**Challenging Questions to Test Your Skills**

** How would you calculate the percentage of missing values in each column of the dataset, and create a bar plot to show these percentages for all columns?**

** What approach would you use to drop rows with missing CustomerID, and how would you confirm the number of rows dropped using a summary table?**

** How would you fill missing Description values with the string “Unknown”, and visualize the count of “Unknown” descriptions versus other descriptions in a pie chart?**

** How would you identify duplicate transactions based on InvoiceNo, StockCode, and Quantity, and create a bar plot showing the number of duplicates found versus unique transactions?**

** What method would you use to remove duplicate transactions, keeping the first occurrence, and confirm their removal by checking the total row count before and after in a table?**

** How would you filter out cancellation transactions (where InvoiceNo starts with ‘C’), and visualize the proportion of cancellations versus non-cancellations in a pie chart?**

** What strategy would you use to create a TotalPrice column by multiplying Quantity and UnitPrice, and plot a histogram to show the distribution of TotalPrice?**

** How would you convert InvoiceDate to a datetime format and extract the Month as a new column, then create a line plot of transaction counts by Month?**

** How would you standardize Description by converting it to uppercase and removing extra spaces, and show a table of unique Description values before and after standardization?**

** What method would you use to extract products containing “BAG” from Description into a new IsBag column (True/False), and create a bar plot of transaction counts for IsBag True versus False?**

** How would you convert the Country column to a categorical data type, and display a table comparing memory usage before and after the conversion?**

** What approach would you take to create dummy variables for the top 5 countries by transaction count, and visualize their transaction counts in a stacked bar plot?**

** How would you bin TotalPrice into three fixed categories (‘Low’ ≤50, ‘Medium’ 50-200, ‘High’ >200), and create a count plot to show the number of transactions in each bin?**

** What method would you use to detect outliers in Quantity using the IQR method, and visualize the distribution of Quantity with box plots before and after outlier removal?**

** How would you replace negative UnitPrice values with zero, assuming they are errors, and create a histogram to compare UnitPrice distributions before and after replacement?**

** What approach would you use to rename the InvoiceNo column to OrderID for clarity, and display a sample of the dataset to confirm the change?**

** How would you create a HighValueOrder column by mapping TotalPrice >100 to True and ≤100 to False, and visualize the count of high-value versus low-value orders in a bar plot?**

** What method would you use to extract the first word from Description into a new FirstWord column using string splitting, and create a bar plot of the top 5 most frequent first words?**

** How would you bin Quantity into quartiles using equal-frequency binning, and visualize the distribution of transactions across these quartiles with a count plot?**

** How would you check for inconsistencies in Description by identifying rows where Description does not match the most common Description for each StockCode, and create a table summarizing the number of inconsistent rows by StockCode?**