SECTION A — Tractor Credit Risk Analysis

(5+10 marks)

Use *tractor_credit_data.csv* to complete the following tasks:

Q1. Create a file to showing the top 5 upazilas with the highest proportion of customers who paid less than 50% of their EMIs. [Excel/SQL]

Q2. Create a risk scoring system using at least 3 features from the credit dataset. Segment customers into Low, Medium, High Risk. Show: [Python/ R]

a. Number of customers in each segment
b. Average EMI delay per segment

SECTION B — Fraud Pattern in Dealership Transactions

(5+10 marks)

Use *dealer_transaction_data.csv* to complete the following tasks:

Q3. Find dealers who have more than 3 transactions flagged with either: [Excel/SQL]

a. customer_nid_duplicate_flag = 1
b. or same_phone_multiple_buyers_flag = 1

Return: dealer_id, fraud_count, total_transactions

Q4. Analyze the following [Python/ R]

a. A horizontal bar chart of top 10 dealers by suspicious activity rate.
b. Comment: What do you observe? Are there clusters?

SECTION C – Location-Based Segmentation & Marketing Targeting

(15+10 marks)

Use *yamaha_customer_data_with_geo.csv* to complete the following tasks:

Q5. Segment & Visualize Customers

You are tasked with identifying high-value customer segments for Yamaha motorcycle marketing.

1. Use Clustering Algorithm

Use features: monthly_income, purchase_intent_score, and visit_frequency to create 3 customer segments.

Then:

a. Plot the clusters using a scatterplot (e.g., Income vs Intent Score, colored by cluster)

b. Provide business names for each cluster (e.g., "Young Urban Explorers", "Budget Commuters")

2. Geospatial Visualization (NEW)

Use the latitude and longitude columns to:

a. Plot customer density using a scatter plot or heatmap. (Hint - you can use matplotlib/folium/plotly (python) or ggplot2/leaflet/ggplotly (R-programming))

b. Highlight which **upazilas** appear most attractive for targeting **sports bikes** or **premium campaigns**

Q6. High-Potential Upazilas [SQL or Python]

Find the **top 3 upazilas** where:

a. Average monthly income > 50,000

b. Average intent score > 0.6

c. AND product interest is in **Sports** or **Scooter**

Provide:

d. Upazila name, customer count, average income, and product type(s)

SECTION D — Financial Monitoring & Insight Dashboard

(20 marks)

You are part of the Yamaha Motors analytics team tasked with building **monthly financial performance dashboards** for management.

Use the provided dataset *financial_tracking_data.csv* to answer the following questions in Python dash/Tableau/Power BI/any BI tools you like.

Q7. Profitability & ROI Insights -

Choose any 2 of the following tasks. Create individual pages for each task (10 marks each):

1. Regional ROI Analysis

a. Plot average ROI by region (bar chart).

b. Highlight any underperforming regions with low ROI.

c. What strategic decision would you recommend?

2. Profit vs Marketing Spend Relationship

a. Plot a scatter plot with marketing_spend on the x-axis and net_profit on the y-axis.

b. Fit a simple linear trendline.

c. Briefly describe the strength and direction of the relationship.

3. Dealer Performance Dashboard Summary

compute the following for each dealer:

a. Total Units Sold

b. Average ROI

c. Total Net Profit

d. Region

Then show the **top 5 dealers by net profit** in a formatted table or horizontal bar chart.

SECTION E — Analyze and Visualize Customer Feedback

(25 marks)

Use the provided dataset *yamaha_mock_customer_feedback.csv* to answer the following questions

Q8. Customer Comment Analysis Using NLP -

a. Preprocessing (5 marks)

1. Clean the text: remove punctuation, lowercase, stop words, and tokenize.
2. Use stemming or lemmatization (your choice).

b. Sentiment Analysis (5 marks)

1. Use a basic lexicon-based method (e.g., VADER, TextBlob) or Hugging Face models.
2. Plot overall sentiment distribution.
3. Show sentiment by product category (bar chart or pie).

c. Topic Modeling (7.5 marks)

1. Use LDA or simple TF-IDF with KMeans to find 3–5 themes from feedback.
2. Print the top 5 keywords per topic.
3. Label the topics with a business-friendly name (e.g., "Engine Issues", "Dealer Behavior").

d. Word Cloud or Keyword Trends (5 marks)

1. Show a word cloud of most frequent keywords in positive and negative reviews.
2. Optional: Time series of frequent complaint words (if date is present).

e. Insight Summary (2.5 marks)

1. Write 3 key insights that Yamaha Motors' marketing team should know from this data.
2. Propose 1 action based on your findings.