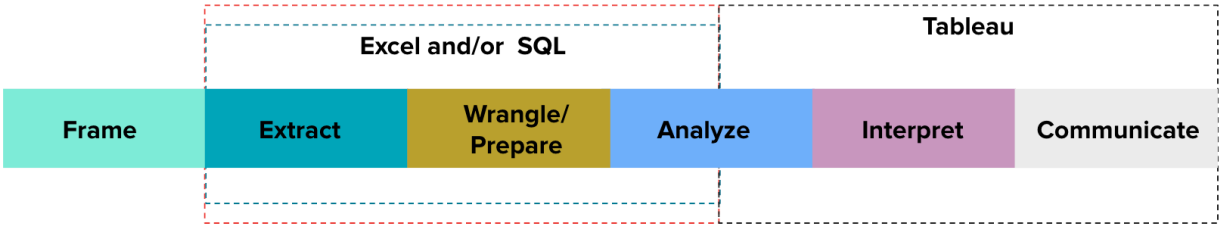
Capstone Project

For your capstone project, you'll practice a combination of Excel, SQL, and Tableau skills as you work through the Data Analytics Workflow to dissect a business problem. Here is how you'll use them in the workflow:



Deliverables

You will be asked to share your findings with your class in a five- to seven-minute presentation and submit the applicable deliverables:

1. **[Required] Data Handling Summary** (*.doc, *.txt, *.csv): Explain how you cleaned the data and why (e.g., handled nulls, removed duplicates, reformatted columns, dropped current columns, added new columns, etc.).
2. **[Required] Data Dictionary** (*.doc, *.txt, *.csv): Provide the name, description, and data type for each column.
3. **If you use Excel**, submit an Excel workbook including:
    - **Analysis and charts** used in your project, which should identify the main characteristics of the data set with supporting visualizations, tables, statistics, and correct SQL syntax where appropriate.
    - **A reference sheet** that shows analysis and charts used in your exploration that did not provide relevant insights or get added to your presentation.
4. **If you use SQL**, submit a SQL file (*.sql) containing:
    - All exploratory and analytical queries used for this project, written with correct and efficient syntax and commented throughout with context for both the query and result.
    - **If you pursue Level 3**, you should include your CREATE TABLE and COPY DATA queries as well.
5. **If you use Tableau**, submit a Tableau packaged workbook (.twbx), which you'll use for your presentation, covering:

- Background information on the problem and your approach to solving it. Your problem statement should include: goals, criteria, intended audience, and data.
- Visualizations to support your analysis.
- A list of recommendations based on your findings.
- [**Required**] Extract all data sources before saving your workbook as a Tableau packaged workbook (.twbx).

Process

You will use the Data Analytics Workflow to structure your project. Your finished deliverables should include evidence of:

- **Framing the Problem**
  - Establish a problem statement that you will explore through your analysis. A good problem is actionable and combines data fields in a meaningful and logical way. Ask yourself:
    i. What fields can I combine to find interesting insights?
    ii. What actions can someone take as a result of my analyses?
  - You can use the prompt below, or write your own!

The Problem Statement
The Regional Sales Director has noticed that — even though sales continue to experience year-over-year growth — profitability in certain regions continues to decline. Conduct an analysis to identify the root cause of the problem in each region and make data-driven recommendations based on region, customer segments, and product categories.

- **Extracting the Data in Excel or SQL**

Tool	**Data Extraction Instructions**
Excel	Use the Excel workbook (included in the Day 1 lesson) as a starting point. “Orders” should be your main data set and “Returns” can be a secondary data source. Don’t forget to review

	the data dictionary.
SQL	Access the data set in pgAdmin and review the data dictionary.

- **Wrangling and Preparing the Data in Excel and/or SQL**
  - Clean the data as appropriate by removing duplicates, addressing nulls, and reformatting data types (e.g., string to date) before you export the query results as a file and import it into Tableau.
  - Select only the columns that will be used in your analysis before you export the query results as a file and import it into Tableau.
  - Add new columns for categorical or numerical data as appropriate using your other features before you export the query results as a file (e.g., math, string, date functions).
  - Describe the steps you took to wrangle the data in a separate handling summary file.

- **Analyzing the Data in Tableau**
  - Load your data into Tableau's data source and JOIN on a unique key where appropriate.
  - **Exploratory data analysis**: Use these questions to get to know your data. Record your answers in your Tableau workbook:
    - What's the count of observations (rows) and features (columns)?
    - Are there any features that are dependent on other features in your data?
  - What's the data type of each feature — categorical or numerical?

- **Interpreting the Data**
  - Summarize relevant numbers for various data points (e.g., the average shipping speed, the product with the highest profit margin, etc.).
  - Create visualizations using the best graph format to highlight insights.
  - Create a set of individual Tableau worksheets, then build a visual analysis to address your business problem.
    - You can use the built-in Stories feature or a separate presentation tool (PowerPoint, Google Slides, etc.).

- Create at least one dashboard that combines related individual worksheets and uses Actions to highlight additional insights.
- Use the results from your analysis to compile a business recommendation that addresses the problem you sought to solve.

- **Communicating the Data**
  - Prepare a five- to seven-minute presentation to address the following:
    i. The problem you wanted to solve.
    ii. How you solved it following the Data Analytics Workflow.
    iii. Your top insights, recommended actions, and what you'd plan to do if given more time.