**	1. Clean the data with rationale-backed handling of null or missing values. 2. Join the data sets together into a single DataFrame. 3. Correlate participation rates in SAT tests with other data points of your choosing. 4. Explore and compare your state's situation with others.

	Based on your analysis: Recommend how to increase participation in your chosen state.

Grading

Your lab will be graded against the criteria laid out in this rubric.