

Center of Excellence

Big Data Technology Project - Requirements

In this document, all rules, and requirements on how to proceed with defining and implementing the Revature Big Data project 2 are defined.

In this project the Cohort will be divided into four teams.

- Each team will be tasked with the analysis of a large dataset (~10,000 records) using our E-Commerce data structure.
- You will have 3 weeks to complete this project.

The project timeline process:

1. By the end of the first week, each team should have produced a data generator.
 - a. This generator should output a .csv file with ~10k records.
 - b. The generator should produce data that matches the schema below.
 - c. The data should have rogue records (slightly altered data).
 - i. This can be done manually or programmatically.
2. By the end of the second week, each team should have their data visualization tool running.
 - a. You can use any visualization tool that you are comfortable with.
3. By the end of the final week, each team should have completed their data analysis.
 - a. Refer to the 'Marketing Questions' card on the Trello board.
 - b. The visualization tool should display the analytics graphically that provide clear answers to the questions that the marketing department asked.

Data Analysis Questions – Marketing Department:

- What is the top selling category of items? Per Country?
- How does the popularity of products change throughout the year? Per Country?
- Which locations see the highest traffic of sales?
- What times have the highest traffic of sales? Per Country?

E-Commerce Data Structure:

Fields (Schema)

Field name	Description
order_id	Order Id
customer_id	Customer Id
customer_name	Customer Name
product_id	Product Id
product_name	Product Name
product_category	Product Category
payment_type	Payment Type (card, Internet Banking, UPI, Wallet)
qty	Quantity ordered
price	Price of the product
datetime	Date and time when order was placed
country	Customer Country
city	Customer City
ecommerce_website_name	Site from where order was placed
payment_txn_id	Payment Transaction Confirmation Id
payment_txn_success	Payment Success or Failure (Y=Success. N=Failed)
failure_reason	Reason for payment failure

Sample Data (CSV)

```
1,101,John Smith,201,Pen,Stationery,Card,24,10,2021-01-10 10:12,India,Mumbai,www.amazon.com,36766,Y,  
2,102,Mary Jane,202,Pencil,Stationery,Internet Banking,36,5,2021-10-31 13:45,USA,Boston,www.flipkart.com,37167,Y,  
3,103,Joe Smith,203,Some mobile,Electronics,UPI,1,4999,2021-04-23 11:32,UK,Oxford,www.tatacliq.com,90383,Y,  
4,104,Neo,204,Some laptop,Electronics,Wallet,1,59999,2021-06-13 15:20,India,Indore,www.amazon.in,12224,N,Invalid CVV.  
5,105,Trinity,205,Some book,Books,Card,1,259,2021-08-26 19:54,India,Bengaluru,www.ebay.in,99958,Y,
```