**Background**

Welcome to this module challenge focused on data analysis with Python's Pandas library. As a crucial component in various fields, such as data science, machine learning, and business intelligence, practical data analysis involves examining, cleaning, processing, and extracting useful information from large datasets. This assignment provides a hands-on opportunity to develop these skills.

In this challenge, we'll dive into a dataset from a fictional e-commerce company, exploring and analyzing data to address real-world business questions. Your mission will involve identifying top customers, popular product categories, calculating profits, and more. By the end of this task, you'll have a practical understanding of data exploration, transformation, and analysis, preparing you for more complex data scenarios in your future career.

**Before You Begin**

Before starting the assignment, be sure to complete the following steps:

* Create a new repository for this project called pandas-challenge-1. **Do not add this homework assignment to an existing repository**.
* Clone the new repository to your computer.
* Inside your local Git repository, add the starter files from your file downloads.
* Push these changes to GitHub or GitLab.

**Challenge Instructions**

**Part 1: Explore the Data**

In this part, you will import the data and use Pandas to learn more about the dataset.

* Import the data from the CSV file.
* View the column names.
* Use the describe function to gather some basic statistics.
* Use the provided space to explore and make yourself familiar with the data. Feel free to create more cells as needed.
* Answer the following questions using Pandas:
  + What three item categories had the most entries?
  + For the category with the most entries, which subcategory had the most entries?
  + Which five clients had the most entries in the data?
    - Store the client ids of those top 5 clients in a list.
  + How many total units (the qty column) did the client with the most entries order?

**Part 2: Transform the Data**

Now that you have a better understanding of the data you will be asked to transform the data for better and easier analysis.

* Create a column that calculates the subtotal for each line using the unit\_price and the qty.
* Create a column for shipping price. Assume a shipping price of $7 per pound for orders over 50 pounds and $10 per pound for items 50 pounds or under.
* Create a column for the total price using the subtotal and the shipping price along with a sales tax of 9.25%.
* Create a column for the cost of each line using unit cost, qty, and shipping price (assume the shipping cost is exactly what is charged to the client).
* Create a column for the profit of each line using line cost and line price.

**Part 3: Confirm Your Work**

After transforming data, it's always a good idea to verify the results. You have email receipts showing the total prices for 3 orders.

Order ID 2742071 had a total price of $152,811.89 Order ID 2173913 had a total price of $162,388.71 Order ID 6128929 had a total price of $923,441.25

* Confirm that your calculations match the receipts. Remember, each order has multiple lines.

**Part 4: Summarize and Analyze**

Use the new columns with confirmed values to find the following information.

* How much did each of the top 5 clients by quantity spend? Use your work from Part 1 for client ids?
* Create a summary DataFrame showing the totals for the top 5 clients with the following information: total units purchased, total shipping price, total revenue, and total profit.
* Create a function to change the currency to millions of dollars. Format the data and rename the columns to names suitable for presentation. Then, sort the DataFrame in descending order by total profits.
* Write a brief 2-3 sentence summary of your findings.

**Hints and Considerations**

* Use the Pandas DataFrame [documentationLinks to an external site.](https://pandas.pydata.org/docs/reference/frame.html" \t "_blank) to find helpful functions. Feel free to use functions that you didn’t learn in class if you find they make the work easier!
* Remember to make functions whenever appropriate and name them well. Functions are helpful both for performing operations and for making code more readable.
* Don’t forget to use all the steps in the analytical process! It can be especially hard to remember the steps that come before the coding starts. Define the question before you begin, and make sure to read to the end of the instructions first. If there are steps that don’t make sense yet, use the space reserved for data exploration to try and understand it.

**Requirements**

**Explore the Data (30 points)**

* View the column names. (4 points)
* Use the describe function. (4 points)
* Correctly identify the category with the most entries. (4 points)
* For the category with the most entries, correctly identify the subcategory with the most entries. (5 points)
* Correctly identify the 5 clients with the most entries in the data. (5 points)
* Store the client ids of those top 5 clients in a list. (4 points)
* Display the total units (the qty column) that the client with the most entries ordered. (4 points)

**Transform the Data (30 points)**

* Create a column that calculates the subtotal for each line using the unit\_price and the qty. (6 points)
* Create a column for shipping price. Assume a shipping price of $7 per pound for orders over 50 pounds and $10 per pound for items 50 pounds or under. (6 points)
* Create a column for the total price using the subtotal and the shipping price along with a sales tax of 9.25%. (6 points)
* Create a column for the cost of each line using unit cost, qty, and shipping price (assume the shipping cost is exactly what is charged to the client). (6 points)
* Create a column for the profit of each line using line cost and line price. (6 points)

**Part 3: Confirm Your Work (15 points)**

* Confirm that Order ID 2742071 had a total price of $152,811.89. (5 points)
* Confirm that Order ID 2173913 had a total price of $162,388.71. (5 points)
* Confirm that Order ID 6128929 had a total price of $923,441.25. (5 points)

**Part 4: Summarize and Analyze (25 points)**

* Calculate the total revenue from each of the top 5 clients in Part 1. (5 points)
* Create a summary DataFrame showing the totals for the top 5 clients with the following information: total units purchased, total shipping price, total revenue, and total profit. (5 points)
* Create a function to change the currency to millions of dollars. Format the data and rename the columns to names suitable for presentation. Then, sort the DataFrame in descending order by total profits. (5 points)
* Write a brief 2-3 sentence summary of your findings. (10 points)