Data Scientist – Assessment

Thank you for your interest in the Data Scientist role. As the next step in our process, we are sharing the following assessment. Please complete the tasks using the provided Dataset and submit your work.

Instructions:

- Use Python and SQL for your solutions.
- You may use libraries such as Pandas, Scikit-learn, PyTorch (or TensorFlow), and visualization tools.
- Ensure your solution is reproducible and well-documented.
- Document your entire thought process, including:
  * Workflow followed
  * Points where you got stuck and how you overcame them
- Submit your code, SQL queries, and a README/notes file explaining your approach.

Task 1: Data Cleaning & Modeling (SQL + Python)

1. Write SQL queries to:
   - Identify the top 5 most purchased products in the last 30 days.
   - Find customers who purchased more than 3 times in the last week.
   - Calculate the average order value per customer.

2. Using Python, design a customer segmentation model (e.g., clustering based on purchase frequency, monetary value, and recency). Provide reasoning for your choice of features and algorithm.

Task 2: Product Recommendation (ML)

1. Using transaction history data, build a collaborative filtering or content-based recommendation model.
2. Explain your approach, key trade-offs, and how you would scale it for production.
3. Implement at least one model using PyTorch (preferred) or another ML framework.

Task 3: Forecasting & Business Impact

1. Build a time-series forecasting model (ARIMA, Prophet, or ML-based) to predict daily sales for the next 30 days.
2. Visualize predictions and explain how business teams (marketing, operations) could use these forecasts.

Bonus (Optional)

- Describe how you would deploy your ML models in a cloud environment (AWS, GCP, or Azure).
- Outline monitoring metrics to track model performance over time.

Deliverables:

- Python notebooks/scripts with working code
- SQL queries in a separate file or included in your notebook
- Documentation (README/notes) covering:
  * Workflow & reasoning
  * Where you got stuck and how you overcame it