Pandas Practice Problem

Problem Statement: Analyzing Customer Purchase Behavior

You have been provided with a dataset containing information about customer transactions from an online retailer. The data includes information such as customer ID, product ID, purchase date, price, and more. The goal is to use Pandas to preprocess, clean, and analyze the data to gain insights into customer purchase behavior and product popularity.

The specific tasks you might undertake could include:

Cleaning and preprocessing the data to remove duplicates, missing values, and other errors.

Merging and aggregating the data to create summary tables and metrics, such as total sales by product or customer.

Analyzing the data to identify trends and patterns in customer purchase behavior, such as which products are frequently purchased together, which products are popular at different times of year, or which customers are most valuable to the business.

Creating visualizations to help communicate the results of your analysis, such as scatter plots, bar charts, or heat maps.

This problem statement is a practical and relevant task that requires expertise in data manipulation, cleaning, and analysis using Pandas. It provides a good opportunity to practice your skills in handling complex datasets and exploring them using Pandas. Additionally, the problem is relevant and practical, making it a great task for an advanced data science learner who is interested in working with Pandas.

There are various datasets that you could use for the problem statement I suggested. Here are a few possible sources:

The Online Retail Data Set: This dataset contains transactional data of a UK-based online retailer from 2010 to 2011. You can download it from the UCI Machine Learning Repository at <https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>.

Instacart Market Basket Analysis: This dataset contains anonymized data of customer orders from the online grocery store Instacart. It includes information on orders, products, and customers. You can download it from Kaggle at <https://www.kaggle.com/c/instacart-market-basket-analysis/data>.

Amazon Customer Reviews: This dataset contains reviews and ratings of products sold on Amazon. It includes information such as product ID, customer ID, review text, and rating. You can download it from the Amazon Customer Reviews Dataset page at <https://s3.amazonaws.com/amazon-reviews-pds/readme.html>.

Note that these datasets may require some preprocessing before you can analyze them using Pandas. You may also want to explore additional datasets to find one that best suits your interests and expertise.