# ETL – Capstone Requirements

Finals Outline

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

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

- One team, a combination of the E-Commerce Data P2 teams, will begin by combining the data they generated in Project 2. Out of the two full-stack applications created during P2, your team must select one to take into the finals.
- The other team, a combination of the Insurance Data P2 teams, will begin by combining the data they generated in Project 2. Out of the two full-stack applications created during P2, your team must select one to take into the finals.
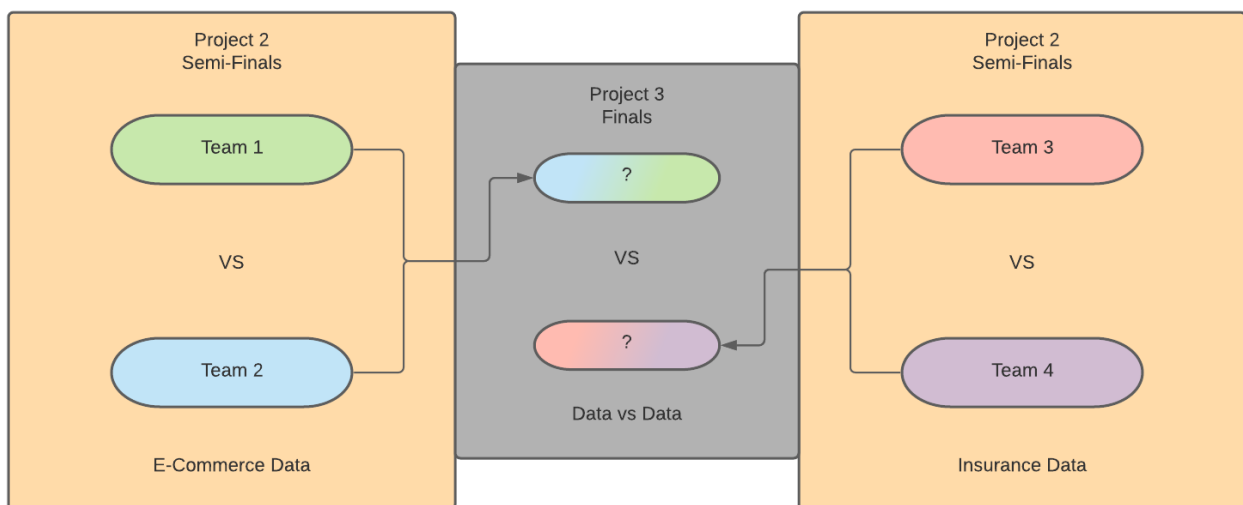
Capstone Process:

Once the data from each P2 team has been merged into its final form, the following will happen:

- The team with E-Commerce data will pass over their chosen UI and ALL the merged data generated by both teams during Project 2 to the Insurance data team.

- The team with Insurance data will pass over their chosen UI and ALL the merged data generated by both teams during Project 2 to the E-Commerce data team.

Please follow the respective backlogs for your team. The requirements that follow are contained within the 'Sprint Backlog' Column of the Trello boards.

The Center of Excellence team will designate a panel of judges for the Capstone presentations to determine the champion!

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