

Task -1:

For **Task 1: Exploratory Data Analysis (EDA) and Business Insights**, the evaluation criteria are **Model Accuracy and Logic** and **Quality of Recommendations and Similarity Scores**. However, since Task 1 is focused on EDA and deriving business insights (not building a predictive model), the evaluation criteria need to be interpreted differently. Below, I'll break down how to address these criteria in the context of EDA.

1. Model Accuracy and Logic

What is Being Evaluated?

- The logic behind the data exploration and analysis.
- The accuracy of the insights derived from the data.

How to Ensure Good Model Accuracy and Logic:

1. Data Cleaning and Preparation:

- Ensure that the data is clean and ready for analysis. Handle missing values, duplicates, and inconsistencies.
- Perform necessary transformations, such as converting date columns to datetime format or encoding categorical variables.

2. Exploratory Data Analysis (EDA):

- Perform both **univariate** and **multivariate** analysis to understand the distribution of data and relationships between variables.
- Use statistical summaries (e.g., mean, median, standard deviation) and visualizations (e.g., histograms, box plots, scatter plots) to explore the data.

3. Business Insights:

- Derive insights that are actionable and relevant to the business. For example:
 - Identify the most profitable products or regions.
 - Analyze customer behavior, such as purchasing patterns or signup trends.
 - Detect anomalies or outliers in the data.

4. Validation of Insights:

- Ensure that the insights are supported by the data. For example:
 - If you claim that a certain product is the most popular, verify it by checking the sales data.
 - If you identify a trend, ensure that it is statistically significant.

2. Quality of Recommendations and Similarity Scores

What is Being Evaluated?

- The quality of the business insights derived from the EDA.
- The clarity and relevance of the recommendations.

How to Ensure High-Quality Recommendations:

1. Actionable Insights:

- The insights should be actionable and relevant to the business. For example:
 - If a certain region has low sales, recommend targeted marketing campaigns.
 - If a product category is underperforming, suggest promotions or discounts.

2. Clarity and Conciseness:

- Present the insights in a clear and concise manner. Use bullet points or short sentences to make the insights easy to understand.
- Avoid jargon and ensure that the insights are accessible to non-technical stakeholders.

3. Visualizations:

- Use visualizations to support your insights. For example:
 - Use bar charts to show sales by region.
 - Use line charts to show trends over time.
- Ensure that the visualizations are clear and well-labeled.

4. Relevance:

- Focus on insights that are relevant to the business. For example:
 - If the goal is to increase sales, focus on insights related to customer behavior and product performance.
 - If the goal is to improve customer retention, focus on insights related to customer satisfaction and loyalty.

Task -1 :

```
[35]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[43]: customers = pd.read_csv(r"C:\Users\badam\Downloads\Customers.csv")
products = pd.read_csv(r"C:\Users\badam\Downloads\Products.csv")
transactions = pd.read_csv(r"C:\Users\badam\Downloads\Transactions.csv")

print(customers.head())
print(products.head())
print(transactions.head())
```

	CustomerID	CustomerName	Region	SignupDate
0	C0001	Lawrence Carroll	South America	2022-07-10
1	C0002	Elizabeth Lutz	Asia	2022-02-13
2	C0003	Michael Rivera	South America	2024-03-07
3	C0004	Kathleen Rodriguez	South America	2022-10-09
4	C0005	Laura Weber	Asia	2022-08-15

	ProductID	ProductName	Category	Price
0	P001	ActiveWear Biography	Books	169.30
1	P002	ActiveWear Smartwatch	Electronics	346.30
2	P003	ComfortLiving Biography	Books	44.12
3	P004	BookWorld Rug	Home Decor	95.69
4	P005	TechPro T-Shirt	Clothing	429.31

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	\
0	T00001	C0199	P067	2024-08-25 12:38:23	1	
1	T00112	C0146	P067	2024-05-27 22:23:54	1	
2	T00166	C0127	P067	2024-04-25 07:38:55	1	
3	T00272	C0087	P067	2024-03-26 22:55:37	2	
4	T00363	C0070	P067	2024-03-21 15:10:10	3	

	TotalValue	Price
0	300.68	300.68
1	300.68	300.68
2	300.68	300.68
3	601.36	300.68
4	902.04	300.68