Prompt :  
hi aku mendapatkan tugas untuk mentransofrmasi data dengan case seperti ini :

The client wants to better understand user behavior and improve its ecommerce platform. The management has identified several scenarios where more data could be helpful:

1. Identify top products based on the total transaction revenue per day.

2. Detect any anomalies, such as a sharp decrease or increase in the number of transactions for a specific product.

3. Identify the most profitable city or province based on the total transaction revenue.

ini info datasetnya :

The dataset it's a part of the 'data-to-insights.ecommerce.all\_sessions' dataset from Google Big Query. Kolom pada datasetnya adalah 'fullVisitorId', 'channelGrouping', 'time', 'country', 'city',

'totalTransactionRevenue', 'transactions', 'timeOnSite', 'pageviews',

'sessionQualityDim', 'date', 'visitId', 'type', 'productRefundAmount',

'productQuantity', 'productPrice', 'productRevenue', 'productSKU',

'v2ProductName', 'v2ProductCategory', 'productVariant', 'currencyCode',

'itemQuantity', 'itemRevenue', 'transactionRevenue', 'transactionId',

'pageTitle', 'searchKeyword', 'pagePathLevel1', 'eCommerceAction\_type',

'eCommerceAction\_step', 'eCommerceAction\_option'

dan ini aturan bagaimana mengerjakannya :

Present your work in a Jupyter notebook or equivalent. For each step in the process, provide comments and explanations detailing your thought process and decision-making rationale. Show not only your final results but also your approach to each task and any challenges you encountered along the way.

Remember, we value process just as much as the final results. Therefore, it is crucial to include the specific prompts you've given to ChatGPT, the outputs it has generated, and the improvements you've made based on those outputs. Use your findings to provide insights on user behavior and recommendations for improving the ecommerce platform.

bisakah kamu membantu ku?

Response :

Sure, I can help you with this! Here's a plan to approach the tasks in your case:

**1. Identify Top Products by Total Transaction Revenue Per Day**

* **Steps:**
  1. Group the data by v2ProductName (or productSKU) and date.
  2. Sum up the transactionRevenue for each product and date.
  3. Sort the results by revenue per day.
* **Explanation**: This will help you understand which products are generating the most revenue daily, which can inform decisions on product promotions or stock levels.

**2. Detect Anomalies in Product Transactions**

* **Steps:**
  1. Group the data by v2ProductName and date, then count the number of transactions (transactions or use itemQuantity).
  2. Use statistical methods like Z-score or rolling averages to detect sharp increases or decreases in transactions.
  3. Visualize the trends for each product using line plots, highlighting any anomalies.
* **Explanation**: Identifying anomalies can help management detect unexpected changes in sales performance that might require further investigation.

**3. Identify the Most Profitable City or Province**

* **Steps:**
  1. Group the data by city (or country/province if available).
  2. Sum up the transactionRevenue for each city.
  3. Rank cities based on the total revenue.
* **Explanation**: This will provide insight into which geographic regions are driving the most revenue, potentially leading to better-targeted marketing or localized strategies.

**Structuring the Jupyter Notebook:**

For each task:

* Start with a markdown cell to describe the objective.
* Write code cells with detailed comments explaining your approach, challenges, and how you're handling the data.
* After each result, summarize your findings in markdown cells.

**Sample Code for Task 1 (Top Products by Revenue):**

# Importing necessary libraries

import pandas as pd

# Assuming 'data' is the dataframe containing the loaded dataset

# Grouping by product and date, and summing the transactionRevenue

top\_products\_per\_day = data.groupby(['v2ProductName', 'date'])['transactionRevenue'].sum().reset\_index()

# Sorting by revenue in descending order

top\_products\_per\_day = top\_products\_per\_day.sort\_values(by='transactionRevenue', ascending=False)

# Display the result

top\_products\_per\_day.head()

You can follow a similar process for the other tasks. Let me know if you'd like specific code for anomaly detection or profitable cities, or if you want to dive deeper into any part of this project!

Improvement :

* 1. Saya mempelajari terlebih dahulu terkait data yang akan saya proses, baik dari strukturnya dan datanya itu sendiri. Dari sini saya baru membuat prompt untuk ChatGPT agar response yang diberikan oleh ChatGPT lebih sesuai dengan apa yang saya butuhkan
  2. Mengganti transactionRevenue menjadi totalTransactionRevenue karna kolom transactionRevenue tidak memiliki nilai sama sekali
  3. Menambahkan visual agar data yang ditampilkan lebih mudah dilihat