**Scenarios**

Data Preprocessing

1. In the given CSV file do the preprocessing.
2. Write the steps you followed.

* import warnings
* import pandas as pd
* import matplotlib.pyplot as plt
* import seaborn as sns
* warnings.filterwarnings('ignore')
* data = pd.read\_csv(r"C:\Data Science\Yuma\Data Source (sales\_transactions).csv")

1. What was your thought process when you first saw the data.

* When first encountering the data, my initial focus was on identifying inconsistencies, such as missing values, negative numbers, and nan values. I then categorized these issues and formulated a plan to clean the data, focusing on retaining as much useful information as possible while ensuring the dataset was ready for analysis. My approach was systematic, starting with understanding the data, handling missing values, correcting errors, and performing consistency checks, all while considering the implications of each step on the quality of the final dataset.

Data Aggregation and Grouping

1. What all fields among them you think can be aggregated? Name them.

* **CustomerID, ProductCategory, PaymentMethod, Weekday**

1. What kind of aggregation (for every column) would make sense and why?

* Mean, Sum, Count

Data Validation

1. How do you know, your preprocessing was correct?
2. How will you validate your results?
3. Do you follow any specific validation process for all questions? Explain.
4. What are the edge cases you can think of?
5. What all data integrity points you want to mention for the given scenario?

Data Visualisations

1. What all projections are possible out of the data.
2. How would be know if the data is linearly projected?
3. For all the different combinations of possible projections, what are the suitable graphical representation? (Eg: Line Chart or Bar Graph)

Output everything in Python. Also, other than visualization, insert the csv in sql and do everything in sql. Attach the code.