

## Retail Data Intelligence Project – Capstone

A capstone project is your proof that you are a professional — not just a student.

### 💡💼 Why companies value capstone projects

Recruiters don't trust only certificates.

They ask:

"Show me what you can do."

A capstone project:

- Works like experience
- Goes in resume
- Used in interviews
- Used in portfolio

### ✳️ Difference between Assignment and Capstone

Assignment	Capstone Project
Small task	Full business case
Topic based	End-to-end project
Few tools	Multiple tools
Marks focused	Job focused
Academic	Industry level

This **Cypher Analytica retail analytics assignment** is fully aligned with **industry-level retail analytics & BI projects** used in hiring interviews 💼📊

This is a **real-world retail transaction dataset**, perfect for testing students on **EDA, SQL, Excel, Power BI, Python, R, and Business Analytics**.

### 1. Evaluation Rubric (100 Marks)

Section	Criteria	Marks
<b>Part A</b>	Data understanding & variable explanation	10
<b>Part B</b>	Data cleaning logic & correctness	15
<b>Part C</b>	Descriptive analysis accuracy	20
<b>Part D</b>	RFM computation & segmentation	20
<b>Part E</b>	Diagnostic reasoning	15
<b>Part F</b>	Visualizations quality & insights	10
<b>Part G</b>	Business recommendations	10
<b>Total</b>		<b>100</b>

### Grading Bands

- **90–100** → Industry ready
- **75–89** → Employable analyst
- **60–74** → Needs practice
- **< 60** → Weak analytical foundation

### Columns

Invoice No, Stock Code, Description, Quantity, Invoice Date, UnitPrice, CustomerID, Country

### Business Scenario

You are hired as a **Data Analyst** by an international online retail company.

The management wants to understand:

- Who their **best customers** are



- Which **products and countries** generate the most revenue
- How to **improve customer retention & sales**

You are given historical invoice data to analyze and provide insights.

### **Part A – Data Understanding (10 Marks)**

1. Identify:
  - Total number of rows and columns
  - Data types of each column
2. Explain the meaning of:
  - InvoiceNo
  - StockCode
  - Quantity
3. How many:
  - Unique customers?
  - Unique products?
  - Unique invoices?
  - Countries?

---

### **Part B – Data Cleaning (15 Marks)**

Perform the following:

1. Identify missing values in each column
2. Remove or treat:
  - Null CustomerID
  - Negative Quantity or UnitPrice
3. Check for:
  - Duplicate invoices
  - Zero or negative Amount

- 
4. Convert **InvoiceDate** into proper date format

---

### **Part C – Descriptive Analytics (20 Marks)**

Answer the following:

1. Total:
    - Sales (Sum of Amount)
    - Quantity sold
  2. Top 10:
    - Customers by Revenue
    - Products by Revenue
    - Countries by Revenue
  3. Monthly:
    - Sales trend
    - Number of invoices
  4. Average:
    - Order value per invoice
    - Items per invoice
- 

### **Part D – Customer Analysis (RFM) (20 Marks)**

Using **CustomerID**:

1. Compute:
  - Recency
  - Frequency (No. of invoices)
  - Monetary (Total Amount spent)
2. Segment customers into:
  - High Value

- Medium Value
  - Low Value
3. Identify:
- Top 5 loyal customers
  - Customers at risk of churn
- 

#### **Part E – Diagnostic Analysis (15 Marks)**

1. Why do some customers have very high Recency?
  2. Which products are often returned or have negative quantity?
  3. Which country shows:
    - High sales but low number of customers?
  4. Identify seasonal patterns in sales.
- 

#### **Part F – Visualization (10 Marks)**

Create at least **5 visuals**:

1. Sales by Country
2. Monthly Sales Trend
3. Top 10 Customers
4. Top 10 Products
5. RFM Segmentation chart

(Excel / Power BI / Python / R allowed)

---

#### **Part G – Business Insights (10 Marks)**

Write a short report (1–2 pages) answering:

1. Who are the most valuable customers?
2. Which products should be promoted?



3. Which country has highest growth potential?
  4. What strategy would you suggest to improve repeat purchases?
- 

### **Bonus (Optional)**

- Predict next month's sales using time-series
  - Build a churn risk flag based on Recency
  - Create a Power BI dashboard
- 

### **Deliverables**

Students must submit:

- Cleaned dataset
- Analysis file (Excel / Python / R / Power BI)
- Business insight report (PDF / Word)