E-Commerce Data Analysis: Project Plan

Epic 0: Environment & Scoping

Setup Cloud Environment
Create BigQuery dataset, load raw data tables, and configure permissions.
Done

Setup Analysis Environment
Configure Google Colab notebook and establish connection to BigQuery.
Done

Map the Data Landscape
Perform initial data profiling and create an Entity-Relationship Diagram (ERD).
Done

Epic 1: Data Foundation

Define Analytical Layers
Architect the analysis flow: Raw data (DW) -> SQL Exploration -> Python EDA.
Done

Validate Discount Logic
Confirm discount is per-product, not per-order, and is a percentage value.
Done

Test Break-Even Hypothesis
Prove that products are never sold at their exact break-even point.
Done

Formulate Implied Cost
Develop and validate the master formula for `implied_cost_per_unit`.
Done

Epic 2: Advanced Cost Analysis

Develop Master Analytics CTE
Build a reusable SQL query joining all tables and calculating `implied_cost`.
Next Up

Analyze Cost vs. Time
Test the primary hypothesis that cost is dynamic over time using `order_date`.
To Do

Analyze Cost vs. Product Dims
Investigate cost variance grouped by `product_category` and `sub_category`.
To Do

Synthesize Cost Drivers
Document the primary factors that influence the `implied_cost` of products.
To Do

Epic 3: Business Insights & KPIs

Create Customer Profitability View
Analyze `SUM(profit)` and `SUM(sales)` grouped by `customer_segment`.
To Do

Create Product Performance View
Analyze `SUM(profit)` and `AVG(discount)` grouped by `product_category`.
To Do

Formulate Actionable Recommendations
Translate data findings into concrete business strategy suggestions.
To Do

Epic 4: Data Modeling & Governance

Establish Naming Conventions
Define and document standards for tables, views, and columns (e.g., `view_`, `dim_`, `fact_`).
To Do

Create BigQuery VIEW
Write and deploy the `CREATE VIEW` statement for the `MasterAnalyticsView`.
To Do

Document the Data Model
Create documentation explaining the view's columns, logic, and purpose.
To Do

Epic 5: Statistical Analysis (EDA)

Setup Python Environment
Connect Google Colab to BigQuery and import data from the new VIEW.
To Do

Visualize Distributions
Create histograms and box plots for key metrics to identify outliers and skew.
To Do

Generate Correlation Matrix
Create a heatmap to visualize relationships between all quantitative variables.
To Do

Epic 6: BI & Reporting

Connect Looker Studio
Establish a data source connection from Looker Studio to the BigQuery VIEW.
To Do

Design KPI Dashboard
Create mockups and layout for the main business intelligence dashboard.
To Do

Build & Deploy Dashboard
Develop the interactive charts, filters, and scorecards for stakeholders.
To Do