# Assignment 3: Retail Store Chain Analysis

Dataset: [Superstore Sales Dataset](https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting)

## Core Requirements:

1. Data Pipeline

* Clean retail data using Python (handling missing values, duplicates)
* Process sales history with PySpark
* Design efficient PostgreSQL schema for retail analytics

1. Transformations

* Product hierarchy organization
* Sales aggregations by region/category
* Inventory turnover calculations
* Customer purchase patterns

1. Analysis Features

* Product performance metrics
* Regional sales comparison
* Seasonal trend analysis
* Customer segment profitability
* Build KPI summary tables

## Deliverables:

* ETL pipeline code
* Analysis queries
* Project documentation