

VTI Python + SQL + RFM + KNN + Power BI + GitHub

PAST DATA Source: [RFM Analysis for Online Retail](#)

Question 1: Import Historical Labeled Data into SQL

Hint: Use flat-file import or bulk insert.

Question 2: Import Data from SQL into Python

Hint: Use pandas read_sql with SQLAlchemy or pyodbc.

Question 3: General Information

Hint: Use pandas structure/summary functions.

Question 4: Create Recency Column

Hint: Reference_date – last_purchase_date.

Question 5: Create Frequency Column

Hint: Count number of transactions per customer.

Question 6: Create Monetary Column

Hint: Sum transaction amounts per customer.

Question 7: Apply Min–Max Scaler

Hint: Use sklearn MinMaxScaler.

Question 8: Calculate Weighted RFM Score

Hint: Apply weights 0.3, 0.5, 0.2 (invert Recency).

Question 9: Assign VIP Labels

Hint: If score > 0.75 → VIP.

Question 10: Train KNN Using Historical Data

Hint: Features = R, F, M; Label = VIP; use KNeighborsClassifier.

Question 11: Predict New User IN CURRENT

Sample Historical Records (10 Customers)

Customer_ID	Recency	Frequency	Monetary	LABEL KNN ???
C001	0.92	0.10	0.15	
C002	0.88	0.20	0.30	

C003	0.40	0.60	0.70
C004	0.30	0.80	0.80
C005	0.95	0.05	0.10
C006	0.50	0.50	0.55
C007	0.25	0.90	0.85
C008	0.70	0.30	0.40
C009	0.35	0.75	0.75
C010	0.60	0.40	0.45

Question 12: Visualize in Power BI

Hint: Add RFM charts, segmentation, VIP distribution.

Question 13: Push Project to GitHub

Hint: Include README, SQL scripts, notebooks, Power BI.

ADVANCED Question 14: Assume 4 groups users, run K-Means Classification for PAST DATA SOURCE