Curriculum - Product-Driven Learning with PySpark for Beginners

Day 1: PySpark Basics & Data Ingestion

- Goals: Understand SparkSession, DataFrames, CSV ingestion

- Tools: PySpark, Google Colab

- Lab: Load sales.csv, explore schema

- Assessment: 5-question quiz on PySpark basics

- Takeaway: Able to load & inspect data

Day 2: Data Cleaning & Transformation

- Goals: Handle nulls, format columns, validate types

- Tools: dropna(), withColumn(), to_date()

- Lab: Clean date column, filter bad rows

- Assessment: Fix faulty date formats

- Takeaway: Can prep data for analysis

Day 3: Aggregation & SQL Reporting

- Goals: Perform groupBy, aggregations, Spark SQL queries

- Tools: groupBy(), agg(), sql(), Parquet

- Lab: Generate 3 insights: monthly sales, top customers, avg discount

- Assessment: Write 2 SQL queries

- Takeaway: Build batch ETL pipeline