

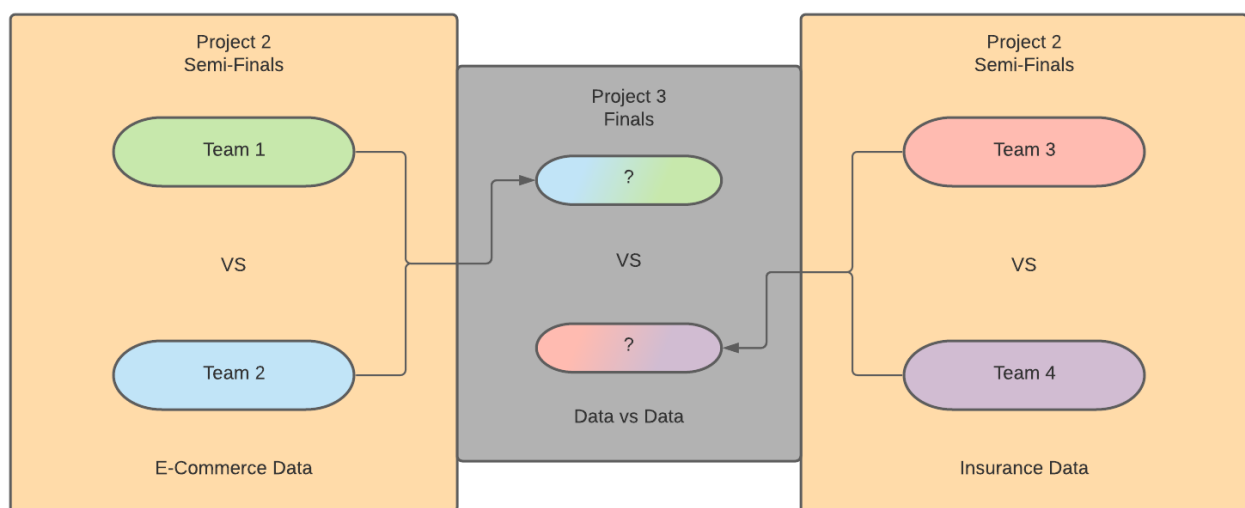
ETL – Project 2 Requirements

Outline

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

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

- Two teams will begin by creating a data producer that will generate real time data simulating orders from an E-Commerce application.
- The other two teams will begin by creating a data producer that will generate real time data simulating claims from an Insurance application.



The Center of Excellence team will designate a panel of judges for Project 2 and Project 3 presentations to determine the winners of each round. The winner of the Semi-Final round will move on to the Finals and absorb the developers of the losing team.

The panel of judges will consist of:

- The Respective Trainer
- The Respective QC Analyst
- Center of Excellence Team Member
- A Special Guest (optional)

Each team will consume the output data from the other team and run a series of ETL data movements and analysis.

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,

Insurance Data Structure:

Fields (Schema)

Field name	Description
claim_id	Claim Id
customer_id	Customer Id
customer_name	Customer Name
Customer_age	Customer Age
agent_id	Insurance Agent Id
agent_name	Insurance Agent Name
claim_category	Claim Category (Dental, Vision, Medical, Life, etc)
amount	Claim Amount (\$)
reason	Reason for Claim
agent_rating	Customer rating for their respective agent
datetime	Date and time when the claim is filed
country	Customer Country
state	Customer State
approval	Reimbursement Approved or Denied (Y=Approved. N=Denied)
reimbursement_id	Reimbursement Confirmation Id
failure_reason	Reason for reimbursement failure