

Data Analysis with Pandas: Problems 01

This document contains a series of data analysis problems to be solved using the Pandas library in Python. The problems are based on a real-world use case involving daily operational data. A dataset has been provided for this assignment.

The Use Case: Analyzing Daily Operations

A business owner, Ms. Kavita, wants to analyze the daily sales data to better understand customer behavior and product performance. The data is available in a CSV file named `sales_data.csv`. Your task is to use Pandas to answer her questions. The dataset contains records of orders with details like customer name, product, quantity, unit price, and the date of the order.

Instructions

For each problem, write and execute the Python code using Pandas. The problems are designed to be solved sequentially. You can load the data once and use the same DataFrame for all questions.

Problem 1: Data Loading and Initial Inspection

Your first step is to load the provided CSV file into a Pandas DataFrame and perform an initial check to ensure the data is loaded correctly. This is a crucial first step in any data analysis workflow.

Write Python code to:

- Load the `sales_data.csv` file into a DataFrame.
- Display the first 5 rows to get a quick look at the data structure.
- Display the last 5 rows to see the end of the data.
- Print a concise summary of the DataFrame, including the data types of each column and the number of non-null values.

Problem 2: Basic Descriptive Analysis

Ms. Kavita wants to get a general overview of the dataset. Use basic Pandas functions to get a high-level summary of the data.

Write Python code to:

- Calculate the total number of orders.
- Find the total quantity of products sold.
- Calculate the total revenue (sum of 'total_price_inr').
- Find the number of unique products sold.
- Determine how many times each unique product was sold.

Problem 3: Answering Specific Business Questions with Filtering and Grouping

Ms. Kavita has some specific questions about her operations. Use filtering and grouping techniques to find the answers.

Write Python code to:

- Filter the DataFrame to show all orders made by 'Aarav'.
- Find the total revenue from 'Aarav's orders.

- Identify the product that generated the most revenue.
- Calculate the average order value for each unique customer.
- Sort the data to show the top 5 orders by revenue, from highest to lowest.

Problem 4: Combining DataFrames (Concatenation) and Time-Series Analysis

Ms. Kavita has a new dataset representing an additional week of sales data. She also wants to analyze sales trends over time.

Write Python code to:

- Create a new DataFrame for an additional week of sales. Make sure its columns match the original DataFrame.
- Concatenate the new DataFrame with the original one.
- Convert the 'order_date' column to a proper datetime format if not already done.
- Calculate the total daily revenue over the entire period.
- Find the day of the week with the highest sales on average.

Hint: The new DataFrame can be small, for example, 5 rows. You can create it manually using `pd.DataFrame()`.