Final Project Plan: Generative AI for Olist Business Reports

🔷 Step 1: Dataset Overview

The Olist dataset (from Kaggle) contains **100k+ orders** across Brazil with these tables:

- **orders** → order_id, customer_id, status, timestamps (purchase, delivery)
- **order_items** → product_id, seller_id, price, freight_value
- **products** → product_category_name, weight, dimensions
- **customers** → location (city, state, zip)
- **geolocation** → mapping customers & sellers to latitude/longitude
- **payments** → payment type, value, installments
- **reviews** → review_score, comments

This richness means you can analyze **sales, operations, customers, logistics, and sentiment**.

🔷 Step 2: KPIs & Metrics to Track

Your AI-generated reports will summarize these metrics:

Sales & Revenue

- Monthly revenue, average order value (AOV), # of orders
- Top product categories & underperforming ones

Operations & Logistics

- Average delivery time vs promised time
- Late delivery %
- Freight cost analysis

Customer Metrics

- Repeat customers vs one-time buyers
- Regional performance (state, city, region heatmaps)

Review & Sentiment

- Average review score per month

- NLP sentiment from review text
- Correlation: late delivery → lower review score

◆ Step 3: Pipeline Architecture

1. Data Prep (Python + Pandas)

- Clean timestamps, merge tables (orders + items + customers + payments).
- Aggregate monthly KPIs.

2. Dashboard (Tableau or Plotly)

- Revenue trends, delivery lead time, churn curves.
- Regional heatmaps of sales.

3. Generative AI Integration (LangChain + GPT API)

- Convert KPIs into structured JSON summary (facts + anomalies).
- Feed into GPT prompt:

  > *"Write an executive report for October 2018: include revenue growth, category performance, late delivery insights, and customer sentiment. End with recommendations for operations and marketing."*

4. Streamlit App (Optional)

- Upload CSV → AI auto-writes a monthly report.
- Include a dashboard tab + report tab.

◆ Step 4: Example AI Output

Executive Summary – October 2018

- Revenue increased **12% MoM**, reaching R$3.1M.
- **Electronics** grew fastest (+18%), while **Furniture** declined (-6%).
- Average delivery time was **14.2 days**, with **8% late deliveries**, mainly in the Northeast.
- **Review sentiment dropped** slightly (avg. 3.9/5), with delays mentioned in 32% of negative reviews.

Recommendations

- Strengthen logistics partners in the Northeast to reduce late deliveries.
- Offer Furniture promotions or bundle discounts to recover sales.
- Launch a "fast delivery guarantee" for electronics buyers to build loyalty.

◆ Step 5: Deliverables for Your Professor

1. **Report/Paper**
   o Intro, problem statement, methodology, results, AI report examples, conclusion.
2. **Jupyter Notebook**
   o Data cleaning, analysis, prompt pipeline.
3. **Optional Streamlit App**
   o Live demo: upload data → AI-generated report.
4. **PowerPoint**
   o Slides with system architecture, Tableau dashboards, and sample reports.