Project 1 Checkpoint

Sample Superstore Dataset

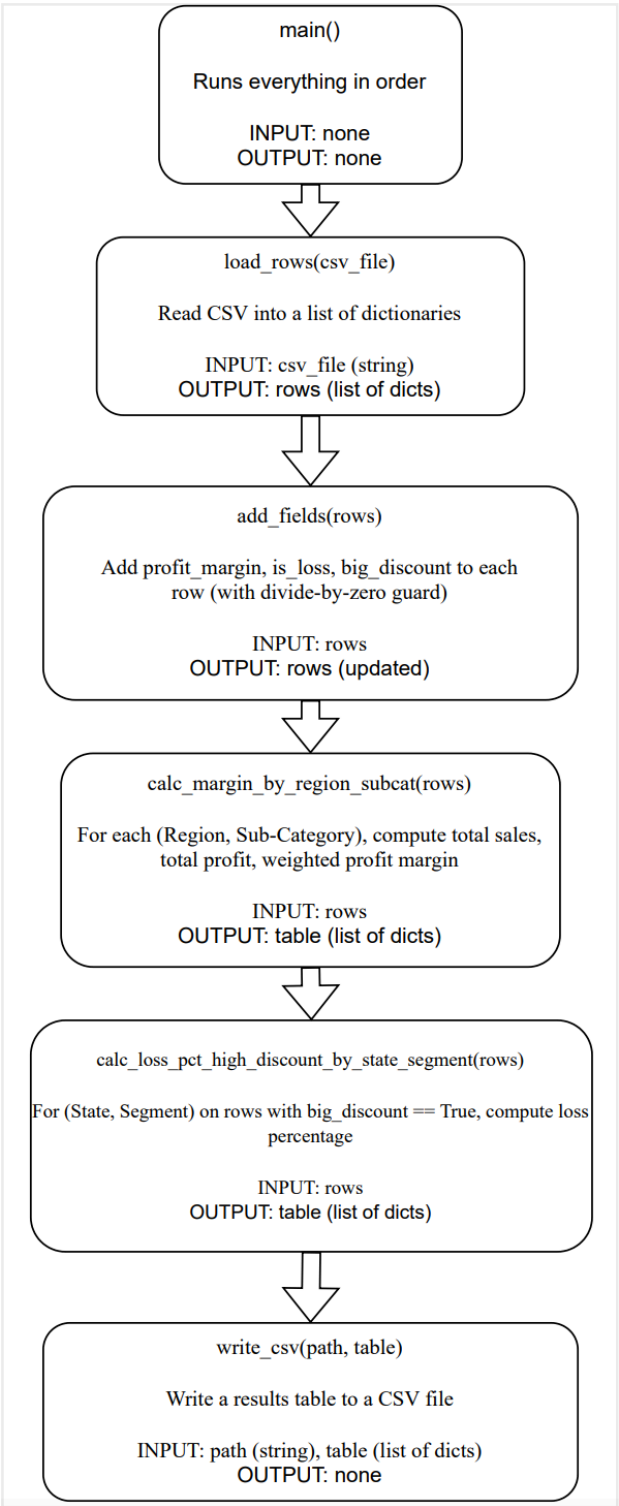
Columns I'll be working on:

- **Measures**: Sales, Profit, Quantity, Discount
- **Dimensions**: Region, State, Segment, Sub-Category, Category, Ship Mode
- **Derived fields I'll create**: profit_margin = Profit / Sales (guard Sales=0), is_loss = Profit < 0, big_discount = Discount >= 0.20

Calculations I'll perform:

- **Calculation 1**: Average profit margin by sub-category within each region
  - **Uses**: Profit, Sales, Sub-Category, Region
  - **Method**: For each (Region, Sub-Category), compute weighted margin = sum(Profit) / sum(Sales)
  - **Output file**: margin_by_region_subcategory.csv with columns: Region,Sub-Category,total_sales,total_profit,profit_margin
- **Calculation 2**: Loss rate for high-discount lines by state and segment
  - **Uses**: Discount, Profit, State, Segment
  - **Method**: Filter rows where Discount >= 0.20; for each (State, Segment), compute percent with Profit < 0
  - **Output file**: loss_pct_high_discount_by_state_segment.csv with columns: State,Segment,num_lines,num_losses,loss_pct

Diagram:



Collaborator:
Alberto Puliga